HARQ Buffer Management: An Information-Theoretic View

Wonju Lee*, Osvaldo Simeone†, Joonhyuk Kang*, Sundeep Rangan‡ and Petar Popovski§
*Dept. of EE, KAIST, South Korea
†ECE Dept., NJIT, USA
‡ECE Dept., NYU-Poly, USA
§Dept. of Electronic Systems, Aalborg Univ., Denmark

***Abstract*—A key practical constraint on the design of Hybrid automatic repeat request (HARQ) schemes is the modem chip area that needs to be allocated to store previously received packets. The fact that, in modern wireless standards, this area can amount to a large fraction of the overall chip has recently highlighted the importance of HARQ buffer management, that is, of the use of advanced compression policies for storage of received data. This work tackles the analysis of the throughput of standard HARQ schemes, namely Type-I, Chase Combining and Incremental Redundancy, under the assumption of a finite-capacity HARQ buffer by taking an information-theoretic standpoint based on random coding. Both coded modulation, via Gaussian signaling, and Bit Interleaved Coded Modulation (BICM) are considered. The analysis sheds light on questions of practical relevance for HARQ buffer management such as on the type of information to be extracted from the received packets and on how to store it.**

***Index Terms*—HARQ buffer management, quantization, BICM.**

I. Introduction

Hybrid automatic repeat request (HARQ) is an integral part of modern wireless communication standards such as LTE. With HARQ, the receiver can store previously received packets for joint processing with the last received packet in order to enhance reliability. Three HARQ mechanisms are conventionally used, namely HARQ type I (HARQ-TI), HARQ chase combining (HARQ-CC), and HARQ incremental redundancy (HARQ-IR) (see, e.g., [1]).

Previous theoretical work on HARQ has assumed unrestricted HARQ buffers to be available at the receivers or has imposed limits on the number of packets that can be stored (see, e.g., [1], [5] and references therein). In this paper, instead, we assume a generic capacity constraint for the HARQ buffer in terms of number of bits, and we aim at addressing the following questions: (*i*) How is the relative performance of the different HARQ schemes, namely HARQ-TI, HARQ-CC and HARQ-IR, affected by the amount of available HARQ buffer capacity? (*ii*) Are there more efficient alternatives to the conventional implementation in which the buffered packets are represented by quantizing the log-likelihood ratios (LLRs) of the coded bits (see [2], [4])?

One of the key challenges in implementing HARQ is the need to store data from previously received packets. In LTE and LTE-Advanced, the HARQ buffer often becomes the main driver of the overall modem area and power consumption, as well as a key determinant of the User Equipment (UE) category level [2], [3]. Placing the HARQ buffer off chip can also be challenging due to the large bandwidth requirements on the external memory interface. This makes HARQ buffer management, meaning the use of advanced compression policies for storage of received data, of critical importance for the feasibility of HARQ in modern wireless standards [2], [4].

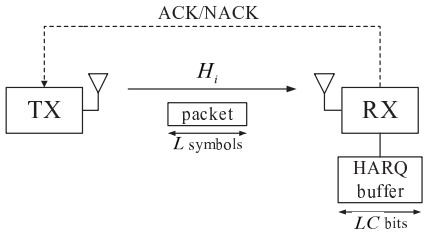

Fig. 1. HARQ with a limited-capacity HARQ buffer.

This works makes some steps towards answering these questions by taking an information-theoretic approach based on random coding. Specifically, in order to address question (*i*), we first consider a baseline system that uses an ideal coded modulation scheme via Gaussian signaling, and study the performance of HARQ-TI, HARQ-CC and HARQ-IR with baseband compression of the previously received packets in Sec. III. Then, in order to tackle question (*ii*), we investigate the more complex case of a link employing Bit Interleaved Coded Modulation (BICM) [6] and study the performance with both baseband compression and the more conventional LLR compression of the previously received packets in Sec. IV. Sec. V presents numerical results and Sec. VI some concluding remarks.

II. System Model and Performance Criteria

We consider a communication link with a single-antenna transmitter and a single-antenna receiver operating over a quasi-static fading channel via an HARQ mechanism. As illustrated in Fig. 1 and further discussed below, we make the assumption that the receiver has a limited HARQ buffer to store information extracted from the packets received in the previous (re)transmissions. Time is slotted and each slot accommodates the transmission of a packet of length L symbols. The received signal in a channel use of the i-th slot

is given by

$$Y_i = \sqrt{\text{snr}} H_i X_i + Z_i, \tag{1}$$

where the parameter snr represents the average signal to noise ratio; the channel gain H_i has unit power and changes independently slot by slot with a given cumulative distribution function (cdf) F; the input signal X_i is subject to the power constraint $E[|X_i|^2] = 1$; and we have the additive noise $Z_i \sim \mathcal{CN}(0,1)$. The receiver has an HARQ buffer with capacity LC bits, where C is the memory size normalized with respect to the packet length. The channel gain H_i is assumed to be known to the receiver, where, being a single (complex) value per packet, it is stored with a negligible buffer space.

Let us denote the maximum number of retransmission by N_{max} and the transmission rate by R, which is measured in bits/s/Hz or, equivalently, in bits/symbol. Then, the throughput T can be written as (see, e.g., [5])

$$T = \frac{R(1 - P_e^{N_{max}})}{E[N]}, \tag{2}$$

where N is the random variable that measures the number of retransmissions, including the original transmission, which satisfies $E[N] = \sum_{n=1}^{N_{max}} n\text{Pr}[N = n]$; and P_e^n is the probability of an unsuccessful transmission up to, and including, the n-th attempt. We have $\text{Pr}[N = n] = P_e^{n-1} - P_e^n$ for $n < N_{max}$ and $\text{Pr}[N = N_{max}] = P_e^{N_{max}-1}$. Therefore, it is sufficient to calculate the probabilities P_e^n for $n = 1, \ldots, N_{max}$ to characterize the throughput of any given HARQ scheme.

III. Gaussian Signaling and Baseband Compression

In this section, we evaluate the throughput of HARQ-TI, HARQ-CC, and HARQ-IR assuming a baseline scheme whereby the transmitter uses Gaussian signaling and the receiver stores in the memory compressed version of the received baseband packets.

A. HARQ-TI

With HARQ-TI, the transmitter repeatedly sends the same encoded packet and the receiver attempts decoding based solely on the last received packet. HARQ-TI hence does not make use of the receiver's HARQ buffer. The probability of an unsuccessful transmission up to the n-th attempt can be obtained as

$$P_e^n = \text{Pr}\left[\bigcap_{i=1}^{n} \left(\text{snr}|H_i|^2 \leq 2^R - 1\right)\right] = \left(F\left(\frac{2^R - 1}{\text{snr}}\right)\right)^n \tag{3}$$

B. HARQ-CC

With HARQ-CC, the transmitter repeats the same packet at each retransmission, as for HARQ-TI, but the receiver performs decoding on a packet obtained by combining all previously received packets via maximum ratio combining (MRC). HARQ-CC hence requires storage either of all previously received packets or of the current combined packet obtained from all previous transmissions. In the presence of a limited-buffer receiver, these two HARQ buffer management options yield different throughputs and are discussed next.

1) HARQ-CC Store and Combine (S&C)

A first option to implement HARQ-CC in the presence of a limited HARQ buffer is for the receiver to store all the previously received packets. Due to memory limitations, prior to storage, packets need to be compressed. To this end, the receiver divides the available memory size equally among all the packets received up to the given retransmissions and compresses each packet separately. If the n-th transmission is unsuccessful, the receiver then compresses the last received packet to LC/n bits and recompresses the previously stored packets to LC/n bits (from their previous larger size of $LC/(n-1)$ bits). We refer to this scheme as Store and Combine (S&C).

In order to account for the effect of quantization, we use the standard additive quantization noise model. Specifically, if the n-th retransmission is not successful, the quantized signals are given by

$$\hat{Y}_{i,n} = Y_i + Q_{i,n}, \tag{4}$$

for $i = 1, \ldots, n$ and $n = 1, \ldots, N_{max}$, where $Q_{i,n} \sim \mathcal{CN}(0, \sigma_{i,n}^2)$ is the quantization noise for the i-th received packet as stored at the n-th unsuccessful transmission.

Remark 1. *Quantization noise models such as (4) are used throughout this work within the information-theoretic framework of random coding. Accordingly, the results obtained in this paper are to be interpreted as implying the existence of specific (deterministic) coding and compression strategies that achieve the calculated throughput levels as long as they operate over sufficiently long block-lengths (see, e.g., [7]). The choice of a Gaussian distribution for the quantization noise is made with no claim of optimality and may be in practice justified by the fact that dithered lattice vector quantizers are able to approximate (4) with increasing accuracy as the dimensions of the quantizer increases [8].*

Following Remark 1, we relate the quantization noise $\sigma_{i,n}^2$ to the number of allocated bits LC/n via the standard rate-distortion theoretic equality [7] $C/n = I(Y_i; \hat{Y}_{i,n})$, which can be evaluated as

$$\frac{C}{n} = \log_2\left(1 + \frac{\text{snr}|H_i|^2 + 1}{\sigma_{i,n}^2}\right) \tag{5}$$

implying

$$\sigma_{i,n}^2 = \frac{\text{snr}|H_i|^2 + 1}{2^{C/n} - 1}. \tag{6}$$

The equality (5) holds also for recompressed packets, i.e. for all packets (1) with $i < n$, as long as successive refinement compression [7, Ch. 13] is employed. To briefly elaborate, each packet i is first compressed at the i-th transmission (if unsuccessful) with a number $(N_{max} - i)$ of compression layers. At later transmissions, higher layers, corresponding to refinement descriptions, are progressively discarded as n increases in order to satisfy the rate constraint C/n and effectively increasing the quantization noise (6). We refer to [9] for a detailed discussion.

At the n-th retransmission, the decoder performs MRC of the stored $(n-1)$ packets and of the last received packet prior to decoding as

$$\bar{Y}_n = H_n^* Y_n + \sum_{i=1}^{n-1} H_i^* \hat{Y}_i. \quad (7)$$

As a result, the effective SNR can be easily calculated and the probability of an unsuccessful transmission up to the n-th attempt is given by

$$P_e^n = \Pr\left[\bigcap_{j=1}^{n}\left(\frac{\text{snr}\left(\sum_{i=1}^{j}|H_i|^2\right)^2}{|H_j|^2+\sum_{i=1}^{j-1}|H_i|^2\left(\sigma_{i,j}^2+1\right)} \le 2^R-1\right)\right]. \quad (8)$$

Remark 2. *In the absence of buffer restrictions, i.e., with $C \to \infty$, we have $P_e^n = Pr\left[\sum_{i=1}^{n} snr|H_i|^2 \le 2^R-1\right]$. Therefore, under this conventional assumption, there is no need to include the intersection operation in (8). This is because, with $C \to \infty$, the effective SNR (i.e., the ratio in (8)) is a monotonically increasing function of n, while this is generally not the case for finite C due to the increasing quantization noise power (6).*

Remark 3. *The combining (7) does not account for the different noise powers affecting the combined packets due to the quantization noise. Therefore, the combining (7) is suboptimal for finite C and it reflects the operation of a standard Chase combiner (see [9] for further discussion and for improvements).*

2) HARQ-CC Combine and Store (C&S)

Instead of storing all the previously received packets, we now consider compressing and storing directly the MRC-combined packet. Specifically, at each retransmission, the last received packet is combined with the current stored packet in the HARQ buffer. If decoding is unsuccessful, the combined packet is compressed and stored. We refer to this scheme as Combine and Store (C&S).

To elaborate, if decoding is not successful at the first transmission, the compressed packet is given by

$$\hat{Y}_1 = H_1^* Y_1 + Q_1 = \sqrt{\text{snr}}|H_1|^2 X + E_1, \quad (9)$$

where $Q_1 \sim \mathcal{CN}\left(0,\sigma_1^2\right)$ is the quantization noise and $E_1 = H_1^* Z_1 + Q_1 \sim \mathcal{CN}\left(0,\rho_1^2\right)$ is the effective noise. From rate-distortion theory, similar to (6), we have $\sigma_1^2 = \left(|H_1|^2 + \text{snr}|H_1|^4\right)/\left(2^C-1\right)$ and $\rho_1^2 = |H_1|^2+\sigma_1^2$. The combined signal used in decoding at the n-th transmission is given by

$$\bar{Y}_n = H_n^* Y_n + \hat{Y}_{n-1}, \quad (10)$$

for all $n > 1$. Moreover, the stored packet at the n-th attempt, if unsuccessful, can be written as

$$\hat{Y}_n = \bar{Y}_n + Q_n = \sqrt{\text{snr}}\sum_{i=1}^{n}|H_i|^2 X + E_n, \quad (11)$$

with the effective noise given by $E_n = E_{n-1} + H_n^* Z_n + Q_n \sim \mathcal{CN}\left(0,\rho_n^2\right)$. The power of the effective noise can be expressed using the recursive relationship

$$\rho_n^2 = \rho_{n-1}^2 + |H_n|^2 + \left\{\rho_{n-1}^2 + |H_n|^2 + \text{snr}\left(\sum_{i=1}^{n}|H_i|^2\right)^2\right\}\Big/\left(2^C-1\right). \quad (12)$$

Based on (10) and (12), we can finally obtain the probability of an unsuccessful transmission up to the n-th attempt as

$$P_e^n = \Pr\left[\bigcap_{j=1}^{n}\left(\frac{\text{snr}\left(\sum_{i=1}^{j}|H_i|^2\right)^2}{|H_j|^2+\rho_{j-1}^2} \le 2^R-1\right)\right], \quad (13)$$

where we set $\rho_0 = 0$.

Remark 4. *As $C \to \infty$, the effective noise is given by $\rho_n^2 = \sum_{i=1}^{n}|H_i|^2$ and we have $P_e^n = Pr\left[\sum_{i=1}^{n} snr|H_i|^2 \le 2^R-1\right]$. The other considerations made in Remark 2 and Remark 3 apply here as well.*

C. HARQ-IR

With HARQ-IR, at each retransmission, the transmitter sends a packet consisting of new parity bits. We assume here that the receiver stores all the previously received packets as in HARQ-CC S&C. Note that the idea of storing a combined version of the previous packets as in HARQ-CC is more difficult to apply to HARQ-IR. The compressed packets at the n-th retransmission are given by (4) and (6). Since with HARQ-IR the achievable rate is the sum of the achievable rates across all transmissions (see, e.g. [5]), the probability of an unsuccessful transmission up to the n-th attempt can be obtained as

$$P_e^n = \Pr\left[\bigcap_{j=1}^{n}\left(\log_2\left(1+\text{snr}|H_j|^2\right) + \sum_{i=1}^{j-1}\log_2\left(1 + \frac{\text{snr}|H_i|^2}{1+\left(1+\text{snr}|H_i|^2\right)/\left(2^{C/(j-1)}-1\right)}\right) \le R\right)\right]. \quad (14)$$

Remark 5. *With $C \to \infty$, we obtain $P_e^n = Pr\left[\sum_{i=1}^{n}\log_2\left(1+snr|H_i|^2\right) \le R\right]$ [5].*

IV. BICM and Baseband or LLR Compression

In this section, we consider transmission based on BICM with a fixed M-ary constellation $\mathcal{X}$, where $M = 2^m$ for some integer m [6]. Throughout, we make the standard assumptions of ideal interleaving, so that the m bit channels can be assumed to be independent, and of binary i.i.d. Ber$(1/2)$ codewords for all bit channels [6]. To elaborate, we define the j-th bit in the binary label of $X \in \mathcal{X},\ j = 1,\cdots,m$, according Gray mapping, as $X(j)$, and the set $\mathcal{X}_b^j = \left\{x \in X \middle| X(j) = b\right\}$, for $b \in \{0,1\}$, of all constellation points in which the j-th bit $X(j)$ equals b. With these definitions and (1), the LLR for the

j-th bit of a symbol within the i-th retransmitted packet can be written as

$$L_i^j = \log_2 \frac{\sum_{x\in\mathcal{X}_1^j} \exp\left(-\left|Y_i - \sqrt{\text{snr}}H_i x\right|^2\right)}{\sum_{x\in\mathcal{X}_0^j} \exp\left(-\left|Y_i - \sqrt{\text{snr}}H_i x\right|^2\right)}. \quad (15)$$

In the rest of this section, we study the performance of HARQ-TI, HARQ-CC and HARQ-IR with BICM and LLR compression at the receiver. We refer to [9] for a discussion on the case of baseband compression.

A. HARQ-TI

Similar to the discussion in Sec. III-A, with HARQ-TI, the receiver decodes based only on the LLRs (15) calculated from the last received packet. With HARQ-TI, the probability of an unsuccessful transmission up to the n-th attempt can be then written as $P_e^n = \Pr\left[\sum_{j=1}^m I(X_n(j); L_n^j) \le R\right]$, which is given as

$$P_e^n = \Pr\Bigg[\frac{1}{2}\sum_{j=1}^{m}\sum_{b=0}^{1}\int f_{L_n^j|X_n(j)}(l|b) \times \log_2\left(\frac{f_{L_n^j|X_n(j)}(l|b)}{f_{L_n^j}(l)}\right) dl \le R\Bigg], \quad (16)$$

where $f_{L_n^j|X_n(j)}(l|b)$ is the conditional probability density function (pdf) of the LLR (15) given that $X_n(j) = b$, and $f_{L_n^j}(l) = 1/2\sum_{b=0}^{1} f_{L_n^j|X_n(j)}(l|b)$ is the pdf of the LLR (15). While a closed-form expression for the conditional pdf $f_{L_n^j|X_n(j)}(l|b)$ appears to be difficult to obtain, this quantity, and hence also (16), can be estimated numerically through Monte-Carlo simulations.

B. HARQ-CC

1) HARQ-CC Store and Combine (S&C)

With LLR compression, similar to Sec. III-B, HARQ-CC S&C divides the available memory equally to store the compressed LLRs of the previous received packets for all bits channels. Specifically, at the n-th transmission, if unsuccessful, the compressed LLR for the i-th transmissions and bit channel j is given as

$$\hat{L}_{i,n}^j = L_i^j + Q_{i,n}^j, \quad (17)$$

for $i = 1, \ldots, n$ and $n = 1, \ldots, N_{max}$, where we follow the same standard additive quantization noise model used in Sec. III and the quantization noise is modelled as $Q_{i,n}^j \sim \mathcal{N}(0, \sigma_{i,n,j}^2)$ (see Remark 1 for a discussion on this model). To evaluate the quantization noise variance $\sigma_{i,n,j}^2$, we resort to the information-theoretic equality $I(L_i^j; \hat{L}_{i,n}^j) = C/(mn)$, which accounts for the fact that each bit channel is allocated a memory size equal to $LC/(mn)$. Since L_i^j is not Gaussian, we leverage the following well-known upper bound (see, e.g. [7, Ch. 9])

$$I\left(L_i^j; \hat{L}_{i,n}^j\right) \le \frac{1}{2}\log_2\left(1 + \frac{\text{var}(L_i^j)}{\sigma_{i,n,j}^2}\right). \quad (18)$$

This bound allows us to obtain the conservative estimate of (i.e., upper bound on) the quantization noise power $\sigma_{i,n,j}^2$ by imposing the equality $1/2\log_2(1 + \text{var}(L_i^j)/\sigma_{i,n,j}^2) = C/(mn)$, which yields

$$\sigma_{i,n,j}^2 = \frac{\text{var}(L_i^j)}{\left(2^{2C/(mn)} - 1\right)}. \quad (19)$$

The variance $\text{var}(L_i^j)$ does not appear to admit a closed-form expression but it can be easily evaluated numerically. We observe that the estimate (19) is valid for the recompressed packets, i.e., for $i < n$, if the decoder employs successive refinement compression as discussed in Sec. III.

With HARQ-CC S&C, the combined LLR for j-th bit at the n-th attempt is given by

$$\bar{L}_n^j = L_n^j + \sum_{i=1}^{n-1} \hat{L}_{i,n}^j, \quad (20)$$

hence summing the current LLR with the previously compressed LLRs. This corresponds to the optimal combiner in the absence of quantization noise (see Remark 3). The probability of an unsuccessful transmission for HARQ-CC S&C is finally obtained as $P_e^n = \Pr\left[\bigcap_{i=1}^n \left(\sum_{j=1}^m I(X_i(j); \bar{L}_i^j) \le R\right)\right]$, which can be written as

$$P_e^n = \Pr\Bigg[\bigcap_{i=1}^{n}\Big(\frac{1}{2}\sum_{j=1}^{m}\sum_{b=0}^{1}\int f_{\bar{L}_i^j|X_i(j)}(l|b) \times \log_2\left(\frac{f_{\bar{L}_i^j|X_i(j)}(l|b)}{f_{\bar{L}_i^j}(l)}\right) dl \le R\Big)\Bigg] \quad (21)$$

and evaluated similar to (16).

2) HARQ-CC Combine and Store (C&S)

Instead of storing all the previously received LLRs, similar to Sec. III-B, HARQ-CC C&S stores the compressed value of the combined LLRs at each transmission. Specifically, if decoding of the first transmission is not successful, the stored LLR is given by $\hat{L}_1^j = L_1^j + Q_1^j$, where $Q_1^j \sim \mathcal{N}\left(0, \sigma_{1,j}^2\right)$ is the quantization noise. From the information-theoretic upper bound used in (18), we have $\sigma_{1,j}^2 = \text{var}(L_1^j)/\left(2^{2C/m} - 1\right)$. Similar to (20), combined LLR at the n-th attempt can be written as

$$\bar{L}_n^j = L_n^j + \hat{L}_{n-1}^j \quad (22)$$

for all $m > 1$. Moreover, if the n-th attempt is unsuccessful, the compressed combined LLR is given as $\hat{L}_n^j = \bar{L}_n^j + Q_n^j$, where $Q_n^j \sim \mathcal{N}(0, \sigma_{n,j}^2)$ with quantization noise power $\sigma_{n,j}^2 = \text{var}(\bar{L}_n^j)/\left(2^{2C/m} - 1\right)$, since HARQ-CC C&S allocates the available memory to store only the currently combined LLR (22). Similar to (21), the probability of an unsuccessful transmission up to the n-th retransmission is finally obtained as $P_e^n = \Pr\left[\bigcap_{i=1}^n \left(\sum_{j=1}^m I(X_i(j); \bar{L}_i^j) \le R\right)\right]$, which can be evaluated similar to (16).

C. HARQ-IR

With HARQ-IR, as discussed in Sec. III-C, the transmitter sends new parity bits at each transmission and the receiver stores the previously received LLRs by allocating the available memory as done for HARQ-CC S&C. Therefore, the compressed LLRs are given as (17) with (19). Moreover, using the fact that the achievable rate is the sum of all achievable rates in previously received packets [5], the probability of an unsuccessful transmission up to the n-th attempt can be calculated as $P_e^n = \Pr\Big[\bigcap_{i=1}^{n}\Big(\sum_{j=1}^{m}\Big(I(X_i(j);L_{i,i}^{j})+\sum_{k=1}^{i-1}I(X_k(j);\hat{L}_{k,i}^{j})\Big)\leq R\Big)\Big]$, which can be evaluated similar to (16).

V. Numerical Results

Here, we evaluate the throughput performance with Rayleigh fading. Firstly, we consider Gaussian signaling. In order to illustrate the importance of accounting for the available HARQ buffer capacity when designing the HARQ strategy (see, e.g., limited buffer rate matching in LTE [3]), we plot the throughput of HARQ-IR versus the transmission rate R with snr $= 10$ dB, $N_{max} = 10$, and different values of C. It can be seen that the optimal value of R depends significantly on the value of C, ranging from around 3.5 bits/s/Hz for $C = 1$ to $R = 8$ bits/s/Hz for $C = 10$ bits/s/Hz.

Next, we consider BICM under both baseband and LLR compression. Fig. 3 shows the throughput with $N_{max} = 10$ for 16-QAM, i.e., $M = 16$, with $R = 3.4$ bits/s/Hz and snr $= 10$ dB. The throughput gain of HARQ-IR is seen to increase with C. Moreover, HARQ-CC C&S is observed to outperform HARQ-CC S&C, suggesting that C&S is a more effective HARQ buffer management mechanism than S&C. Note that, at lower values of C, the HARQ-CC mechanism suffers from the suboptimal combining mechanism discussed in Remark 3. Finally, it is also seen that baseband compression is generally advantageous over LLR compression for values of $C > 0$ that are not large enough to make the effect of quantization immaterial, and that the relative gain is more pronounced for simpler HARQ strategies such as CC. This suggests that the use of a more sophisticated decoder, as in HARQ-IR, reduces the performance loss of a less effective compression strategy. The performance loss of LLR compression in fact increases as the size of the constellation grows larger due to the larger number of LLR values that need to be compressed [9].

VI. Conclusion

Motivated by the observation that the chip area occupied by the HARQ buffer presents a bottleneck on the performance of modern wireless modems, this work has taken an information-theoretic view of the problem of HARQ buffer management. With reference to the questions asked in the introduction, we have concluded that: (*i*) the amount of available HARQ buffer capacity determines optimal design and performance of HARQ schemes; (*ii*) storing baseband samples is advantageous over the conventional strategy of storing LLRs, suggesting that advanced compression mechanisms have the potential to reduce HARQ memory.

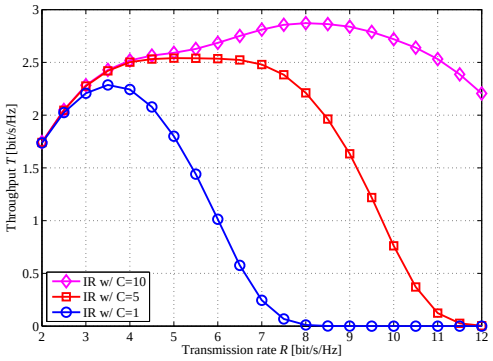

Fig. 2. Throughput T of HARQ-IR versus the transmission rate R with Gaussian signaling and baseband compression (snr $= 10$ dB and $N_{max} = 10$).

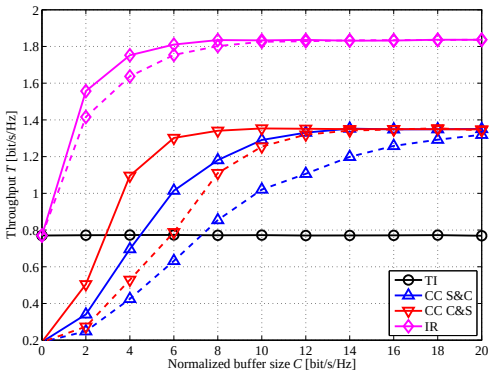

Fig. 3. Throughput T of different HARQ schemes versus the normalized buffer size C for BICM with 16-QAM for baseband compression (solid lines) and LLR compression (dashed lines) ($M = 16$, $R = 3.4$ bit/s/Hz, snr $= 10$ dB, and $N_{max} = 10$).

Acknowledgment

This research was supported by 'The Cross-Ministry Giga KOREA Project' of The Ministry of Science, ICT and Future Planning, Korea. [GK 14N0100, 5G mobile communication system development based on mmWave]. The work of O. Simeone was partially supported by WWTF Grant ICT12-054.

References

[1] S. Lin and D. J. Costello Jr., *Error control coding: Fundamentals and applications*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[2] D. Bai, et al, "LTE-advanced modem design: Challenges and persepectives," *IEEE Commun. Magazine*, vol. 50, no. 2, pp. 178-186, Feb. 2102

[3] S. Sesia, I. Toufik, and M. Baker, *LTE: the UMTS long term evolution*, Wiley Online Library, 2009.

[4] M. Danieli, et al, "Maximum mutual information vector quantization of Log-Likelihood Ratios for memory efficient HARQ implementations," in *Proc. Data Compression Conference (DCC 2010)*, pp. 30-39, Mar. 2010.

[5] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Info. Theory*, vol. 47, no. 5, pp. 1971-1988, July 2001.

[6] G. Caire, T. Politecnico, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Info. Theory*, vol. 44, no. 3, pp. 927-946, May 1998.

[7] A. El Gamal and Y. H. Kim, *Network information theory*. Cambridge University Press, 2011.

[8] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Info. Theory* vol. 42, no. 4, pp. 1152-1159, July 1996.

[9] W. Lee, O. Simeone, J. Kang, S. Rangan, and P. Popovski, "HARQ buffer management: An information-theoretic view," in preparation (see Appendix for an excerpt).