

Finite Block-length Analysis of the Incremental Redundancy HARQ

Behrooz Makki, Tommy Svensson, Michele Zorzi, *Fellow, IEEE*

Abstract—This letter studies the power-limited throughput of a communication system utilizing incremental redundancy (INR) hybrid automatic repeat request (HARQ). We use some recent results on the achievable rates of finite-length codes to analyze the system performance. With codewords of finite length, we derive closed-form expressions for the outage probabilities of INR HARQ and study the throughput in the cases with variable-length coding. Moreover, we evaluate the effect of feedback delay on the throughput and derive sufficient conditions for the usefulness of the HARQ protocols, in terms of power-limited throughput. The results show that, for a large range of HARQ feedback delays, the throughput is increased by finite-length coding INR HARQ, if the sub-codeword lengths are properly adapted.

I. INTRODUCTION

Hybrid automatic repeat request (HARQ) techniques are commonly used in wireless networks to combat the loss of data packets due to channel fading [1]–[4]. The performance of HARQ protocols is addressed in various papers, e.g., [1]–[4], where the results are obtained under the assumption of asymptotically long codewords. On the other hand, in many applications, such as vehicle-to-vehicle and vehicle-to-infrastructure communications for traffic efficiency/safety or real-time video processing for augmented reality, the codewords are required to be short (in the order of ~ 100 channel uses) [5], [6]. Thus, it is interesting to investigate the performance of HARQ protocols in the presence of finite-length codewords [7], [8].

In this letter, we study the data transmission efficiency of HARQ protocols utilizing codewords of finite length. The problem is cast as the maximization of the power-limited throughput in the presence of incremental redundancy (INR) HARQ feedback. The contributions of the paper are two-fold. 1) We use the recent results on the achievable rates of finite block-length codes [9]–[11] to analyze the throughput. With codewords of finite length, we derive closed-form expressions for the outage probabilities of the INR HARQ in different retransmission rounds and evaluate the effect of variable-length coding on the throughput. 2) We investigate the effect of feedback delay on the throughput. Particularly, we present sufficient conditions for the usefulness of HARQ protocols such that the use of HARQ increases the throughput compared to the open-loop communication setups. For a large range of HARQ feedback delays, the results show that the implementation of finite-length INR HARQ leads to throughput improvements.

Behrooz Makki and Tommy Svensson are with the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden, Email: {behrooz.makki, tommy.svensson}@chalmers.se

Michele Zorzi is with the Department of Information Engineering, University of Padova, Padova, Italy, Email: zorzi@dei.unipd.it

This work was supported in part by the Swedish Governmental Agency for Innovation Systems (VINNOVA) within the VINN Excellence Center Chase.

II. SYSTEM MODEL

Consider a point-to-point communication setup following

$$Y = \sqrt{P}hX + Z, \quad (1)$$

where P is the transmit power, X is the unit-variance input message, h denotes the fading coefficient and $Z \sim \mathcal{CN}(0, 1)$ is the independent and identically distributed (iid) complex Gaussian noise added at the receiver. We consider an INR HARQ protocol in which each data packet is sent using a maximum of M transmissions.

The system performance is studied in quasi-static conditions, e.g., [3], [4], where the channel coefficients remain constant during a packet transmission, and then change to other values according to the fading probability density function (pdf). As a motivation for this model, consider the LTE standard; as discussed in [1], each HARQ round corresponds to one transmission slot which is 0.5 milliseconds in the LTE standard. Also, for systems operating at a carrier frequency around 2.5 GHz and in the case that the receiver is moving with a speed of 2 km/h the coherence time is equal to 200 ms [1]. This coherence time is 400 times larger than the time slot duration in LTE. Thus, with proper selection of the maximum number of retransmissions, the quasi-static case can properly model the channel characteristics. (For more motivations for the quasi-static models, see [3], [4].)

Let us define the *channel gain* as $g \doteq |h|^2$. The results are given for Rayleigh fading channels where $h \sim \mathcal{CN}(0, 1)$ and, as a result, $f_g(x) = e^{-x}$ with f_g denoting the channel gain pdf. In each slot, the channel coefficient is assumed to be known by the receiver, which is an acceptable assumption in quasi-static conditions [1]–[4], [9]–[11]. However, no instantaneous channel state information is assumed to be available at the transmitter except the HARQ feedback bits.

III. ANALYTICAL RESULTS

In this section, we study the throughput of the INR HARQ protocol utilizing codewords of finite length. We first obtain a closed-form expression for the throughput, following the same procedure as in [2]–[4], except that we add the effect of feedback delay into the analysis. The throughput, given in (5), is a function of a set of probabilities that depend on the sub-codewords' length. Hence, we use the results of finite-length codes [9]–[11] to obtain the probability terms. To find the probabilities, we need to use approximation and bounding techniques, as stated in Lemmas 1 and 2. Finally, we derive bounds for the desired HARQ feedback delays such that the implementation of HARQ increases the throughput compared to the open-loop communication setup. The details are presented as follows.

Using INR HARQ, K information nats are encoded into a parent codeword of length $l_{(M)} = \sum_{m=1}^M l_m$ channel uses. Then, the parent codeword is divided into M sub-codewords of length $l_m, m = 1, \dots, M$, channel uses which are sent in the successive transmission rounds. Thus, the equivalent data rate at the end of round m is $R_{(m)} = \frac{K}{l_{(m)}}, l_{(m)} = \sum_{n=1}^m l_n, R_{(0)} \doteq \infty$. In each round, the receiver combines all received sub-codewords to decode the message. The retransmission continues until the message is correctly decoded or the maximum permitted transmission round is reached.

Let D (in channel uses) denote the HARQ feedback delay in each round. If the data transmission is stopped at the end of the m -th round, the total number of channel uses is

$$\tau_{(m)} = \begin{cases} l_{(m)} + mD, & \text{if } m \neq M \\ l_{(M)} + (M-1)D, & \text{if } m = M \end{cases} \quad (2)$$

which is based on the fact that in each retransmission round, except the last round, an acknowledgement/negative acknowledgement (ACK/NACK) signal is fed back to the transmitter. In this way, with some manipulations, the expected number of channel uses in each packet transmission period is found as

$$\mathcal{T} = \sum_{m=1}^M l_m \Omega_{m-1} + D \sum_{m=1}^{M-1} \Omega_{m-1}, \quad (3)$$

where Ω_m denotes the probability that the message is not correctly decoded up to the end of the m -th round and $\Omega_0 \doteq 1$.

If the message is correctly decoded in any round, all K information nats are received by the receiver. Thus, following the same arguments as in [2]–[4], the expected number of received information nats in each packet is given by

$$\mathcal{K} = K(1 - \Pr(\text{Outage})) = K(1 - \Omega_M), \quad (4)$$

where $\Pr(\text{Outage}) = \Omega_M$ is the packet outage probability. Using (3), (4), the *renewal-reward theorem* [2]–[4] and $l_m = \frac{K}{R_{(m)}} - \frac{K}{R_{(m-1)}}$, the throughput (in nats-per-channel-use (npcu)) is obtained as

$$\eta = \frac{\mathcal{K}}{\mathcal{T}} = \frac{1 - \Omega_M}{\sum_{m=1}^M \left(\frac{1}{R_{(m)}} - \frac{1}{R_{(m-1)}} \right) \Omega_{m-1} + \frac{D}{R_{(M)}} \sum_{m=1}^{M-1} \Omega_{m-1}}. \quad (5)$$

Here, we have defined $D^f \doteq \frac{D}{l_{(M)}}$ which is referred to as the *relative delay*, with respect to the maximum packet length.

To study the power-limited throughput of the INR protocol the final step is to calculate the probabilities $\Omega_m, m = 1, \dots, M$. In the following, we use the recent results of [9]–[11] to find Ω_m for the cases with codewords of finite length. Let us first define an (L, N, P, δ) code as the collection of

- An encoder $\Upsilon : \{1, \dots, N\} \mapsto \mathcal{C}^L$ which maps the message $n \in \{1, \dots, N\}$ into a length- L codeword $x_n \in \{x_1, \dots, x_N\}$ satisfying the power constraint

$$\frac{1}{L} \|x_j\|^2 \leq P, \forall j. \quad (6)$$

- A decoder $\Lambda : \mathcal{C}^L \mapsto \{1, \dots, N\}$ satisfying the maximum error probability constraint

$$\max_{\forall j} \Pr(\Lambda(y) \neq J | J = j) \leq \delta \quad (7)$$

with y denoting the channel output induced by the transmitted codeword according to y .

The maximum achievable rate of the code is defined as

$$R_{\max}(L, P, \delta) = \sup \left\{ \frac{\log N}{L} : \exists(L, N, P, \delta) \text{ code} \right\} \text{ (npcu)}. \quad (8)$$

Considering quasi-static conditions, [10], [11] have recently presented a very tight approximation for the maximum achievable rate (8) as

$$R_{\max}(L, P, \delta) = \sup \{R : \Pr(\log(1 + gP) < R) < \delta\} - \mathcal{O} \left(\frac{\log L}{L} \right) \text{ (npcu)}, \quad (9)$$

which, for codes of rate R npcu, leads to the following error probability [11, eq. (59)]

$$\delta(L, R, P) \approx E \left[Q \left(\frac{\sqrt{L} (\log(1 + gP) - R)}{\sqrt{1 - \frac{1}{(1+gP)^2}}} \right) \right]. \quad (10)$$

Here, $U(x) = \mathcal{O}(V(x)), x \rightarrow \infty$ is defined as $\lim_{x \rightarrow \infty} \sup | \frac{U(x)}{V(x)} | < \infty$ and $E[\cdot]$ is the expectation with respect to the channel gain g . Also, $Q(\cdot)$ denotes the Gaussian Q -function. Since the approximation (10) has been shown to be very tight for sufficiently large values of L [9]–[11], for simplicity we will assume that it is exact in the following (especially when discussing bounds).

From (9)–(10), the probability that the data is not decodable in rounds $n = 1, \dots, m$, of the INR, i.e., Ω_m , is found as

$$\Omega_m = E \left[Q \left(\frac{\sqrt{l_{(m)}} (\log(1 + gP) - R_{(m)})}{\sqrt{1 - \frac{1}{(1+gP)^2}}} \right) \right]. \quad (11)$$

Here, (11) is based on the fact that 1) with a quasi-static condition, the same fading realization g is experienced in all rounds of a packet, 2) the receiver combines all received signals of a packet to decode the message and, 3) for a given value of K nats, $A_m \subset A_n, n < m$ for quasi-static channels, where A_m is the event that the data is not decoded in rounds $1, \dots, m$.

For Rayleigh fading conditions, Ω_m is found as

$$\Omega_m = \int_0^\infty e^{-x} Q \left(\frac{\sqrt{l_{(m)}} (\log(1 + Px) - R_{(m)})}{\sqrt{1 - \frac{1}{(1+Px)^2}}} \right) dx \quad (12)$$

which does not have a closed-form expression. Lemmas 1-2 approximate/bound the probabilities $\Omega_m, \forall m$, as follows.

Lemma 1. The probabilities $\Omega_m, \forall m$, are approximated by

$$\Omega_m = \begin{cases} \frac{1}{2} \left(1 - e^{\frac{1}{P}} \sum_{i=0}^\infty \frac{1}{i!} \left(\frac{-e^{R_{(m)}}}{P} \right)^i \left(1 - \operatorname{erf} \left(\frac{K-i-1}{\sqrt{2l_{(m)}}} \right) \right) \right. \\ \quad \left. - \operatorname{erf} \left(-\frac{R_{(m)} \sqrt{l_{(m)}}}{\sqrt{2}} \right) \right), & \text{For high SNRs} \\ 1 - \frac{b_m}{\sqrt{2\pi}} e^{-\theta_m} \left(e^{\sqrt{\frac{\pi}{2b_m^2}}} - e^{-\sqrt{\frac{\pi}{2b_m^2}}} \right), & \text{For all SNRs} \end{cases}$$

where $\operatorname{erf}(\cdot)$ represents the error function, $\theta_m \doteq \frac{e^{R_{(m)}} - 1}{P}$ and

$$b_m \doteq \sqrt{\frac{l_{(m)} P^2}{e^{2R_{(m)}} - 1}}.$$

Proof: At medium/high signal-to-noise ratios (SNRs), (11) is approximated by $\Omega_m = \int_0^\infty e^{-x} Q(\sqrt{l(m)}(\log(1+xP) - R(m))) dx$ which leads to

$$\begin{aligned} \Omega_m &\stackrel{(a)}{=} \sqrt{\frac{l(m)}{2\pi}} \int_{e^{-R(m)}}^\infty \frac{1 - e^{-\frac{te^{R(m)}-1}{P}}}{t} e^{-\frac{l(m)}{2}(\log t)^2} dt \\ &\stackrel{(b)}{=} \sqrt{\frac{l(m)}{2\pi}} \int_{e^{-R(m)}}^\infty t^{-1} e^{-\frac{l(m)}{2}(\log t)^2} dt - \\ &\quad \sqrt{\frac{l(m)}{2\pi}} e^{\frac{1}{P}} \sum_{i=0}^\infty \frac{1}{i!} \left(\frac{-e^{R(m)}}{P}\right)^i \int_{e^{-R(m)}}^\infty t^{i-1} e^{-\frac{l(m)}{2}(\log t)^2} dt \\ &= \frac{1}{2} \left(1 - \operatorname{erf}\left(-\frac{R(m)\sqrt{l(m)}}{\sqrt{2}}\right) - \right. \\ &\quad \left. e^{\frac{1}{P}} \sum_{i=0}^\infty \frac{1}{i!} \left(\frac{-e^{R(m)}}{P}\right)^i \left(1 - \operatorname{erf}\left(\frac{K-i-1}{\sqrt{2l(m)}}\right)\right)\right). \end{aligned} \quad (13)$$

Here, (a) is obtained by partial integration, variable transform $t = (1+Px)e^{-R(m)}$ and the definition of the Gaussian Q -function with $Q(\infty) = 0$, $\frac{dQ(s(x))}{dx} = \frac{-1}{\sqrt{2\pi}} \frac{ds(x)}{dx} e^{-\frac{s(x)^2}{2}}$. Then, (b) follows from the Taylor expansion of the first exponential term, and the last equality is obtained by some manipulations and the definition of the error function $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$.

For the second approximation approach, we implement $Q\left(\frac{\sqrt{l(m)}(\log(1+xP) - R(m))}{\sqrt{1 - \frac{1}{(1+xP)^2}}}\right) \simeq Z_m$ with

$$Z_m(x) \stackrel{(c)}{=} \begin{cases} 1 & x \leq \theta_m - \sqrt{\frac{\pi}{2b_m^2}} \\ \frac{1}{2} - \frac{b_m}{\sqrt{2\pi}}(x - \theta_m) & x \in [\theta_m - \sqrt{\frac{\pi}{2b_m^2}}, \theta_m + \sqrt{\frac{\pi}{2b_m^2}}] \\ 0 & x \geq \theta_m + \sqrt{\frac{\pi}{2b_m^2}} \end{cases} \quad (14)$$

which results in

$$\Omega_m = \int_0^\infty e^{-x} Z_m(x) dx \stackrel{(d)}{=} 1 - \frac{b_m}{\sqrt{2\pi}} e^{-\theta_m} \left(e^{\sqrt{\frac{\pi}{2b_m^2}}} - e^{-\sqrt{\frac{\pi}{2b_m^2}}} \right). \quad (15)$$

Here, (c) is obtained by using the linearization technique for the function $Q\left(\frac{\sqrt{l(m)}(\log(1+xP) - R(m))}{\sqrt{1 - \frac{1}{(1+xP)^2}}}\right)$ at $x = \theta_m$ and (d) follows from (14) and some manipulations. ■

Lemma 2. The probabilities $\Omega_m, \forall m$, are bounded by

$$\begin{aligned} v_m &\leq \Omega_m \leq u_m, \forall m, \epsilon > 0, \\ v_m &= \frac{1}{2} \left(1 - \operatorname{erf}\left(\frac{-\theta_m b_m}{\sqrt{2}}\right) - e^{-\frac{1-2\theta_m b_m^2}{2b_m^2}} \left(1 - \operatorname{erf}\left(\frac{1-b_m^2 \theta_m}{\sqrt{2} b_m}\right)\right)\right), \\ u_m &= 1 - \frac{e^{-\theta_m + e^{-\psi_m}}}{2} + \frac{1}{2} e^{\alpha_m} P^{-\epsilon l(m)} \Gamma(1 - \epsilon l(m), \psi_m + \frac{1}{P}), \\ \psi_m &\doteq \frac{e^{R(m) + \frac{\epsilon}{2}} - 1}{P}, \alpha_m \doteq \frac{1}{P} + K\epsilon + \frac{l(m)\epsilon^2}{2}. \end{aligned} \quad (16)$$

Proof: $Q(x)$ is a decreasing function and $W(m)(x) = \frac{\sqrt{l(m)}(\log(1+xP) - R(m))}{\sqrt{1 - \frac{1}{(1+xP)^2}}}$ is concave in x . Thus, from (11), a lower bound on Ω_m is obtained if $W(m)(x)$ is replaced by its first-order Taylor expansion at any point. Using the Taylor expansion of $W(m)(x)$ at $x = \theta_m$, we have

$$\begin{aligned} \Omega_m &\geq \int_0^\infty e^{-x} Q(b_m(x - \theta_m)) dx \\ &\stackrel{(e)}{=} \int_0^\infty \frac{b_m(1 - e^{-x})}{\sqrt{2\pi}} e^{-\frac{b_m^2}{2}(x - \theta_m)^2} dx = v_m, \end{aligned} \quad (17)$$

where (e) comes from partial integration and v_m given in (16) is found by some manipulations and the definition of error function.

The upper bound is found by

$$\begin{aligned} \Omega_m &\stackrel{(f)}{\leq} 1 - e^{-\theta_m} + \frac{1}{2} \int_{\theta_m}^\infty e^{-x} e^{-\frac{l(m)}{2}(\log(1+Px) - \frac{K}{l(m)})^2} dx \\ &\stackrel{(g)}{\leq} 1 - \frac{e^{-\theta_m} + e^{-\psi_m}}{2} + \\ &\quad \frac{e^{K\epsilon + \frac{\epsilon^2 l(m)}{2}}}{2} \int_{\theta_m}^\infty (1+Px)^{-\epsilon l(m)} e^{-x} dx = u_m. \end{aligned} \quad (18)$$

Here, (f) is obtained by (12), using the inequality

$$Q(x) \leq \begin{cases} 1, & \text{if } x < 0 \\ \frac{1}{2} e^{-\frac{x^2}{2}}, & \text{if } x \geq 0 \end{cases} \quad (19)$$

and removing the denominator in (12) when $x \geq \theta_m$. Then, (g) follows from $(a-b)^2 \geq \max\{0, 2\epsilon a - 2b\epsilon - \epsilon^2\}, \forall a \geq b, \epsilon > 0$, and some manipulations. Finally, the last equality is given by the definition of the incomplete Gamma function $\Gamma(n, x) = \int_x^\infty t^{n-1} e^{-t} dt$. ■

Depending on the SNR/the sub-codewords length, the approximations/bounds in Lemmas 1-2 are useful in different conditions. Finally, to enjoy the benefits of the INR HARQ the channel code should satisfy the following requirements: 1) A parent code that can be punctured into rate-optimized sub-codewords and 2) a decoder with performance close to (11) for all retransmissions. There exist several practical finite-length code designs, e.g., [12], [13], that satisfy these requirements.

A. On the Effect of Feedback Cost

With no HARQ (open-loop setup), the parent codeword of length $l(M)$ is sent in *one shot*. Thus, with uniform power allocation, on which we focus, the outage probability is Ω_M , the same as in the HARQ-based scheme. Also, as there is no feedback, the throughput of the non-HARQ scheme is found as

$$\eta^{\text{Open-loop}} = \frac{K}{l(M)} (1 - \Pr(\text{Outage})) = R(M)(1 - \Omega_M). \quad (20)$$

From (5), (20), the intuition behind the HARQ protocols is as follows. If the channel quality is low, all possible transmissions of the HARQ are used and the system performance will be the same as in the case with no HARQ (except for the additional feedback delays). But, if the channel quality is high, the message may be correctly decoded at the end of the m -th, $m < M$ round, and the channel uses for rounds $m+1, \dots, M$ are saved. The cost of this *gambling* is the cost for feedback, i.e., the term $D \sum_{m=1}^{M-1} \Omega_{m-1}$ in (3). Hence, depending on the feedback delay and the channel conditions, using HARQ may or may not improve the throughput.

To find the acceptable range of feedback delays, we can maximize (20) for a given power and then sweep on different values of relative feedback delay, i.e., D^f in (5), to find the maximum value of D^f for which the HARQ-based approach leads to higher throughput, compared to non-HARQ scheme. Sufficient conditions for the usefulness of the HARQ, i.e., lower bounds on the acceptable range of feedback delays such that the throughput is improved by the HARQ, are obtained as follows.

For every given value of power and information nats, optimize $l_{(M)}$ in terms of the open-loop throughput (20). Consider the same packet length $l_{(M)}$ and fixed-length coding, i.e., $l_m = \frac{l_{(M)}}{M}, \forall m$, for HARQ, which is not necessarily optimal for HARQ-based scheme, in terms of throughput. Then, as the HARQ and the non-HARQ schemes have the same outage probability, the throughput is increased by HARQ if it results in less expected delay. Hence, from (3), a sufficient condition for the usefulness of HARQ is given by $\mathcal{T} \leq l_{(M)}$ leading to $D^f \leq r$ with

$$r = \frac{1 - \frac{1}{M} \sum_{m=1}^M \Omega_{m-1}}{\sum_{m=1}^{M-1} \Omega_{m-1}}. \quad (21)$$

Using Lemma 2, $\Omega_0 = 1$ and because (21) is a decreasing function of $\Omega_m, \forall m$, we have

$$\frac{M - 1 - \sum_{m=1}^{M-1} u_m}{M(1 + \sum_{m=1}^{M-2} u_m)} \leq r \leq \frac{M - 1 - \sum_{m=1}^{M-1} v_m}{M(1 + \sum_{m=1}^{M-2} v_m)} \quad (22)$$

with u_m and v_m derived in (16). Note that, while (21) provides sufficient conditions for the HARQ feedback delay with fixed-length coding, variable-length HARQ indeed does better, and larger range of feedback delays are tolerated in the variable-length coding scheme. The numerical results are presented in the sequel.

IV. NUMERICAL RESULTS AND CONCLUSIONS

According to [10], [11], the approximations in (9) and (10) are very tight for sufficiently long sub-codewords, and the tightness increases with the sub-codewords' length. For the numerical results, we consider the cases with $l_m \geq 100, \forall m$, channel uses, for which the approximation is tight enough, and we do not consider shorter sub-codewords. Our choice of $l_m \geq 100$ as the minimum possible length is motivated by [11, Fig. 2] where the relative difference of the exact and the approximate achievable rates is less than 2% for the cases with codewords of length ≥ 100 . We are further motivated for our choice of the sub-codewords length by reports such as [5], which suggest the practical codes of interest for, e.g., vehicle-to-vehicle communication to be in the range of 100 – 300 channel uses.

Shown in Fig. 1a is the throughput achieved by the variable- and the fixed-length coding INR HARQ and $D^f = 0$. Here, the results are obtained for the cases with a maximum of $M = 2$ transmissions, both the number of information nats and the sub-codewords' length are optimized in terms of throughput, and the simulation results are compared with the ones obtained via the approximation techniques of Lemma 1. Setting $K = 600$ nats, Fig. 1b demonstrates the throughput gain of the variable-length HARQ, i.e., $\Delta = \frac{\eta - \eta_{\text{Open-loop}}}{\eta_{\text{Open-loop}}} \%$ with η being the throughput in variable-length HARQ. Finally, using fixed-length coding, Fig. 1c shows the acceptable range of feedback delay and compares the results with (21) and the bounds developed in (22), i.e., when the probabilities are bounded via Lemma 2. In harmony with Lemma 1, the approximations in (13) and (15) are tight at high and all SNRs, respectively (Fig. 1a). Also, variable-length coding INR leads to throughput increment, compared to the fixed-length coding INR and the open-loop setup, especially at high SNRs (Figs. 1a-1b). As illustrated in Fig. 1c, the bounds developed in Lemma 2 and the sufficient condition in (21) are very tight for a large range of SNR. Also, the acceptable range of HARQ feedback delay, in terms of throughput, is very low at medium

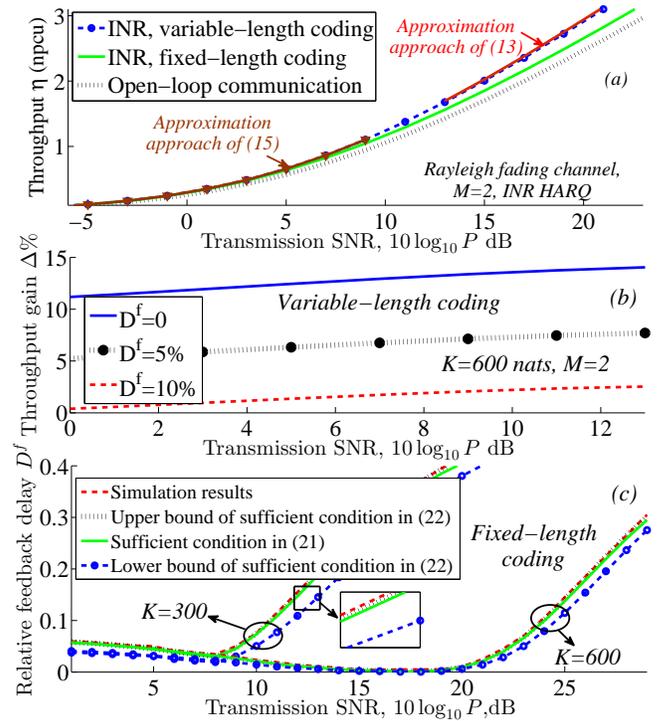


Figure 1. (a): Throughput, (b): throughput gain and (c): acceptable relative feedback delay vs SNR, $M = 2$. In subplot (a), both the number of information nats and the sub-codewords' length are optimized in terms of throughput. In subplots (b) and (c), $K = 600$ and $K = 300, 600$ nats, respectively.

SNRs while its effect is relaxed at low and high SNRs. Finally, the acceptable range of feedback delay decreases with K . To summarize, using INR HARQ with finite-length codes results in throughput increment for a large range of feedback delays, particularly when the SNR increases.

REFERENCES

- [1] A. Chelli and M. Alouini, "Performance of hybrid-ARQ with incremental redundancy over relay channels," in *GLOBECOM*, Dec. 2012, pp. 116–121.
- [2] L. Szczecinski, C. Correa, and L. Ahumada, "Variable-rate transmission for incremental redundancy hybrid ARQ," in *GLOBECOM*, Dec. 2010, pp. 1–5.
- [3] 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.
- [4] C. Shen, T. Liu, and M. P. Fitz, "On the average rate performance of hybrid-ARQ in quasi-static fading channels," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3339–3352, Nov. 2009.
- [5] K. Bilstrup, E. Uhlemann, E. G. Strom, and U. Bilstrup, "Evaluation of the IEEE 802.11p MAC method for vehicle-to-vehicle communication," in *VTC Fall*, Sept. 2008, pp. 1–5.
- [6] EU FP7 INFSO-ICT-317669 METIS project, Deliverable D1.1, "Scenarios, requirements and KPIs for 5G mobile and wireless system" Tech. Rep., April 2013. Online: <https://www.metis2020.com>.
- [7] R. Devassy, G. Durisi, P. Popovski, and E. G. Strom, "Finite-blocklength analysis of the ARQ-protocol throughput over the Gaussian collision channel," in *ISCCSP*, 2014, accepted.
- [8] B. Makki, T. Svensson, and M. Zorzi, "Green communication via Type-I ARQ: Finite block-length analysis," in *GLOBECOM*, 2014, accepted.
- [9] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [10] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static SIMO fading channels at finite blocklength," in *ISIT*, July 2013, pp. 1531–1535.
- [11] —, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, July 2014.
- [12] S. Sesia, G. Caire, and G. Vivier, "Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes," *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1311–1321, Aug. 2004.
- [13] N. Varnica, E. Soljanin, and P. Whiting, "LDPC code ensembles for incremental redundancy hybrid ARQ," in *ISIT*, Sept. 2005, pp. 995–999.