Incremental Redundancy Hybrid ARQ Schemes based on Low-Density Parity-Check Codes

Stefania Sesia[†⋆], Giuseppe Caire[⋆] and Guillaume Vivier[†]

10th February 2003

†Centre de Recherche de Motorola - Paris, Espace Technologique - Commune de Saint Aubin
91193 Gif-sur-Yvette Cedex - France
Email: {stefania.sesia,guillaume.vivier}@crm.mot.com

⋆ Institut Eurecom, 2229 Route des Crétes, B.P. 193, 06904 Sophia-Antipolis, France.
Email: {stefania.sesia,giuseppe.caire}@eurecom.fr

Note: the content of this paper has been partially presented in the Information Theory Workshop (ITW02), Bangalore, India, 2002.

Abstract

Packet-oriented data transmission is gaining more and more importance in wireless communication systems. Typically, data transmission is not strictly delay-sensitive but requires a virtually error-free link. In order to provide such level of reliability over wireless channels, affected by propagation impairments such as fading, Automatic Retransmission reQuest (ARQ) schemes can be combined with channel coding (Hybrid ARQ). In brief, when fading varies slowly over the duration of a codeword, coding takes care of the channel noise while retransmissions take care of bad channel conditions (deep fades).

In this work we study the throughput achievable by H-ARQ schemes based on *incremental redundancy* over a block-fading channel. We provide an information-theoretic analysis assuming binary random coding and typical-set decoding. Then, we study the performance of Low-Density Parity-Check code ensembles with iterative belief-propagation decoding and show that, assuming infinite block length, LDPC codes yield almost optimal performance. Unfortunately, practical finite-length LDPC codes incur a considerable performance loss with respect to their infinite-length counterpart. In order to reduce this performance loss, we propose two very effective methods: 1) using special LDPC ensembles designed to provide good *frame-error rate* (rather than just good *iterative decoding threshold*); 2) using an outer selective-repeat protocol acting on smaller packets of information bits. Surprisingly, these two apparently very different methods yield almost the same performance gain and recover a considerable fraction of the optimal throughput, thus making practical finite-length LDPC codes very attractive for data wireless communications based on incremental redundancy H-ARQ schemes.

Keywords: Hybrid ARQ Protocols, Incremental Redundancy, Data Transmission, LDPC codes, Fading Channels.

1 Introduction, system model and background

We consider a frequency-flat block-fading Gaussian channel [1] where transmission is slotted. In each slot of duration T the transmitter sends $L \approx WT$ dimensions, where W is the two-sided signal bandwidth and we assume $WT \gg 1$, and the channel gain over each slot is random but constant for the whole slot. For simplicity, we assume that the channel gains over different slots are statistically independent. Let $\mathbf{x}_s, \mathbf{y}_s \in \mathbb{C}^L$ denote the input and output signals over slot s. The block-fading channel is expressed by

$$\mathbf{y}_s = c_s \mathbf{x}_s + \nu_s \tag{1}$$

where $\nu_s \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, N_0\mathbf{I})$ is an i.i.d. complex circularly-symmetric Gaussian noise vector with per-component variance N_0, and c_s is the (scalar) fading coefficient during slot s.

The energy per symbol is constant and given by E and the fading is normalized so that $\mathbb{E}[|c_s|^2] = 1$. Therefore, the average received SNR is given by $\gamma \stackrel{\Delta}{=} E/N_0$. For later use, we define also the fading power gain $\alpha_s \stackrel{\Delta}{=} |c_s|^2$ and the instantaneous received SNR over slot s, $\beta_s \stackrel{\Delta}{=} \alpha_s\gamma$.

Although very idealized, the simple block-fading model (1) captures several aspects of wireless communications over fading channels (see the thorough discussion in [1], [2]). For example, this model applies to narrow-band transmission over a multipath fading channel with slow frequency hopping (e.g., a GSM/GPRS system [3]). As illustrated in [1, 2], when fading is slowly-varying with respect to the duration of a codeword, each codeword experiences a fixed number of fading states (M values, in our block-fading model). Under the realistic assumption of large L and small M,[1] the channel is not *information stable* and outage capacity, rather than standard ergodic capacity, describes best the limits of reliable communications. If a feedback channel is available, Automatic Retransmission reQuest (ARQ) schemes can be used in order to trade-off delay for reliability. Roughly speaking, a codeword is retransmitted until it is correctly decoded. The

[1] For example, in GSM $M = 8$ and $L \approx 100$, and in 64kbps downlink reference data channel for UMTS data-transmission modes, codewords are interleaved over $M = 2$ frames, and each frame may contains up to ≈ 1000 dimensions [4].

information-theoretic performance of Hybrid ARQ schemes [5] (i.e., schemes combining ARQ with channel coding) was studied in [6] for random Gaussian codes. In [7], the performance of H-ARQ schemes with binary convolutional codes is examined. In this work, we shall consider the very powerful class of *irregular* Low-Density Parity-Check codes, [8], [9], and show that they are very good candidates for efficient H-ARQ schemes.

The H-ARQ scheme under analysis sends additional coded symbols (redundancy) until successful decoding is achieved. For this reason, it is referred to as Incremental Redundancy (IR) protocol. The transmitter encodes information messages of b bits by using a channel code with codebook $\mathcal{C} \in \mathbb{C}^n$ of length $n = LM$ and coding rate $R = b/n$ bit/symbol. The codewords are divided in M subblocks of length L symbols. Each subblock is sent over one slot. Let $\mathcal{C}_m$ denote the punctured code of length Lm obtained from $\mathcal{C}$ by deleting the last $M-m$ subblocks. Without loss of generality, we enumerate the slots as $s = 1, 2, \ldots, M$. In order to transmit the current codeword, the transmitter sends the first subblock of L symbols on slot $s = 1$. The receiver decodes the code $\mathcal{C}_1$, by processing the corresponding received signal $\mathbf{y}_1$. If decoding is successful, a positive acknowledgement (ACK) is sent on a delay-free error-free feedback channel, the transmission of the current codeword is stopped and the transmission of the next codeword will start in the next slot (say, $s = 2$). On the contrary, if a decoding error is detected, a negative acknowledgement (NACK) is sent back and the next subblock of the current codeword is transmitted on slot $s = 2$. In this case, the receiver decodes $\mathcal{C}_2$ by processing the received signal $\{\mathbf{y}_1, \mathbf{y}_2\}$ and the same ACK/NACK procedure is repeated, until either successful decoding occurs, or all M subblocks of the current codeword are transmitted without successful decoding (see figure 1).

If successful decoding occurs after $m \leq M$ subblocks, the effective coding rate for the current codeword is $\frac{r}{m}$ bit/symbol, where we define the rate of the first block as $r \stackrel{\Delta}{=} b/L$. Therefore, the IR protocol implements easily an adaptive rate scheme that takes advantage of good instantaneous channel conditions. The throughput of the IR protocol is defined as the average number of bit/s/Hz successfully received. As far as the throughput is concerned, it is irrelevant whether codewords not successfully decoded after M subblocks are retransmitted in some successive slots or if they are just

discarded [6]. On the contrary, the packet loss rate and the average delay of the system are affected by the policy for handling decoding-failures. In general, for some information packet arrival model and some delay constraint we might seek a policy minimizing the delay subject to a packet loss probability constraint. This topic is out of the scope of the present paper and, for simplicity, we shall assume that the transmitter has an infi nite number of information packets available (no packet arrival process) and applies the IR procedure to the current packet until decoding is successful.

In the rest of this paper we are concerned with the calculation of the throughput of the IR protocol for certain random coding ensembles and deterministic LDPC code constructions. For the sake of completeness, we recall here the general throughput analysis of IR protocols given in [6]. By defi nition, the throughput is given by

$$\eta = \lim_{t\to\infty} \frac{b(t)}{Lt} \quad \text{bit/s/Hz} \tag{2}$$

where $b(t)$ is the number of successfully decoded information bits up to slot t. The event $\mathcal{E} = \{\text{The user stops transmitting the current codeword}\}$ is recognized to be a *recurrent event*[2] [10, 11, 6]. A random *reward* $\mathcal{R}$ is associated to the occurrence of the recurrent event: $\mathcal{R} = r$ bit/symbol if transmission stops because successful decoding and $\mathcal{R} = 0$ bit/symbol if it stops because at step M successful decoding has not occurred. As an application of the Renewal Theorem [10], we obtain

$$\eta = \frac{\mathbb{E}[\mathcal{R}]}{\mathbb{E}[\tau]} \tag{3}$$

where τ is the inter-renewal time expressed in slots, i.e., it is the number of slots between two consecutive occurrences of the recurrent event.

[2]**Definition:** *The $\mathcal{E}$ is a recurrent event if: (a) In order that $\mathcal{E}$ occurs at the n-th and at the $(n+m)$-th positions of the sequence $(\mathrm{E}_{j_1}, \mathrm{E}_{j_2}, \ldots, \mathrm{E}_{j_{n+m}})$ it is necessary and sufficient that $\mathcal{E}$ occurs at the last place in each of the two subsequences $(\mathrm{E}_{j_1}, \mathrm{E}_{j_2}, \ldots, \mathrm{E}_{j_n})$ and $(\mathrm{E}_{j_{n+1}}, \mathrm{E}_{j_{n+2}}, \ldots, \mathrm{E}_{j_{n+m}})$; (b) If $\mathcal{E}$ occurs at position n-th then*

$$\Pr(\mathrm{E}_{j_1}, \mathrm{E}_{j_2}, \ldots, \mathrm{E}_{j_{n+m}}) = \Pr(\mathrm{E}_{j_1}, \mathrm{E}_{j_2}, \ldots, \mathrm{E}_{j_n}) \cdot \Pr(\mathrm{E}_{j_{n+1}}, \mathrm{E}_{j_{n+2}}, \ldots, \mathrm{E}_{j_{n+m}})$$

By appropriately defining the recurrent event and the associated random reward, and by computing the expectations $\mathbb{E}[\mathcal{R}]$ and $\mathbb{E}[\tau]$, we obtain expressions for the throughput of the IR protocol under various assumptions.

2 Throughput of binary random coding

We assume perfect channel knowledge at the receiver, i.e., the receiver knows perfectly the fading coefficients $\{c_s : s = 1, \ldots, M\}$. Let the *instantaneous* mutual information per input symbol on slot s be given by

$$J(\beta_s) \stackrel{\Delta}{=} I(\mathbf{x}_s; \mathbf{y}_s | c_s) = \frac{1}{L} \mathbb{E}\left[\log_2 \frac{p(\mathbf{y}_s | \mathbf{x}_s, c_s)}{p(\mathbf{y}_s | c_s)} \right] \tag{4}$$

where $\mathbf{x}_s$ is distributed according to some input distribution $Q(\mathbf{x})$ and where

$$p(\mathbf{y}|\mathbf{x}, c) = \frac{1}{(\pi N_0)^L} e^{-\frac{1}{N_0}|\mathbf{y} - c\mathbf{x}|^2}$$

is the channel transition pdf for given fading gain c. Given the sequence of fading gains $\mathcal{F}_m \stackrel{\Delta}{=} \{c_s : s = 1, \ldots, m\}$, we define the conditional probability of decoding error after m received slots $\Pr(\text{error}|\mathcal{F}_m, \mathcal{C}_m)$ given the code $\mathcal{C}$ and the fading sequence $\mathcal{F}_m$. In [6] it is shown that there exist families of codes $\mathcal{C}$ with increasing subblock length L such that

$$\lim_{L \to \infty} \Pr(\text{error}|\mathcal{F}_m, \mathcal{C}_m) = 0 \tag{5}$$

if $I_m \stackrel{\Delta}{=} \sum_{s=1}^{m} J(\beta_s) > r$. Moreover, for any L the error probability of any code is bounded away from zero if $I_m \stackrel{\Delta}{=} \sum_{s=1}^{m} J(\beta_s) < r$. Finally, assuming typical-set decoding [12] the conditional probability of an undetected decoding error vanishes as $L \to \infty$ for any code $\mathcal{C}$ and any fading sequence $\mathcal{F}$.

Eventually, we can say that for large number of dimensions per slot L (i.e., large product WT) the error probability of the best possible code at each IR step m, for given fading sequence $\mathcal{F}_m$, is given by $\Pr(\text{error}|\mathcal{F}_m, \mathcal{C}_m) = 1\{I_m \leq r\}$ where $1\{.\}$ is the indicator function. Hence, the average error probability (where average is with respect to the fading statistics), is given by

$$\Pr(\text{error}|\mathcal{C}_m) = \Pr(I_m \leq r) \tag{6}$$

The probability $\Pr(I_m \leq r)$ will be referred to as the *information outage probability* [2] at step m. Moreover, for large L, decoding errors are revealed with arbitrarily large probability.

We define the probability $q(m)$ of successful decoding with m transmitted slots as

$$\begin{aligned} q(m) &\stackrel{\Delta}{=} \Pr(I_1 \leq r, I_2 \leq r, \ldots, I_{m-1} \leq r, I_m > r) \\ &= p(m-1) - p(m) \end{aligned} \tag{7}$$

where $p(m)$ is defined as

$$p(m) \stackrel{\Delta}{=} \Pr(I_1 \leq r, I_2 \leq r, \ldots, I_m \leq r) = 1 - \sum_{i=1}^{m} q(m) \tag{8}$$

Hence, from (3) it is immediate to obtain the throughput

$$\eta = RM \frac{1 - p(M)}{1 + \sum_{m=1}^{M-1} p(m)} \tag{9}$$

As for the average delay (in slots), we obtain[3]

$$\mu = \frac{1 + \sum_{m=1}^{M-1} p(m)}{1 - p(M)} \tag{10}$$

We apply the above throughput analysis to random binary codes, i.e., when the input distribution $Q(x)$ puts uniform probability on the binary antipodal alphabet $\{-\sqrt{E}, \sqrt{E}\}$. Since, because of non-negativity of mutual information, $(I_1, I_2, \ldots, I_m)$ is a non-decreasing sequence for all fading sequence realization, we have

$$p(m) = \Pr(I_1 \leq r, \ldots, I_m \leq r) = \Pr(I_m \leq r) = \Pr\left(\sum_{s=1}^{m} J(\beta_s) \leq r \right) \tag{11}$$

For binary inputs the instantaneous mutual information $J(\beta_s)$ is given by

$$J(\beta_s) = 1 - \int_{-\infty}^{\infty} \log_2 \left(1 + e^{4\sqrt{\beta_s}\left(z - \sqrt{\beta_s}\right)} \right) \frac{e^{-z^2}}{\sqrt{\pi}} \mathrm{d}z \tag{12}$$

[3]Expression (10) can be obtained either by simple direct calculation, or by noticing that the IR scheme where, in the presence of a decoding failure after M slot, the protocol is reset and the current codeword is transmitted again, corresponds to a newly defined renewal-reward process with deterministic reward RM. Therefore, from (3) and (9) it follows that the average inter-renewal time (i.e., the average delay) of this new process is clearly given by (10).

Since the β_s's are i.i.d. random variables, the cumulative distribution function (cdf) (11) is obtained from the m-fold convolution of the probability density function (pdf) of $J(\beta_s)$, given by

$$f(x) = \frac{1}{\gamma} f_\alpha \left(J^{-1}(x)/\gamma \right) \left(\frac{\mathrm{d}J^{-1}(x)}{\mathrm{d}x} \right) \tag{13}$$

where $f_\alpha(x)$ is the pdf of the fading power gain α.

In order to reduce the computation complexity of (11) for large m, we propose to use the Gaussian Approximation (GA)

$$p(m) \approx 1 - \mathrm{Q}\left(\frac{r - m\mu}{\sqrt{m}\sigma} \right) \tag{14}$$

where μ and σ^2 are the mean and the variance of $J(\beta_s)$.

Conventional Coded ARQ. We take a short detour to compute the throughput of conventional ARQ schemes; this will be used in section 4 to motivate the effectiveness of IR with respect to these conventional protocols. We shall consider two variants of conventional coded ARQ. In the first case, codewords of length L and rate $R = b/L$, spanning a single fading block, are used for transmission. In the presence of a decoding error (detected with arbitrarily large probability in the limit of large L), the codeword is retransmitted in some successive slot. The throughput and average delay (in slots) of this scheme with random binary code ensembles are clearly given by

$$\begin{aligned} \eta_{\mathrm{SR}-1} &= R(1 - p(1)) \\ \mu_{\mathrm{SR}-1} &= \frac{1}{1 - p(1)} \end{aligned} \tag{15}$$

where $p(1) = \Pr(J(\beta_1) \leq R)$ (consistently with (11)), and where the subscript "SR-1" indicates "*selective repeat* with coding over one block".

In the second case, codewords of length $n = LM$ and rate R are transmitted over M fading blocks and decoding is performed only after all M blocks are received. In the presence of a decoding error, the codeword is retransmitted in some successive group of M slots. The resulting throughput and

average delay are given by

$$\begin{aligned} \eta_{\text{SR}-\text{M}} &= R(1-p(M)) \\ \mu_{\text{SR}-\text{M}} &= \frac{M}{1-p(M)} \end{aligned} \tag{16}$$

where the subscript "SR-M" indicates "*selective repeat* with coding over M blocks". In Section 4, we show by some examples that the IR scheme outperforms the above SR-1 and SR-M schemes.

3 Throughput of infinite-length LDPC ensembles

LDPC codes [13] are a class of very powerful randomlike binary codes suited to low-complexity iterative decoding via the belief propagation (BP) algorithm [14]. Their bit-error rate (BER) performance under BP, in the limit of large block length, can be obtained via the *Density Evolution* (DE) method [8]. These codes exhibit a threshold phenomenon: as the block length tends to infinity, an arbitrarily small BER can be achieved if the SNR is larger than a certain threshold [8]. Otherwise, the BER is bounded away from zero for any number of decoder iterations.

In our analysis, we make the optimistic assumption that decoding is successful (the frame is error-free) with high probability if, after m received slots, the BER under BP decoding vanishes with the number of decoder iterations. Notice that vanishing BER does not necessarily implies vanishing frame-error rate (FER) in the limit of infinite block-length. However, arguments based on concatenation of LDPCs with outer *expander* codes [15] with very large rate show that, in principles, vanishing BER implies vanishing FER at least for such concatenated constructions. Furthermore, we assume that the convergence of the decoder to vanishing BER can be detected by the decoder, so that decoding failure is always revealed. Under these optimistic assumptions, we can use the same throughput formula (9) by redefining $p(m)$ as

$$p(m) = \Pr\left(\lim_{l\to\infty} \text{BER}^{(l)}(1) > 0, \ldots, \lim_{l\to\infty} \text{BER}^{(l)}(m) > 0 \right) \tag{17}$$

where $\text{BER}^{(l)}(m)$ is the BER at BP decoder iteration l with m received slots.

We assume that the reader is familiar with the standard terminology and notation of irregular LDPC codes, BP decoding and DE analysis (for details, see [8, 9, 16, 17]). The parity-check matrix of a randomly selected instance $\mathcal{C}$ in a given LDPC ensemble is conveniently represented by a bipartite graph with the nodes on the left (bitnodes) corresponding to the coded symbols and the nodes on the right (checknodes) corresponding to parity-check equations (see figure 2). A bitnode v is connected to a checknode c if the corresponding v-th symbol participates in the c-th parity equation. The LDPC ensemble is defined by its left and right degree distributions $\lambda(x) \stackrel{\Delta}{=} \sum_{i=2}^{\ell_{\max}} \lambda_i x^{i-1}$ and $\rho(x) \stackrel{\Delta}{=} \sum_{i=2}^{r_{\max}} \rho_i x^{i-1}$, where λ_i (resp., ρ_i) is the fraction of edges in the graph connected to bitnodes (resp., checknodes) of degree i. The rate of the ensemble is given by

$$R = 1 - \frac{\int_0^1 \rho(x)dx}{\int_0^1 \lambda(x)dx}$$

We assume that the coded symbols are randomly assigned to the M subblocks so that the fraction of bitnodes of degree i on each m-th subblock is the same as for the total code. In order words, the fraction of edges connected to bitnodes of degree i on subblock m is equal to λ_i/M, for all $m = 1,\ldots,M$. Numerical examples (not shown in this work for the sake of space limitation) supported our choice of distributing "uniformly" the left degrees on the subblocks.

In order to compute $\lim_{l\to\infty} \mathrm{BER}^{(l)}(m)$ for given fading coefficients $(\alpha_1,\ldots,\alpha_m)$, we resort to a Gaussian Approximation (GA) of DE, which is accurate and computationally very simple. In Appendix A we show that the condition of vanishing BER limit for given instantaneous SNRs $(\beta_1,\ldots,\beta_M)$ can be approximated by the condition that the one-dimensional dynamical system

$$z^l = \Psi\left(z^{l-1}, \beta_1, \ldots, \beta_M\right), \quad l = 1,2,\ldots \tag{18}$$

with initial condition $z^0 = 0$ has a unique fixed-point $z^\infty = 1$. The mapping function $\Psi(.)$ is given by

$$\Psi(z, \beta_1, \ldots, \beta_M) \stackrel{\Delta}{=} \frac{1}{M} \sum_{s=1}^{M} F_\lambda\left(1 - F_\rho(1-z, 0), \beta_s\right) \tag{19}$$

where, for a given distribution $g(x) = \sum_{i\geq 2} g_i x^{i-1}$, the function $F_g(z,b)$ is defined in (30).

At step m of the IR protocol, the decoder treats the not-yet received subblocks $s = m+1, \ldots, M$ as erasures, i.e., as if the received signal was zero. In the DE-GA (Gaussian Approximation applied to Density Evolution), this corresponds to letting $\beta_s = 0$ for $s = m+1, \ldots, M$. The condition that (18) has unique fi xed-point equal to 1 is equivalent to (see Appendix A)

$$\Psi(z, \beta_1, \ldots, \beta_M) > z, \quad \forall\ z \in [0,1) \tag{20}$$

Moreover, it is immediate to see that $\Psi(z, \beta_1, \ldots, \beta_M)$ is an increasing function of $\beta_s, s = 1, \ldots, M$ for any given $z \in [0,1]$. Hence, for any fading gain sequence $(\alpha_1, \ldots, \alpha_M)$ and any $z \in [0,1]$ we have that, for any m, the condition

$$\exists\, z \in [0,1) \ \text{ such that } \ \Psi\left(z, \gamma\alpha_1, \ldots, \gamma\alpha_m, \underbrace{0, \ldots, 0}_{M-m}\right) < z$$

implies that

$$\exists\, z \in [0,1) \ \text{ such that } \ \Psi\left(z, \gamma\alpha_1, \ldots, \gamma\alpha_{m'}, \underbrace{0, \ldots, 0}_{M-m'}\right) < z$$

for all $1 \leq m' < m$. Due to this inclusion, $p(m)$ in (17), under the DE-GA approximation, reduces to

$$p(m) = 1 - \Pr\left(\Psi\left(z, \gamma\alpha_1, \ldots, \gamma\alpha_m, \underbrace{0, \ldots, 0}_{M-m}\right) > z, \ \ \forall\, z \in [0,1)\right) \tag{21}$$

The probability in the RHS of (21) can be evaluated by Monte Carlo simulation by generating the sequence $\alpha_1, \ldots, \alpha_M$ i.i.d., distributed according to the fading pdf $f_\alpha(x)$.

4 Results: Achievable Throughput

In all our numerical examples we assume Rayleigh fading, i.e., $f_\alpha(x) = e^{-x}$, and $M = 10$ fading blocks. Figures 3 and 4 show the throughput of binary random codes when the IR and SR scheme are used, as a function of the coding rate R for $\gamma = 3$ and 10 dB, respectively. The throughput of binary random codes with IR is evaluated by computing the $p(m)$'s via convolution. For the

sake of comparison, the results of GA are also shown: remarkably, GA yields a very accurate approximation.

The comparison between IR and SR protocols is more evident by plotting the average delay vs. the throughput (see figure 5). From the analysis of Section 2, η and or μ are given as functions of the code rate $R \in [0,1]$ (for given number of fading blocks M, fading gain statistics and SNR γ). Hence, the curve $\mu = \mu(\eta)$ can be obtained in parametric form, by letting R varying in the interval $[0,1]$. Since η is a non-monotone function of R, each value of η corresponds to possibly multiple values of μ. Clearly, in the presence of multiple values only the minimum is relevant. Figure 5 shows that SR-M is not convenient. On the contrary, for a certain range of throughputs SR-1 achieves almost the same average delay of IR. However, there is a range of high throughput that is not achievable by SR-1 while it can be achieved by the IR protocol at the cost of a very small average delay (6 or 7 slots).

Figures 3 and 4 show also results for infinite block length LDPC codes. Each mark ($*$) in figures 3 and 4 is obtained by using an irregular LDPC ensemble with degree distributions λ, ρ optimized for the corresponding rate R and for the standard unfaded AWGN channel [18, 8]. No attempt was made to optimize the degree distributions to take into account the block-fading channel. Nevertheless, these results show that AWGN-optimized ensembles perform close to optimal and not much can be gained by further ensemble optimization.

5 LDPCs with finite block length

At this point, it is natural to ask how practical finite-length LDPC code perform on the block-fading channel under the IR protocol, by removing the optimistic assumptions (limit for large L, vanishing BER $\Rightarrow$ vanishing FER) that led to the outstanding results of the previous section. Figures 3 and 4 show also the throughput obtained by simulation of the IR protocol by using actual finite-length LDPC codes of length $n = 5000$ and $n = 10000$. The finite-length results are obtained by averaging over the channel fading, the noise and the ensemble of codes, i.e., a new parity-

check matrix is randomly generated according to the given left and right degree distributions λ, ρ defining the ensemble for each transmitted information packet. The throughput formula for finite-length codes is still given by (9) where $p(m)$, for a given LDPC ensemble with degree distributions λ, ρ, is expressed by

$$p(m) = \mathop{\mathbb{E}}_{\mathcal{C}(n,\lambda,\rho),\boldsymbol{\alpha}} [Pr(\overline{\mathcal{A}}_1, \overline{\mathcal{A}}_2, \ldots, \overline{\mathcal{A}}_m \,|\boldsymbol{\alpha}, (\lambda, \rho)\,]} \tag{22}$$

where, $\boldsymbol{\alpha}$ is the sequence of fading gains, $\mathcal{A}_s$ is the event of successful decoding at step s and where the code parity-check matrix is randomly generated with uniform probability over all bipartite graphs with degree distributions λ, ρ (see [8]). Successful decoding is defined by the event that, after a given maximum number of BP decoder iterations, all information bits are correct.

The throughput performance loss of finite-length ensembles with respect to their infinite-length counterpart can be explained by observing that, typically, irregular finite-length LDPC codes with bitnodes of degree 2 have very poor FER performance, despite the fact that they perform well in terms of BER. This is because typical decoding errors involve a very small number of bit-errors per frame error.

Another remarkable fact evidenced by figures 3 and 4 is that codes with block length $n = 5000$ slightly outperform codes with $n = 10000$. This is surprising since in standard AWGN settings (without ARQ) BER is known to improve with the code block length [8]. Indeed, irregular LDPC codes are commonly believed to provide good performances only for extremely large block length. The above results show that in the presence of time-varying channels and retransmission schemes this is not the case, as FER and not BER determines the throughput performance.

Next, we propose two approaches to improve the performance of IR with finite-length practical LDPC codes. The first approach acts directly on the code design and leaves the IR protocol unchanged: it consists of selecting the code parity-check matrix in some appropriate ensemble with good FER properties. The second approach acts on the IR protocol and leaves the code design unchanged: it consists of dividing the information packet into subpackets, performing error detection on each of the subpackets and using an outer selective-repeat protocol only for the subpackets

in error. Interestingly, although quite different, these approaches yield very similar performance improvement and recover a considerable fraction (up to 80% at SNR $= 10$ dB) of the loss due to finite block length.

Special graph construction. Solutions to improve the FER performance of LDPCs consist of finding special constructions based on expander graphs (see for example [15]), or a deterministic arrangement of the edges adjacent to degree-2 bitnodes [18]. Due to its simplicity, we follow this second method. Good FER codes can be obtained constructing the graph such that the edges emanating from a bitnode of degree 2 are placed semi-deterministically. Let R denote the rate of the code and $\tilde{\lambda}_2 = \frac{\lambda_i/i}{\sum_j \lambda_j/j}$ be the fraction of bitnodes of degree 2. If $\frac{(1-R)}{2} < \tilde{\lambda}_2 < 1-R$ we can arrange the $\tilde{\lambda}_2 n$ deg-2 bitnodes and $\tilde{\lambda}_2 n$ checknodes into a cycle of girth $2\tilde{\lambda}_2 n$, as shown in the example of figure 6.

As an example of this construction, consider a standard unfaded AWGN channel and the ensemble with λ and ρ given in [18], for rate $R = 0.3$ bit/symbol, maximum left degree $d_v = 100$, average right degree $a_r = 6.9$, and block length $n = 10000$. Figure 7 shows the BER and the FER obtained by averaging over all graphs with given degree distributions (Total ensemble) and by averaging over all graphs with special cyclic arrangement of the edges connected with degree-2 bitnodes (Modified ensemble). It is clear that the modified ensemble yields much better FER.

Outer Selective Repeat System. Our second approach stems from the following observation: for standard irregular LDPC codes, most frame errors involve a very small number of bit errors. Therefore, by dividing the information packet into smaller subpackets, only a few of them will contain errors after decoding. Hence, an Outer Selective-Repeat (OSR) protocol acting on these smaller subpacket units can recover subpacket errors without having to retransmit the whole codeword. For the sake of simplicity, we make the optimistic assumption that subpackets errors can be prefectly detected. The concept of the concatenated selective-repeat scheme is represented in figure 8.

Let P denote the subpacket length in bits, and $n_p = b/P$ be the number of subpackets per LDPC codeword. At step m of the IR protocol, after a given number of decoder iterations, let e_m denote the number of subpackets in error. We shall consider "successful" decoding (i.e., the IR protocol stops the transmission of the current codeword at step m) if $e_m \leq \delta$. Otherwise, if $e_m > \delta$, a NACK is sent and the block $m+1$ of the current codeword is sent on the next slot. The system throughput can be optimized with respect to the threshold $\delta \in [0, n_p]$. Notice that setting $\delta = 0$ is equivalent to the IR alone, without the OSR. Therefore, this method can only improve the throughput with respect to the basic IR protocol.

The throughput of the concatenated OSR-IR protocol can be evaluated by using again the Renewal-Reward formula (3), by appropriately defining the random reward $\mathcal{R}$ and the inter-renewal time τ. Let $\mathcal{E} = \{\text{The user stops transmitting the current codeword }\}$ be again the recurrent event, and $\hat{q}(m)$ be the probability that the BP algorithm ends with a number of erroneous subpackets $e_m \leq \delta$. Defining $\mathcal{B}_s = \{e_s < \delta\}$ for $s = 1, \ldots, M$, we have

$$\hat{q}(m) = \Pr\left(\overline{\mathcal{B}}_1, \ldots, \overline{\mathcal{B}}_{m-1}, \mathcal{B}_m\right) \tag{23}$$

The recurrent event probability is given by

$$\begin{cases} \Pr(\mathcal{E}_m) = \hat{q}(m) & \text{if } m \leq M-1, \\ \Pr(\mathcal{E}_M) = 1 - \sum_{m=1}^{M-1} \hat{q}(m) & \text{if } m = M. \end{cases} \tag{24}$$

Defining $\hat{p}(m) = \Pr\left(\overline{\mathcal{B}}_1, \ldots, \overline{\mathcal{B}}_{m-1}, \overline{\mathcal{B}}_m\right)$, we have $\hat{q}(m) = \hat{p}(m-1) - \hat{p}(m)$ and substituting this in (24) we get $\Pr(\mathcal{E}_M) = \hat{p}(M-1)$.

The average inter-renewal time (in slots) is given by:

$$\mathbb{E}[\tau] = \sum_{m=1}^{M} m \cdot \Pr(\mathcal{E}_m) = \sum_{m=1}^{M-1} m\hat{q}(m) + M\hat{p}(M-1) = 1 + \sum_{m=1}^{M-1} \hat{p}(m) \tag{25}$$

The reward $\mathcal{R}$ is a random variable that takes values in the range $\{0, P/L, \ldots, n_p P/L\}$. Recalling

the definition of e_m as the number of erroneous packets after decoding at IR step m, we can write

$$\begin{aligned}\mathbb{E}[\mathcal{R}] &= \frac{P}{L}\sum_{m=1}^{M}\sum_{e=0}^{n_p}(n_p-e)\Pr(e_m=e\,|\,\mathcal{E}_m)\Pr(\mathcal{E}_m) \\ &= \frac{Pn_p}{L}\left(1-\sum_{m=1}^{M-1} r_m\hat{q}(m)-r_M\hat{P}(M-1)\right) \end{aligned} \tag{26}$$

where we define

$$r_m=\frac{1}{n_p}\sum_{e=0}^{n_p} e\Pr(e_m=e|\mathcal{E}_m)$$

to be the average fraction of subpackets in error after decoding at step m, given the recurrent event. Recalling that $Pn_p/L = RM$, we obtain the desired throughput expression as

$$\eta = RM\frac{1-\sum_{m=1}^{M-1} r_m\hat{q}(m)-r_M\hat{p}(M-1)}{1+\sum_{m=1}^{M-1}\hat{p}(m)} \tag{27}$$

The above formula can be evaluated after computing by Monte Carlo simulation the probabilities $\hat{p}(m)$ and the quantities r_m.

5.1 Results

In this section we show the throughput resulting from the modified LDPC ensemble, from the use of an OSR protocol, or from a combination of both techniques. In all the following examples, we fixed the subpacket length of the OSR protocol equal to $P = 48$ bits (6 bytes).

Clearly, the throughput achieved by OSR depends on the threshold δ. Analytical optimization of δ is difficult if not impossible. Hence, we exhaustively searched for the best threshold value. Figure 9 shows the throughput as a function of $\delta \in [0,1]$ for the same setting as in figure 7 and $\gamma = 10$dB. We notice that the performance of the OSR is quite insensitive to the value of δ (unless δ is either very close to 0 or very close to 1). We plotted also the throughput achieved by the same ensemble with infinite length, with finite length without any countermeasure and with finite length by averaging over the modified ensemble. These results are shown as horizontal lines as they do not depend on δ.

Both the OSR and the modified ensemble are able to recover a large fraction of the loss incurred by finite length LDPCs. It is natural to wonder about the benefit of using jointly the OSR protocol and a modified LDPC ensemble. Unfortunately, the answer to this question is negative. In figure 9, the curve labelled by "OSR-Modified Ensemble" refers to this case and we notice that the obtained throughput is slightly inferior to that obtained by using OSR with the total ensemble. This fact can be explained by noticing that for a typical code in the modified ensemble a frame-error corresponds to a large number of bit errors (i.e., a large number of subpackets to retransmit). Hence, using an outer SR protocol does not improve the throughput.

The almost constant behavior of throughput of OSR over a wide range of values of the threshold δ is explained by observing the statistics of the number of subpackets in error e_m after decoding. For example, figure 10 shows the probability mass function of e_m conditioned on the event that the decoder works above its iterative threshold decoding (i.e., subject to the event that *DE* with m received blocks converges to vanishing BER), with $m = 4$ received blocks. We notice that the number of packets in error is mostly concentrated below 10% and above 90%. This behavior can be observed for all m. Therefore, the throughput is almost constant for $\delta \in (0.1, 0.9)$.

6 Conclusions

This paper extends previous analysis of Hybrid ARQ incremental redundancy schemes based on infinite-length Gaussian random codes of [6] to infinite-length binary codes (Random Binary and LDPC codes) and to practical finite block length LDPC codes. We showed that, under the assumption of very large (infinite) block length and that vanishing BER implies vanishing FER, irregular LDPC ensembles with degree distribution optimized for the standard AWGN channel [8] provide performance very close to the information-theoretic limit of random binary codes.

Practical finite-length LDPC codes under no optimistic assumption incur a considerable performance loss. Therefore, we proposed two methods to overcome this problem and to make practical LDPC codes effective for the IR protocol: the first method consists of constructing the LDPC

code with a special arrangement of the edges of left-degree 2, in order to improve the FER performance. The second method is based on the concatenation of an outer selective-repeat loop acting on smaller information packet units. Both methods are able to recover a significant fraction of the throughput loss due to finite-length and yield approximately equivalent performance. Hence, they can be regarded as two valuable alternatives for the system designer.

APPENDIX

A Gaussian Approximation of Density Evolution

Consider the LDPC ensemble defined by the left and right degree distributions $\lambda(x)$ and $\rho(x)$. Denote by a the messages sent from bitnodes (v, see figure 2) to checknodes (c), and by u the messages sent from checknodes to bitnodes. Let $\mathcal{L}$ denote the channel observation message, in the form of the log-likelihood ratio for the symbol associated to the given bitnode, given the channel output. Assuming, without loss of generality, that the all-zero codeword is transmitted, if the symbol corresponding to the bitnode is transmitted on the s-th slot,[4] then $\mathcal{L} \sim \mathcal{N}(4\beta_s, 8\beta_s)$.

In order to simplify the DE, we use the following Gaussian Approximation: we assume that all messages generated by the BP decoder at any iteration are Gaussian distributed, and we enforce the *symmetry condition* [8, 16] that must be satisfied by the true distribution of messages generated by BP. The symmetry condition applied to a Gaussian distribution implies that, at each iteration, the variance of the messages is equal to twice the conditional mean. Therefore, tracking the evolution of the message distribution along the BP iterations is equivalent to tracking the evolution of a single parameter (e.g., the message mean). Following [19, 20], it is convenient to choose as state variable of the resulting one-dimensional dynamical system approximating DE the mutual information between a message and the associated bitnode variable.

[4] It is easy to see that the initial message is equal to $\mathcal{L} = 4\gamma \mathrm{Re}\{yc_s^*\}$, where y is the corresponding channel output and c_s is the fading coefficient.

We define a random variable P that governs the distribution of the variable node belonging to the s-th block, so that P is uniformly distributed over $s = 1, \ldots, M$. Let X denote the bitnode variable and Y denote all the information available at the bitnode at a given iteration. Then, the mutual information between the output of the bitnode and the symbol X is given by

$$I(X,Y\,|\,P) = \sum_{s=1}^{M} \frac{1}{M} I(X,Y\,|\,P=s) \tag{28}$$

From the Gaussian Approximation, it follows that

$$I(X;Y|P=s) = J((d-1)\mu+\beta_s)$$

for a bitnode of degree d transmitted on slot s, where μ denotes the mean divided by 4 of the messages u coming from the checknodes. Hence, the mutual information of a message passed along a random edge from a bitnode to a checknode at iteration l is given by

$$I^{l}_{out,v} = \frac{1}{M} \sum_{s=1}^{M} F_{\lambda}\left(I^{l-1}_{out,c}, \beta_s\right) \tag{29}$$

where, for a general distribution $g(x) = \sum_{i\geq 2} g_i x^{i-1}$ and $b \geq 0$ we define the function

$$F_g(z,b) \stackrel{\Delta}{=} \sum_{i\geq 2} g_i J\left((i-1)J^{-1}(z)+b\right) \tag{30}$$

and where $I^{l-1}_{out,c}$ is the mutual information of messages passed along a random edge from a checknode to a bitnode at iteration $l-1$.

In order to find the mutual information transfer function for the checknodes, we use the so-called "approximate duality" relation [21]. With this approximation, a checknode can be replaced by a bitnode provided that its input mutual information I_{in} is transformed into $1-I_{in}$ and its output mutual information I_{out} is transformed into $1-I_{out}$ (see [20, 22] for a more rigorous motivation of this approximation). Hence, the mutual information transfer of a checknode of degree d is approximated by

$$I^{l}_{out,c} = 1 - J\left((d-1)\, J^{-1}\left(1-I^{l}_{out,v}\right)\right) \tag{31}$$

Therefore, the mutual information of a message passed along a random edge from a checknode to a bitnode at iteration l is given by

$$I^l_{out,c} = 1 - F_\rho\left(1 - I^l_{out,v}, 0\right) \tag{32}$$

By combining equations (29) and (32), we obtain the one-dimensional recursion

$$I^l_{out,v} = \frac{1}{M}\sum_{s=1}^{M} F_\lambda\left(1 - F_\rho\left(1 - I^{l-1}_{out,v}, 0\right), \beta_s\right) \tag{33}$$

with initial condition $I^0_{out,v} = 0$, which is the same as (18). The DE-GA recursion for a given number of received blocks m with fading gains $\alpha_1, \ldots, \alpha_m$ is obtained by letting

$$\beta_s = \begin{cases} \gamma\alpha_s & \text{for } s = 1, \ldots, m, \\ 0 & \text{for } s = m+1, \ldots, M. \end{cases}$$

in (33).

It can be shown that the sequence $\{I^l_{out,v}; l = 1, 2, \ldots\}$ is non-decreasing. Hence, the trajectory of (33) converges to the smallest fixed point in the interval $[0, 1]$. We say that the (approximation of the) DE converges to vanishing BER if the mutual information converges to 1, i.e., (33) has a unique fixed point at 1. Since the function $\frac{1}{M}\sum_{s=1}^{M} F_\lambda\left(1 - F_\rho(1-z, 0), \beta_s\right)$ is non-decreasing with $z \in [0, 1]$ and positive for $z = 0$. The fixed point at 1 is unique if and only if (20) holds.

References

[1] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 2001.

[2] L. H. Ozarow, S. Shamai, and A.D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. on Vehicular Technology*, vol. 43, no. 2, pp. 359–378, May 1994.

[3] M. Mouly and M.-B. Pautet, *The GSM System for Mobile Communications*, Cell&Sys, 1992.

[4] H. Holma and A. Toskala, *WCDMA for UMTS, 2nd Edition*, John Wiley and Sons, 2002.

[5] D. J. Costello and S. Lin, *Error and Control Coding: Fundamentals and Applications*, Prentice Hall, 1983.

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

[7] C.-F. Leanderson and G. Caire, "The performance of incremental redundancy schemes based on convolutional codes in the block-fading Gaussian collision channel," *submitted to IEEE Trans. on Wireless Communication*, Dec. 2001.

[8] T. J. Richardson and R. L Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[9] T. J. Richardson, M. A Shokrollahi, and R. L Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[10] W. Feller, *An Introduction of Probability Theory and Its Applications*, New York: Wiley, 1968.

[11] M. Zorzi and R. R. Rao, "On the use of renewal theory in the analysis of ARQ protocols," *IEEE Trans. on Communications*, vol. 44, no. 9, pp. 1077–1081, Sept. 1996.

[12] T. Cover and J. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[13] R. G. Gallager, *Low-Density Parity-Check Codes*, Ph.D. thesis, Cambridge, MA: MIT Press, 1963.

[14] R. J. McEliece, D. J. C MacKay, and J. F. Cheng, "Turbo decoding as an instance of Pearl' s belief propagation algorithm," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 2, pp. 140–150, Feb. 1998.

[15] D. Burshtein and G. Miller, "Expander graph arguments for message-passing algorithms," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 782–790, Feb. 2001.

[16] S. Y. Chung, T. J. Richardson, and R. L Urbanke, "Analysis of sum-product decoding of design of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. on Information Theory*, vol. 47, no. 2, pp. 657–670, Feb. 2001.

[17] C. D. Matthew, *Error-Correction using Low-Density Parity-Check Codes*, Ph.D. thesis, Gonville and Caius College, Cambridge, 1999.

[18] R.L. Urbanke et al., "http://lthcwww.epfl.ch/research/ldpcopt/," .

[19] S. T. Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. on Communications*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.

[20] A. Roumy, S. Guemghar, G. Caire, and S. Verdu, "Design methods for irregular repeat-accumulate codes," *submitted to IEEE Trans. on Information Theory*, Oct. 2002.

[21] S. Y. Chung, *On the Construction of Some Capacity-Approaching Coding Scheme*, Ph.D. thesis, Massachussetts Institute of Technology, Sept. 2000.

[22] A. Ashikhmin, G. Kramer, and S. T. Brink, "Extrinsic information transfer functions: A model and two properties," *in Proc. IEEE Int. Symposium Information Theory (ISIT 2002), Louzanne, Switzerland*, p. 115, July 2002.

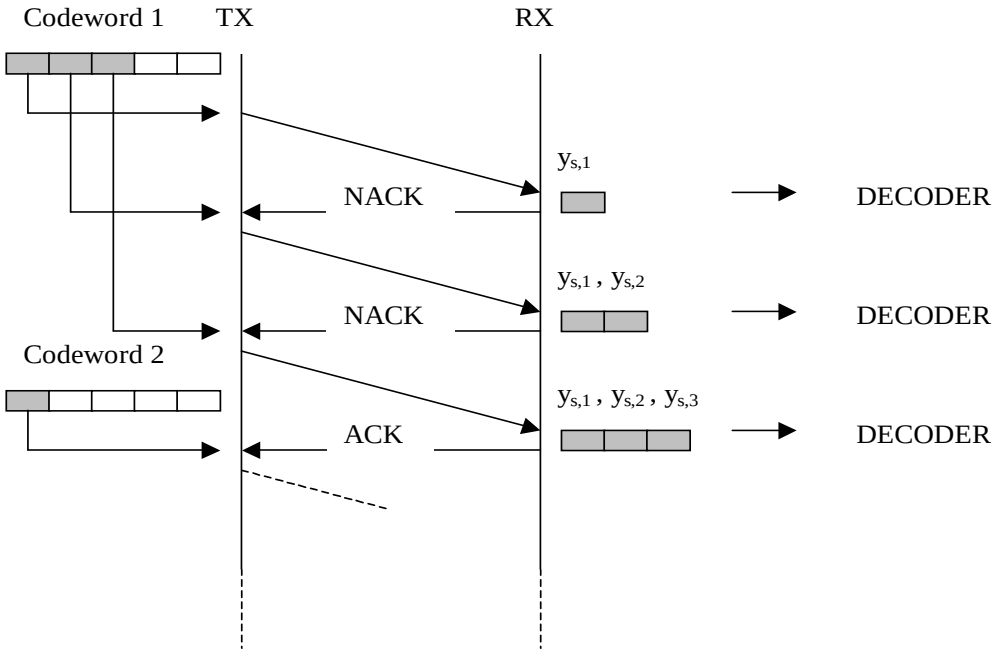

Figure 1: H-ARQ Incremental Redundancy protocol.

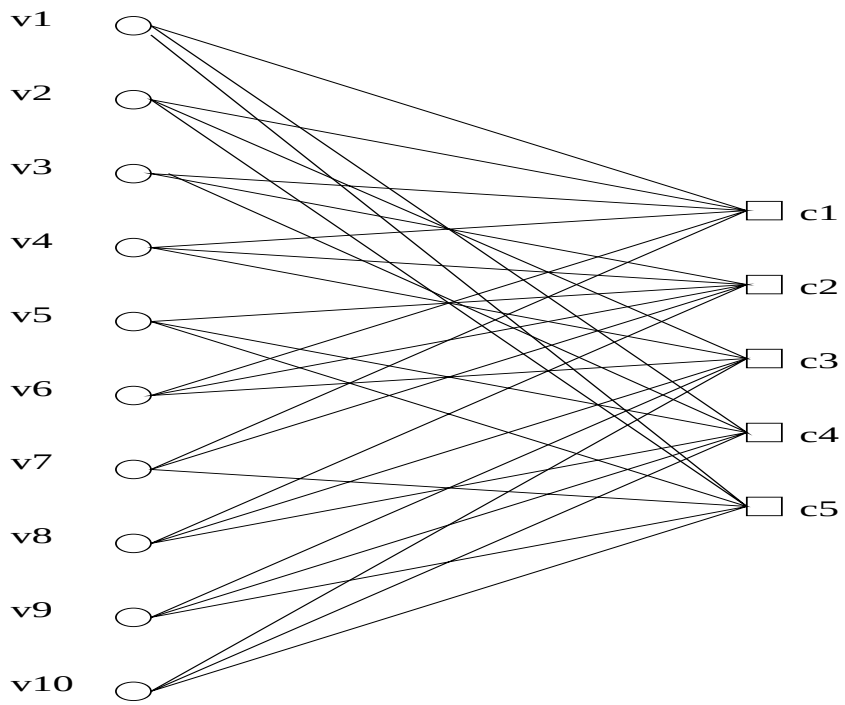

Figure 2: The bipartite graph representing the parity-check matrix of a $(3,6)$-regular LDPC code of length 10.

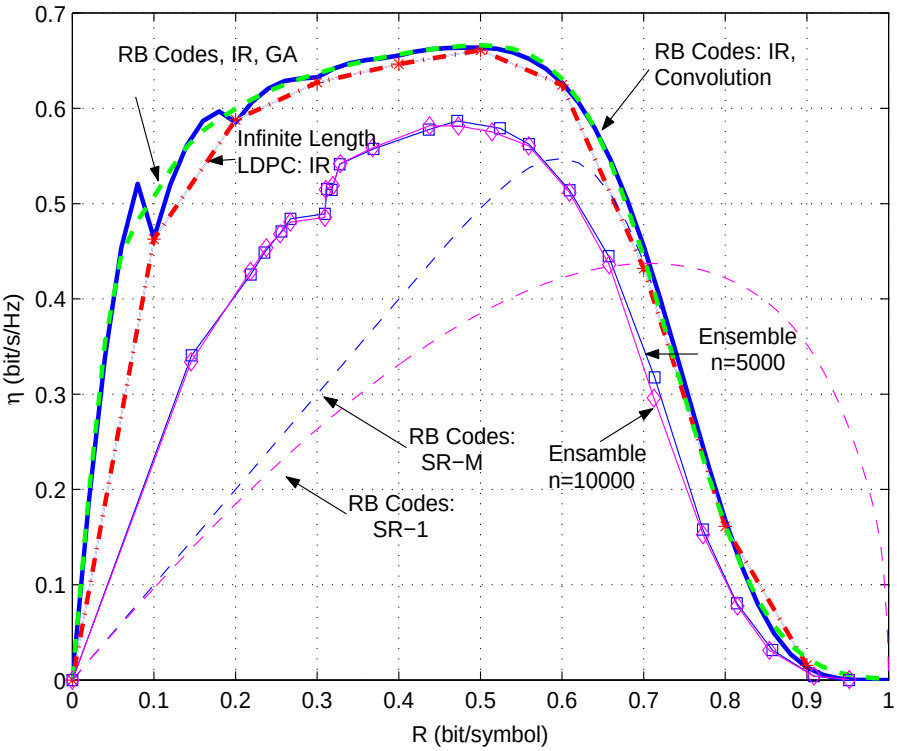

Figure 3: Throughput vs. code rate R for $\gamma = 3$dB. Incremental Redundancy (IR) protocol with Random Binary (RB) codes (the result of the Gaussian Approximation (GA) (14) is shown for comparison), and with infinite length LDPC codes. Selective Repeat protocols (SR-1 and SR-M) with random binary codes are also shown.

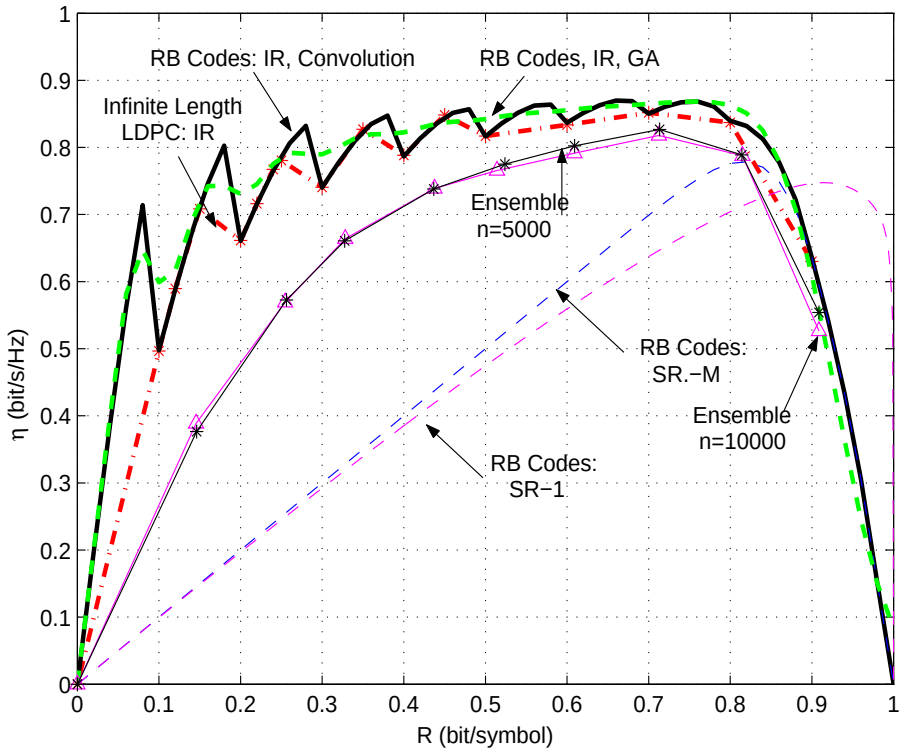

Figure 4: Throughput vs. code rate R for $\gamma = 10$dB. Incremental Redundancy (IR) protocol with Random Binary (RB) codes (the result of the Gaussian Approximation (GA) (14) is shown for comparison), and with infinite length LDPC codes with degree distributions taken from [18]. Selective Repeat protocols (SR-1 and SR-M) with random binary codes are also shown.

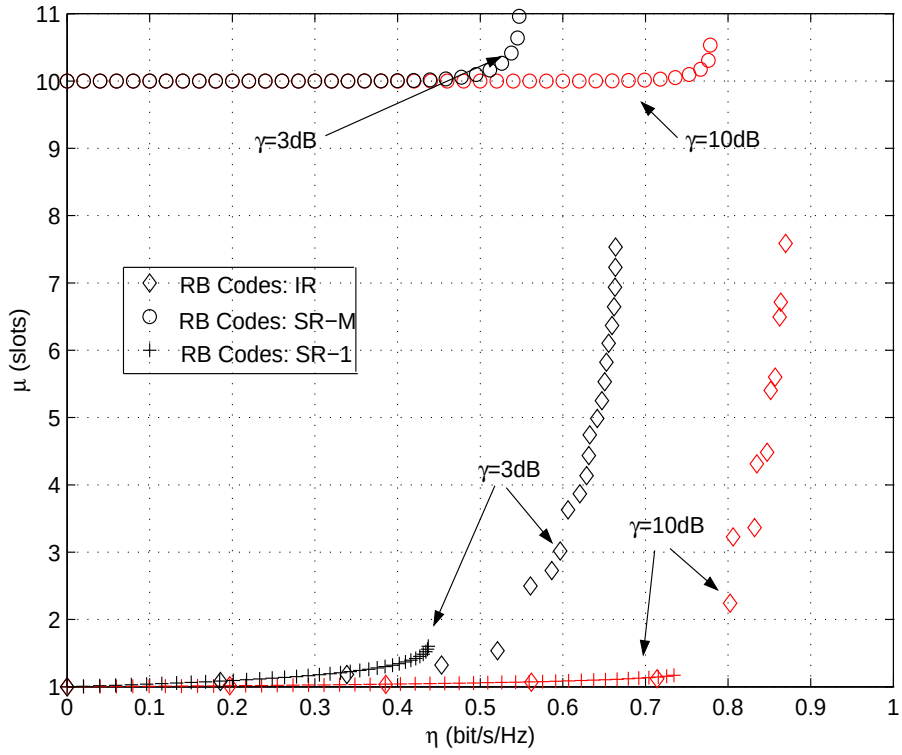

Figure 5: Average delay μ vs. throughput η for IR, SR-1 and SR-*M* protocols with random binary codes for $\gamma = 3$ and 10 dB.

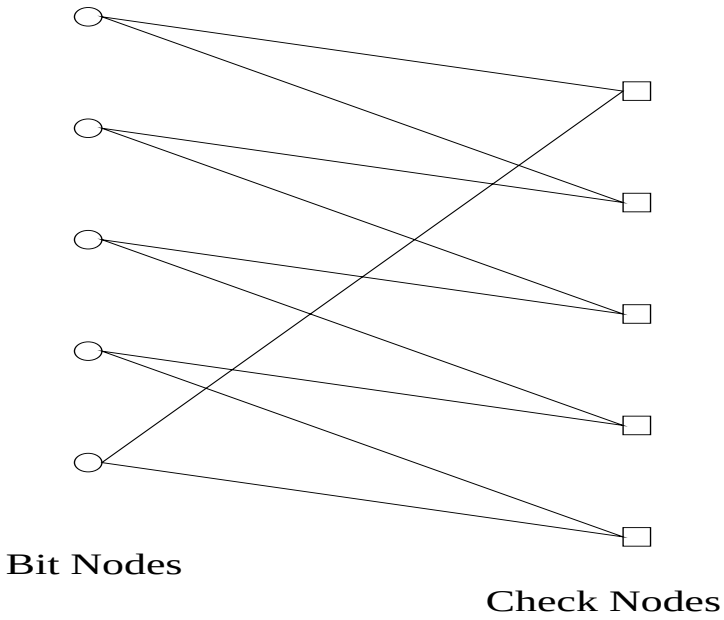

Figure 6: Cyclic arrangement of the edges adjacent to bitnodes of degree 2.

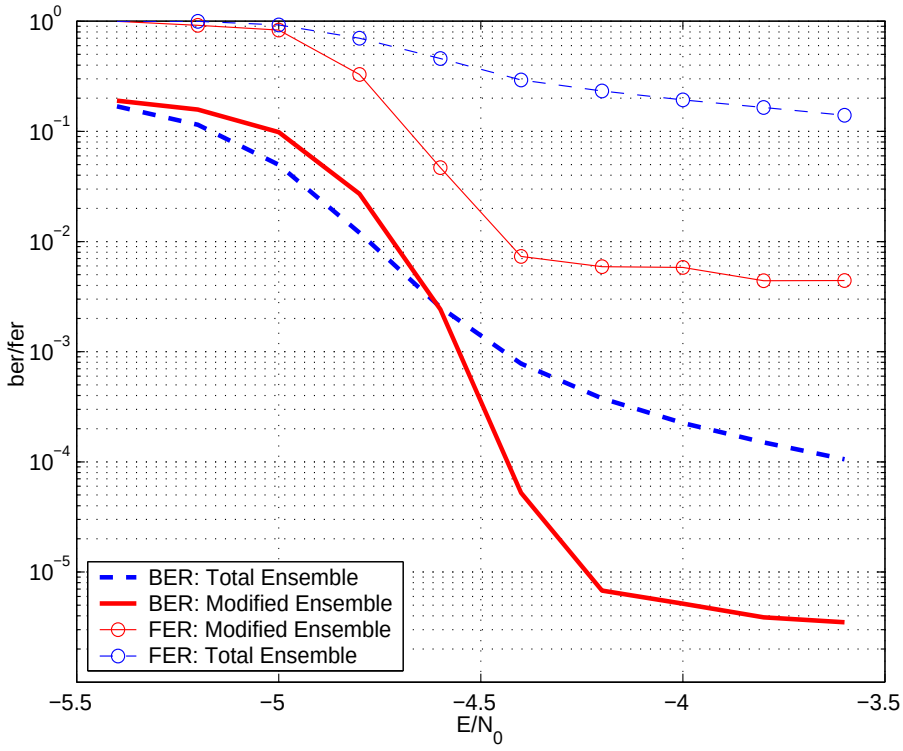

Figure 7: BER and FER of the LDPC ensemble with degree distributions given in [18] for a rate $R = 0.3$ bit/symbol, maximum left degree $d_v = 100$, average right degree $a_r = 6.9$ and length $n = 10000$, over the AWGN channel. The curves labeled as "total ensemble" are obtained by averaging over all code graphs with the given degree distributions. The curves labeled by "modified ensemble" are obtained by averaging over the graphs with degree-2 edges arranged in a cycle, as shown in figure 6.

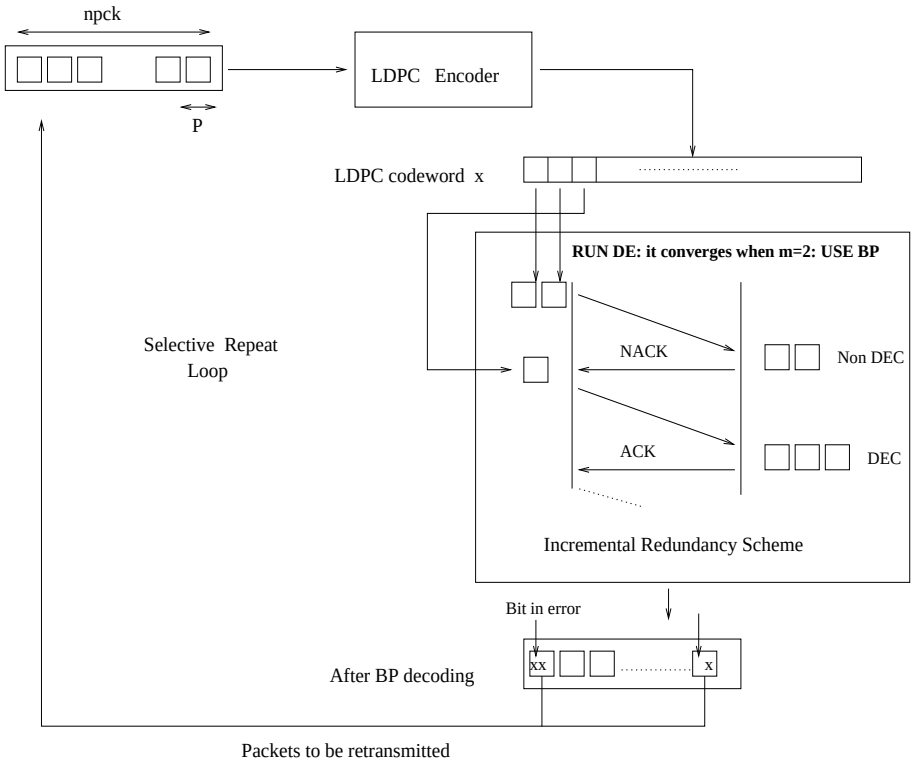

Figure 8: Outer Selective Repeat scheme.

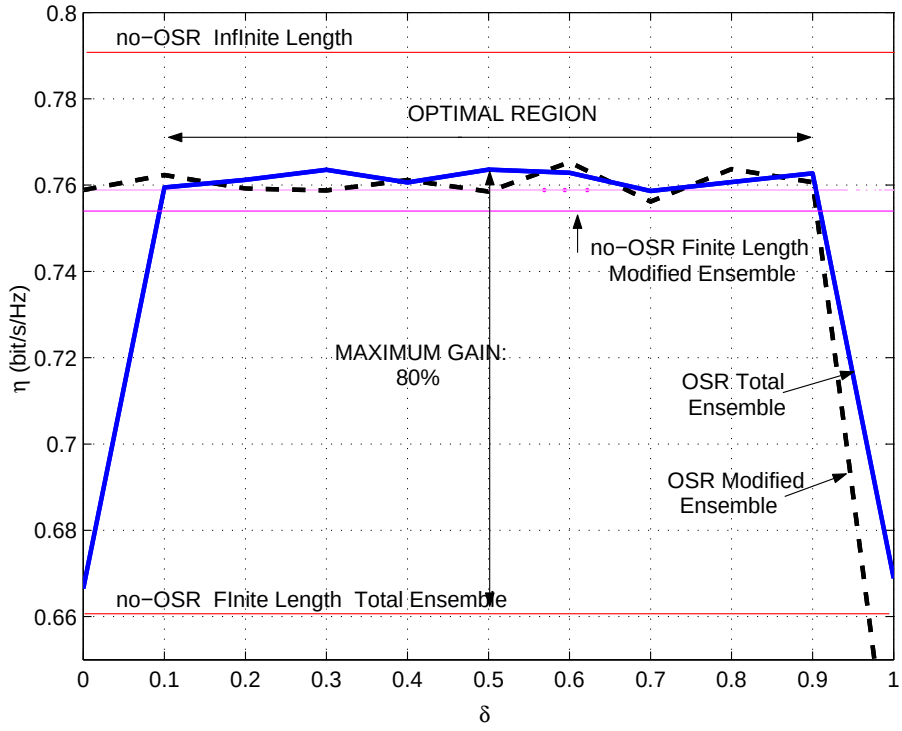

Figure 9: Throughput as a function of the threshold δ for $\gamma = 10$dB and $R = 0.3$bit/symbol for the LDPC codes with length $n = 10000$ with OSR. The throughput without OSR (labeled "no-OSR") for finite and infinite length are shown for comparison as horizontal lines (in these cases the throughput is independent of δ).

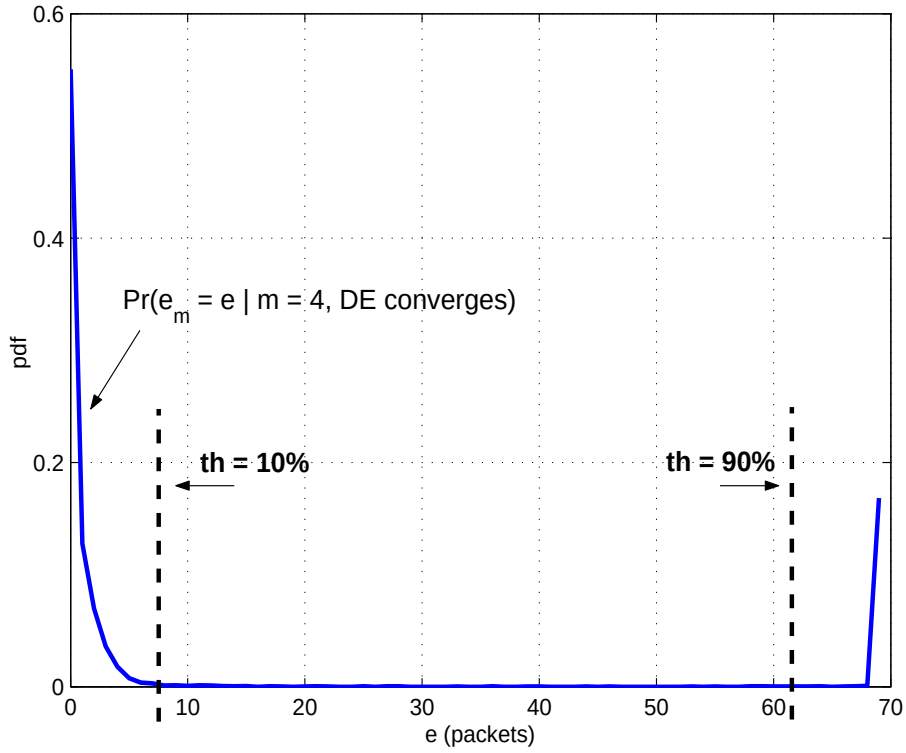

Figure 10: Probability mass function $\Pr(e_m = e \,|\, DE_m \textit{ converges})$ for $m = 4$, $R = 0.3$bit/symbol, $\gamma = 10$dB and $n = 10000$.