Semi-analytical evaluation of concatenated RS/LDPC coding performance with finite block interleaving

Adriaan Suls*, Julie Neckebroek*, Yannick Lefevre°, Mamoun Guenach°*, Marc Moeneclaey*
* Dept. of Telecommunications and Information Processing, Ghent University, Belgium
°Bell Labs, Nokia, Antwerp, Belgium

***Abstract*—The Monte Carlo (MC) simulation of the error performance of a concatenated coding system with finite interleaving depth between the outer and inner codes is time-consuming, especially when targeting low error rates and examining several interleaver settings. In this contribution we present a semi-analytical evaluation of the word error rate (WER) performance of a system with Reed-Solomon (RS) outer coding, finite block interleaving and systematic low-density parity check (LDPC) inner coding. The proposed evaluation method relies on a simple semi-analytical statistical model for the number of bit errors in a segment of the information word after LDPC decoding on the AWGN channel. Only the WER and bit error rate (BER) of the inner subsystem (LDPC code and the considered constellation) are required to compute the WER of the concatenated code, corresponding to different parameters of the RS code and the interleaver. We show that the semi-analytical WER of the concatenated system closely matches the WER resulting from MC simulations, for both the AWGN channel and the Rayleigh block-fading channel.**

I. Introduction

Concatenated coding, involving an outer code and an inner code which are separated by an interleaver, is a common error correction scheme, used to improve the error performance of digital communication systems. The role of the interleaver is to distribute burst errors, occuring at the output of the inner decoder, over multiple codewords of the outer code, hence increasing the chance that the outer decoder can correct the residual errors from the inner decoder. Contemporary technologies using concatenated coding include the G.993.2 (VDSL2) and the G.fast standards (RS outer code, trellis-coded modulation (TCM) as inner code) [1], [2], and the DOCSIS 3.1 and DVB standards (Bose-Chaudhuri-Hocquenghem (BCH) outer code, LDPC inner code) [3] [4].

Evaluating the error performance of a concantenated coding scheme through MC simulations is very time-consuming, especially at low error rates. Alternatively, the error performance of such a scheme can be obtained from a statistical model of the error patterns at the inner decoder output. For TCM and convolutional inner codes, which use the Viterbi algorithm for decoding, these error patterns can be accurately modeled by means of geometrical distributions [5], [6]. However, this model only applies to trellis-based codes and not to block codes such as LDPC codes, which use iterative belief propagation [7], [8] for decoding.

In this contribution we consider the RS/LDPC concatenated coding scheme with a finite block interleaver, described in section II. Based on MC simulations, we provide in section III a simple statistical model for the number of bit errors in a segment of the information word after LDPC decoding on the AWGN channel, involving the binomial distribution. This model allows to evaluate semi-analytically the WER of the concatenated RS/LDPC coding scheme for various interleaver parameter settings, on both the AWGN and the Rayleigh fading channel. The proposed method is validated in section IV, where the WER resulting from the model is compared to the WER obtained from MC simulations. Conclusions are drawn in section V.

Notation: We denote by $B(i;N,p)$ the probability mass function (pmf) of the binomial distribution, i.e., $B(i;N,p) = C_N^i p^i (1-p)^{N-i}$, for $i \in \{0,1,...,N\}$, with $C_N^i = \frac{N!}{i!(N-i)!}$ the binomial coefficient. The operator $\mathbb{E}[.]$ refers to statistical expectation, and $\lfloor x \rfloor$ is the largest integer not exceeding x.

II. System description

We consider a concatenated coding system, with interleaving between the outer and the inner code. The outer code is a systematic RS($N_{\mathrm{RS}}, K_{\mathrm{RS}}$) code defined over a finite field of size 2^S, so that each symbol of the RS codeword can be associated with S bits; the RS code has a length of N_{RS} symbols, representing K_{RS} information symbols, and can correct any combination of up to $t = \left\lfloor \frac{N_{\mathrm{RS}} - K_{\mathrm{RS}}}{2} \right\rfloor$ symbol errors. The RS codewords are applied to a block interleaver. As indicated in Fig. 1, the block interleaver has D rows, each containing one RS codeword; D is referred to as the interleaver depth. The $N_{\mathrm{RS}}DS$ bits contained in the interleaver constitute the information words for L codewords of the inner code. The inner code is a systematic binary LDPC(N_{I}, K_{I}) code, with codewords of length N_{I} bits, representing $K_{\mathrm{I}} = N_{\mathrm{RS}}DS/L$ information bits. The K_{I} information bits from the lth LDPC codeword are contained in the lth column of the interleaver; hence, each column has a width of $J = N_{\mathrm{RS}}/L = K_{\mathrm{I}}/(SD)$ symbols. The N_{I} bits from the lth LDPC codeword are mapped to data symbols from an M-point signal constellation; the resulting data symbols are stacked into the data symbol vector $\mathbf{x}_l$, having $N_{\mathrm{I}}/\log_2(M)$ components.

The components of the symbol vectors are transmitted sequentially over a memoryless channel. The received signal $\mathbf{y}_l$ associated with the data symbol vector $\mathbf{x}_l$ is given by

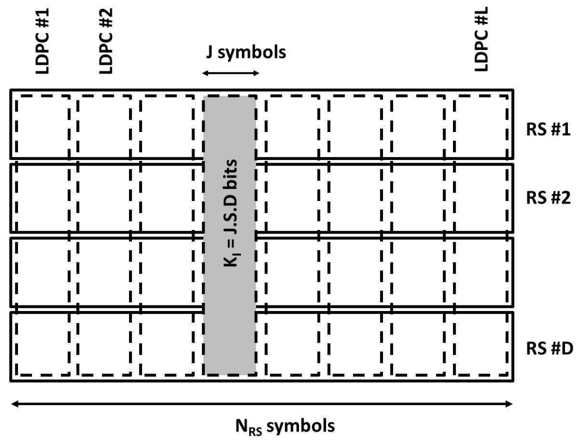

Figure 1. Block interleaving scheme.

$$\mathbf{y}_l = h_l \mathbf{x}_l + \mathbf{w}_l \tag{1}$$

for $l = 1, 2, .., L$, where the components of the noise vector $\mathbf{w}_l$ are independent identically distributed (i.i.d.) complex-valued circular-symmetric (CVCS) Gaussian random variables with zero mean and variance N_0, and h_l is the channel gain experienced by the symbol vector $\mathbf{x}_l$. For an AWGN channel, we have $h_l = 1$ for $l = 1, ..., L$, and the corresponding signal-to-noise ratio (SNR) is defined as $\gamma_{\text{AWGN}} = \mathbb{E}[|\mathbf{x}_l|^2]/\mathbb{E}[|\mathbf{n}_l|^2]$. For Rayleigh block-fading, the channel gains $(h_1, h_2, ..., h_L)$ are i.i.d. CVCS Gaussian random variables with zero mean and variance equal to 1; we denote by $\bar{\gamma} = \mathbb{E}[|\mathbf{x}_l|^2]/\mathbb{E}[|\mathbf{n}_l|^2]$ the average SNR of the fading channel; the instantaneous SNR related to the symbol vector $\mathbf{x}_l$ equals $\gamma_l = |h_l|^2\bar{\gamma}$.

The receiver performs a soft-demapping operation, followed by iterative LDPC decoding based on the sum-product algorithm [7], [8]. The information part of the LDPC decoder output related to the lth LDPC codeword is written into the lth column of a block deinterleaver, which has the same dimensions as the interleaver at the transmitter. An algebraic RS decoder [7], [8], operating on the rows of the deinterleaver, tries to correct the errors at the LDPC decoder output.

III. Error model and performance analysis

A decoding error of the concatenated coding scheme occurs when the RS codeword at the input of the outer decoder is affected by more than t symbol errors. The RS codeword consists of L segments of J symbols, each originating from a different LDPC codeword. A RS decoding error requires at least $L_0 = 1 + \lfloor t/J \rfloor$ segments to contain symbol errors. In the case of Rayleigh block-fading the diversity orders for the systems with and without concatenation therefore equal L_0 and 1, respectively. Hence, to achieve a substantial improvement in error performance with a concatenated scheme, we need $L_0 > 1$, which is obtained by selecting J such that $J \le t$. Based on this observation, we focus on values of JS which are small compared to the number K_I of information bits in the LDPC codeword, as typically we have $St \ll K_\text{I}$.

Let us investigate by means of MC simulation the number of bit errors in segments of the information word, after decoding of the quasi-cyclic binary rate 3/4 LDPC(1120,840) code (DOCSIS standard [3]) with Gray mapping to a 4-QAM constellation on an AWGN channel. Decoding stops when a valid codeword is detected, or after 50 iterations; per SNR value we simulated until 10^4 erroneous codewords occurred. For this specific code, Fig. 2 shows the degree of each variable node versus its position in the codeword. Considering JS-bit segments of the information part of *erroneous* codewords, Fig. 3 (upper) superposes the $D = K_\text{I}/(JS)$ histograms of the number (q_b) of bit errors in such segments, for (JS, D) = (120, 7) and with γ_{AWGN} set such that $\text{WER}_\text{I} = 10^{-2}$. The 7 histograms are not identical, because Fig. 2 indicates that the variable degree distribution *within a segment* depends strongly on which segment has been selected. This dependence gives rise to an unequal WER for the RS codewords in the concatenated system. This can be avoided by applying a *random permutation* to the information bits prior to LDPC encoding (and applying the inverse permutation to the information bits after LDPC decoding). Fig. 3 (lower) shows that the histograms, resulting from applying the same permutation to all transmitted codewords, essentially coincide.

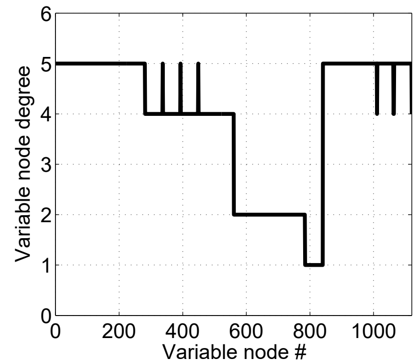

Figure 2. Degree of variable node versus position in LDPC(1120,840) codeword

Denoting by e the event of an LDPC codeword error ($\bar{e}$ indicates correct LDPC decoding) and by q_I the number of information bit errors after LDPC decoding, the random permutation of the information bits gives rise to the following pmf of q_b, the number of bit errors in a segment:

$$\Pr[q_\text{b} = i|e] = \sum_{k=i}^{K_\text{I}-JS+i} H(i; k, JS, K_\text{I}) \Pr[q_\text{I} = k|e] \tag{2}$$

for $i \in \{0, ..., JS\}$. In (2), the function $H(i; k, JS, K_\text{I}) = C^i_{JS} C^{k-i}_{K_\text{I}-JS}/C^k_{K_\text{I}}$ represents the hypergeometrical distribution, which for $K_\text{I} \gg JS$ is well approximated by the binomial distribution $B(i; JS, k/K_\text{I})$. Assuming only a moderate spread of the pmf $\Pr[q_\text{I} = k|e]$ around its mean $\mathbb{E}[q_\text{I}|e]$, so that $H(i; k, JS, K_\text{I})$ (considered as a function of k) is, for any i, much wider than $\Pr[q_\text{I} = k|e]$, we further approximate $\Pr[q_\text{b} = i|e]$ by the following single-parameter model:

$$\Pr[q_\text{b} = i|e] \approx B(i; JS, p_\text{b}) \tag{3}$$

where $p_\text{b} = \mathbb{E}[q_\text{I}|e]/K_\text{I}$. Using $\mathbb{E}[q_\text{I}|\bar{e}] = 0$, and expressing the WER and the BER of the LDPC code as $\text{WER}_\text{I} = \Pr[e] = 1 - \Pr[\bar{e}]$ and $\text{BER}_\text{I} = \mathbb{E}[q_\text{I}]/K_\text{I}$, it follows that

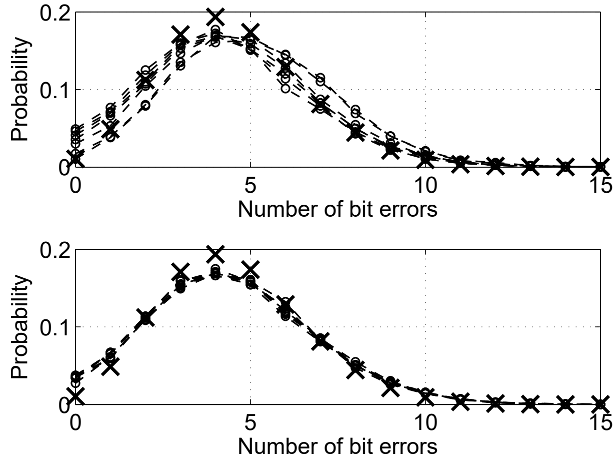

Figure 3. Superposed histograms (small circles, dashed lines) and semi-analytical model (large crosses) for $\Pr[q_\mathrm{b} = i|e]$ with *JS* = 120, showing moderate/good agreement without/with permutation of the information bits (LDPC(1120,840), 4-QAM, AWGN channel, $\mathrm{WER_I} = 10^{-2}$).

$p_\mathrm{b} = \mathrm{BER_I}/\mathrm{WER_I}$, which depends on γ_AWGN. It can be verified from (2) that the model (3) is exact when $\Pr[q_\mathrm{I} = k|e] = B(k; K_\mathrm{I}, p_\mathrm{b})$. According to (3), the bit errors in a segment occur independently with probability p_b in case of a codeword error. Consequently, the unconditional pmf $\Pr[q_\mathrm{b} = i]$ is obtained as

$$\begin{aligned}\Pr[q_\mathrm{b} = i] &= \Pr[q_\mathrm{b} = i|\bar{e}]\Pr[\bar{e}] + \Pr[q_\mathrm{b} = i|e]\Pr[e] \\ &\approx (1 - \mathrm{WER_I})\delta_i + \mathrm{WER_I} B(i; JS, p_\mathrm{b}) \quad (4)\end{aligned}$$

where δ_i is the Kronecker delta function. The model (4) will be instrumental in deriving the WER performance of the concatenated coding scheme for several interleaver settings. Denoting by q_s the number of symbol errors in a J-symbol segment, it follows from the proposed model (4) that

$$\Pr[q_\mathrm{s} = i] \approx (1 - \mathrm{WER_I})\delta_i + \mathrm{WER_I} B(i; J, p_\mathrm{s}) \quad (5)$$

with $p_\mathrm{s} = 1 - (1 - p_\mathrm{b})^S$. As such, p_b and p_s represent the average fraction of erroneous bits and symbols, respectively, in the erroneous codewords.

Fig. 3 also displays the binomial distribution $B(i; SJ, p_\mathrm{b})$, showing a good fit with the histogram corresponding to the permuted information bits. The accuray of the fit of the model (3) with the histograms can be expressed by the Kullback-Leibler divergence (KLD) [9], defined as

$$D_\mathrm{KL}(P|Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)} \quad (6)$$

Here, $Q(i)$ equals $\Pr[q_\mathrm{b} = i|e]$ from the model (3) and $P(i)$ is obtained by averaging the D available histograms corresponding to the JS-bit segments. Table I shows the KLD from (6) for a number of JS and $\mathrm{WER_I}$ values, and two different LDPC codes. The table indicates that for the LDPC(1120,840) and LDPC(5940,5040) codes from [3], the accuracy of the model (3) improves with decreasing JS and increasing $\mathrm{WER_I}$. Denoting by $\sigma_\mathrm{HG}(i)$ the rms spread of $H(i; k, JS, K_\mathrm{I})$ w.r.t. the variable k for given i, we have verified that, as a rule of thumb, the model (3) is very accurate

	LDPC(1120,840)		LDPC(5940,5040)	
γ_AWGN (dB)	4.44	4.77	5.51	5.66
$\mathrm{WER_I}$	1.00E-01	1.00E-02	1.00E-01	1.00E-02
$\mathrm{BER_I}$	3.72E-03	3.81E-04	1.81E-03	2.00E-04
KLD @ JS = 120	2.00E-02	4.16E-02	7.08E-03	9.96E-03
KLD @ JS = 40	1.86E-03	4.44E-03	7.16E-04	9.73E-04

Table I
KLD BETWEEN HISTOGRAM (WITH PERMUTATION OF INFORMATION BITS) AND MODEL INDICATES BETTER ACCURACY FOR SMALLER JS, LARGER K_I AND LARGER $\mathrm{WER_I}$

(i.e., KLD is in the order of 10^{-3} or less), when the standard deviation of $\Pr[q_\mathrm{I} = k|e]$ is less than $(1/2)\min_i \sigma_\mathrm{HG}(i)$, with $\min_i \sigma_\mathrm{HG}(i) \approx D$. We have observed (results not shown for conciseness) that this rule of thumb remains valid for other constellations (e.g., 16-QAM and 64-QAM) and for other LDPC codes (e.g., the (1280, 1024) and (1536, 1024) AR4JA [10] codes, the (1152, 960) and (5184, 4320) G.hn codes [11])). As these G.hn codes are almost regular, their histograms of q_b for the individual segments are essentially the same, even *without* a permutation of the information word.

Now we focus on the error performance of the concatenated coding scheme. Each RS codeword consists of L segments of J symbols, with each segment originating from a different LDPC codeword. Denoting by q_RS the number of symbol errors in a RS codeword at the outer decoder input, we have $q_\mathrm{RS} = \sum_{l=1}^{L} q_\mathrm{s}^{(l)}$, where $q_\mathrm{s}^{(l)}$ is the number of symbol errors in the J-symbol segment from the lth LDPC codeword, which is part of the considered RS codeword. For the AWGN channel, the pmf of q_RS, denoted $\Pr[q_\mathrm{RS} = n]$, equals the L-fold convolution of $\Pr[q_\mathrm{s} = i]$ from (5), because the random variables $\{q_\mathrm{s}^{(l)}, l \in \{1, ..., L\}\}$ are i.i.d. For the Rayleigh block-fading channel with given instantaneous SNRs $\{\gamma_l\}$, the random variables $\{q_\mathrm{s}^{(l)}\}$ are independent, with $\Pr[q_\mathrm{s}^{(l)} = i|\gamma_l]$ given by (5), where WER and p_s correspond to an AWGN channel operating at SNR = γ_l; hence, the pmf of q_RS conditioned on $\{\gamma_l\}$ is obtained as the convolution of $\{\Pr[q_\mathrm{s}^{(l)} = i|\gamma_l], l \in \{1, ..., L\}\}$. As the instantaneous SNRs $\{\gamma_l\}$ are i.i.d., the unconditional pmf of q_RS, denoted $\Pr[q_\mathrm{RS} = n]$, equals the L-fold convolution of the unconditional pmf $\Pr[q_\mathrm{s} = i] = \mathbb{E}_\gamma[\Pr[q_\mathrm{s} = i|\gamma]]$, with $\mathbb{E}_\gamma[.]$ denoting averaging over the instantaneous SNR γ. The resulting WER of the concatenated coding scheme is then obtained as

$$\mathrm{WER_{conc}} = \sum_{n=t+1}^{N_\mathrm{RS}} \Pr[q_\mathrm{RS} = n] \quad (7)$$

with the derivation of $\Pr[q_\mathrm{RS} = n]$ depending on the type of channel (AWGN or Rayleigh block-fading).

The above analysis shows that it is sufficient to know the WER and the BER of the inner LDPC code on the AWGN channel as a function of γ_AWGN for the considered constellation (e.g., in the form of a lookup table), to characterize the semi-analytical model (4) for the pmf of q_b, from which we can compute the WER of the concatenated coding system (for the AWGN channel *and* the Rayleigh block-fading channel) for various values of the interleaver depth D and the RS code parameters; this represents considerable computational savings

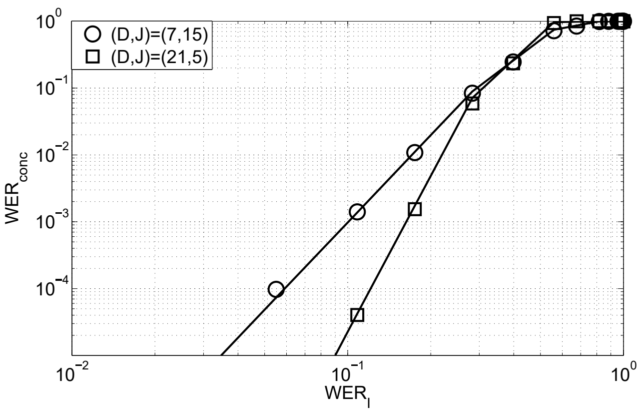

Figure 4. Semi-analytical WER (solid lines) agrees very well with WER from MC simulation (markers) for the concatenated system on AWGN channel with different interleaver depths (RS(210,188) plus LDPC(5940,5040), 4-QAM).

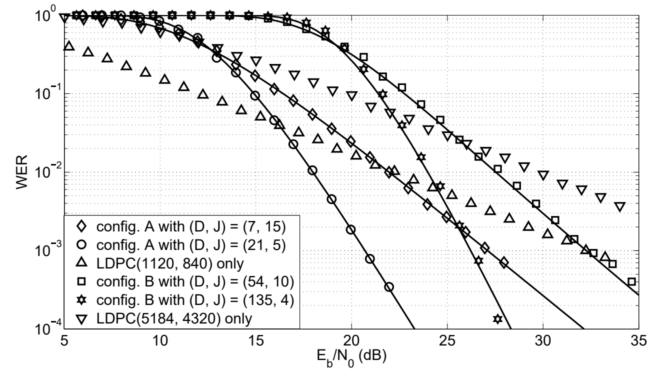

Figure 5. Semi-analytical WER (solid lines) agrees very well with WER from MC simulation (markers) for configurations A and B on Rayleigh block-fading channel with different interleaver depths.

compared to straightforward MC simulations.

IV. Numerical error performance results

We consider a concatenated scheme where the RS outer code is obtained by shortening the RS(255,223) code with symbol size S = 8; the resulting RS outer code can correct up to $t = 16$ symbol errors.

For the AWGN channel, Fig. 4 shows the WER of the concatenated system versus the WER of the inner code, for the RS(210,188) outer code and the LDPC(5940,5040) inner code with 4-QAM mapping, with (D, J) = (7, 15), (21, 5). A very good fit between the simulations and the semi-analytical result is observed, even for settings yielding a KLD (6) up to about 10^{-2}. Note from Fig. 4 that a WER_I in the range $(10^{-2}, 10^{-1})$ is a convenient operating point for the inner code in the concatenated system, as the corresponding $\mathrm{WER}_\mathrm{conc}$ can be made significantly smaller by selecting a proper value of J.

Fig. 5 shows the WER of the concatenated system versus E_b/N_0 (with E_b the energy per information bit) in the case of Rayleigh block-fading for two configurations: in configuration A we use the RS(210,288) outer code and LDPC(1120,840) inner code from [3], with 4-QAM mapping and (D, J) = (7,15) and (21,5), yielding diversity orders L_0 = 2 and 4, respectively; in configuration B we use the RS(240,208) outer code and LDPC(5184,4320) inner code from [11], with 16-QAM mapping and (D, J) = (54,10) and (135,4) yieding L_0 = 2 and 5, respectively. The WER of the LDPC code (without concatenation) shows a diversity order of only 1. The MC simulations and the semi-analytical result agree very well, for both configurations A and B. We also obtained similar results for other LDPC codes (not shown here for conciseness)

V. Conclusions and remarks

In this contribution we have modeled the pmf of the number of bit errors in a segment of the information word after LDPC decoding at a given instantaneous SNR as a mixture of a Kronecker delta function and a binomial distribution; an excellent fit is obtained when the standard deviation of the total number of information bit errors in an erroneous codeword does not exceed half the interleaving depth D. This model allows the fast evaluation of the WER performance of a concatenated system, consisting of outer RS coding, finite block interleaving, inner LDPC coding and mapping to a constellation. For this WER evaluation we only need the WER and the BER performance of the inner subsystem, determined by the LDPC code and the considered constellation. The accuracy of the WER evaluation of the concatenated coding system has been validated by means of MC simulations, for both the AWGN and the Rayleigh block-fading channel.

The proposed method can be straightforwardly generalized to other families of outer codes with known error correcting capabilities (such as BCH codes).

Acknowledgment

This research has been funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office, and by the VLAIO HBC.2016.0055 “Fifth generation broadband access (5GBB)” project.

References

[1] ITU-T recommendation G.993.2: Very high speed digital subscriber line transceivers 2 (VDSL2).

[2] ITU-T recommendation G.9701: Fast access to subscriber terminals (FAST) - physical layer specification.

[3] Data-over-cable service interface specifications: DOCSIS 3.1: Physical layer specification (CM-SP-PHYv3.1-I07-150910).

[4] ETSI TS 102 991 V1.3.1 digital video broadcasting (DVB); implementation guidelines for a second generation digital cable transmission system (DVB-C2).

[5] R.L. Miller, L.J. Deutsch and S.A. Butman. On the error statistics of viterbi decoding and the performance of concatenated codes. *JPL Publications*, 81-9, 1981.

[6] T. Berman, J. Freedman and T. Kaplan. An analytical analysis of a concatenated Reed-Solomon (RS)/Viterbi coding system both with and without RS interleaving. In *Computers and Communications, 1992. Conference Proceedings.*, pages 260–266. IEEE, 1992.

[7] T. Richardson and R. Urbanke. *Modern Coding Theory.* Cambridge University Press, 2008.

[8] W. Ryan and S. Lin. *Channel Codes: Classical and Modern*. Cambridge University Press, 2009.

[9] T.M. Cover and J.A. Thomas. *Elements of Information Theory (2nd ed.).* Wiley, 2006.

[10] CCSDS 131.1-O-2 low density parity check codes for use in near-earth and deepspace applications, CCDS.org, Orange Book, Sept. 2007.

[11] ITU-T recommendation G.9960: Unified high-speed wireline-based home networking transceivers - system architecture and physical layer specification.