

GTTI 2008: sessione Trasmissione Numerica
**Non-binary LDPC codes with good performance on
channels affected by bursty noise**

Andrea Marinoni and Pietro Savazzi
Università degli Studi di Pavia
Dipartimento di Elettronica
Via Ferrata 1, Pavia, Italy
Email: {pietro.savazzi, andrea.marinoni}@unipv.it

Stefano Valle
STMicroelectronics
CPG - Data Storage Division
Via Tolomeo 1, Cornaredo (MI), Italy
Email: stefano.valle@st.com

Abstract

Non-binary Reed-Solomon (RS) codes represent a typical way to counteract the effects of error bursts in hard-disk drives (HDD). Recently, it has been demonstrated how q -ary low-density parity-check (LDPC) codes can outperform RS codes through a proper construction of the parity check (PC) matrix. In this work, two matrix design schemes, based on the main definitions of erasure burst proposed in literature, are presented. The performance of generated codes are evaluated on bursty noise channels, in order to highlight the main constraint that has to be taken into account for designing LDPC codes with a good burst error correction capability. The proposed schemes have been evaluated on both AWGN and magnetic recording channels in presence of erasure bursts, comparing their coding gain against those obtained by using non-binary RS codes.

1 Introduction

Recent research findings propose LDPC as a possible alternative to RS coding in systems such as magnetic recording channels for hard-disk drives (HDD) [1], [3], [4]. Near capacity performance and the ability to correct large error bursts make these codes a good alternative to powerful algebraic decoding of non-binary RS codes with large symbols. In this sense, parallel research works

go toward the direction of improving RS performance introducing the soft decoding concept of LDPC (see [2] and references therein). However, the RS parity check matrix is not suitable for the conventional Message Passing (MP) algorithm; in fact, RS soft decoding is still an open problem, especially when complexity is considered.

Non-binary, or q -ary, LDPC codes may be used in HDD in order to increase the robustness against error bursts, caused by media defects and thermal asperities [3]. The performance of these codes outperforms RS codes, while the major drawback is the complexity which increases with Ntq^2 , where q is the number of symbols defined in the finite field $GF(q)$, N is the coded block length and t is the average weight of parity-check (PC) matrix columns, i.e. the number of non-zero elements per column [6]. In this paper decoder complexity is neglected. However, recent works indicate that it can be reduced either using the Fast Fourier Transform (FFT) [3], [7] or simplifying the decoding rules for high rate codes, considering the dual codes [8].

In [1] the capability to correct an erasure burst, when no errors are expected in the guard band (i.e. all the symbols outside the burst are perfectly known at the receiver), is measured by means of a compact algorithm which provides the "maximum resolvable erasure-burst length", MREBL, defined as "the longest string of erasures that the codes decoder is guaranteed to fill

in with the correct bit values (or symbol values ...), no matter where the burst starts in". The algorithm inspects the matrix H , looking for the error locations in a burst which can be resolved by checks at each iteration. Here, MREBL will be considered in conjunction with the number of iterations the algorithm requires to find MREBL. The capability to correct erasure bursts has been related to the parity check matrix by means of the Minimum Space Distance (MSD), that is the minimum distance between non-zero entries of parity check matrix rows [3], [4]. MSD represents also a lower bound for MREBL. In [5] the authors use a Progressive Edge Growth (PEG) algorithm and look for a good permutation of the variable nodes in order to improve the burst-erasure-correction capability of LDPC codes. In [3], the authors include the MSD maximization in a PEG construction algorithm for the PC matrix. A similar approach has been reported in [9] for binary LDPC. Anyway, there is no difference between the binary and non-binary case as long as the usual way to build a q -ary PC matrix is to multiply the non-zero entries of a binary mother matrix H by random elements taken in $GF(q)$ [3], [4], [6], [7]. It's obvious that the MREBL and MSD as measured on the mother parity check are to be multiplied by $\log_2(q)$ in order to obtain the effective burst length in bits. A similar burst-error-correction capability estimation is reported in [10], [11] for structured LDPC Gilbert codes which are very similar to the so-called array codes [12]. In reality, channels are often affected by random noise and erasures. In this sense, in all the cited works performance has been evaluated by adding error bursts in an additive, white and Gaussian noise (AWGN) channel or in magnetic recording channel.

The main results presented in this work are:

1. the optimization of the PC matrix, monitoring in a PEG algorithm the MREBL and the MSD measures introduced in [1], [3], [4];
2. the comparison among different matrix constructions, optimizing the MSD and the MREBL;
3. validation of the most performable scheme by simulations in burst noise AWGN and magnetic recording channels.

Section 2 is devoted to the description of the modified PEG algorithms for the PC matrix design. The MSD has been inserted in the PEG search in a way similar to [1], while the MREBL has been included progressively in order to reduce

the computation time and maximize the probability of finding a good LDPC matrix without cycles. The subsequent section describes the modeled system with particular emphasis on the perpendicular magnetic channel. Sections 4 and 5 show simulation results on AWGN and perpendicular magnetic recording channel (PMRC) channels, followed by some conclusion and perspective of future works which conclude this paper.

2 Modified PEG algorithms

An empiric procedure to build a Tanner graph maximizing the girth g_0 is suggested in [11], [17], where girth represents the lowest cycle length within the PC matrix: the PEG algorithm is based on a pre-determined weight distribution of symbol and check nodes. The graph construction is based on iterative edge-by-edge steps, maximizing the local girth for given nodes. In the present work, girth ≥ 6 is imposed. The obtained matrix may be either regular or irregular. In the seminal work [7], the author suggests a method, based on Montecarlo simulations, to find the best average column weight of a non-binary LDPC code. Following this reasoning, for code-rate 8/9 $GF(16)$ codes, we observed that an average column weight equaling 2.88 is the best choice. This is the consequence of an empirical optimization based on AWGN channel simulations outcomes obtained with matrices characterized by different average column weights. Figure 1 shows these results as a function of the average column weight t : all the chosen matrices have size (128,1152) and elements living in $GF(16)$, while the average signal to noise ratio per bit (E_b/N_0) value was fixed to 4.2 dB. Ongoing works will be devoted to implementing a more general search algorithm as suggested in [7], [18].

According to this result, the matrices in the present work share an average column weight 2.88; the rate is 8/9 and size is (128,1152) $GF(16)$ symbols. In the following we describe two modified PEG algorithms which include the MSD and the MREBL constraints.

2.1 PEG with the MSD constraint

Let the Tanner graph consists of n symbol nodes s_j , $0 \leq j \leq n - 1$, and m check nodes c_i , $0 \leq i \leq m - 1$. Let d_{s_j} , d_{c_i} be the degree of symbol node s_j and of check node c_i , respectively, where node degree means the number of edges incident to it (this value is defined by weights distributions $\lambda(x)$ and $\rho(x)$), E_x is the set of

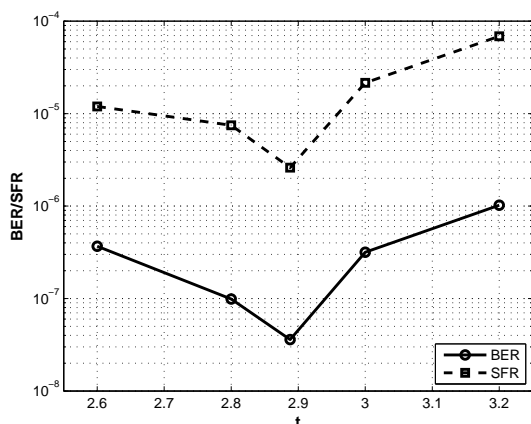


Figure 1: Simulations results on AWGN channel for matrices characterized by different average column weights.

edges incident to node x , while E_x^y represents the y^{th} edge incident to node x . Let $N_{s_j}^l$ be the set of check nodes that can be reached from symbol node s_j by l edges or less and $\overline{N_{s_j}^l}$ the complementary set of $N_{s_j}^l$: in other words, $N_{s_j}^l \cup \overline{N_{s_j}^l} = V_c$, where V_c is the set of all check nodes in the graph. Finally, let B_{s_j} be the set of check nodes $\{c_p\}$ that can be reached from symbol node s_j such that, given two edges incident to c_p , i.e. $E_{c_p}^u \leftarrow (c_p, s_a)$ and $E_{c_p}^v \leftarrow (c_p, s_j) \forall u, v \in \{1, \dots, d_{c_p}\}$, $u \neq v$ and $a \in \{0, \dots, n-1\}$, $a \neq j$, $|a-j| \geq L$, with fixed L : at this point, it is possible to define $\overline{N_{s_j}^l} \cap B_{s_j} = B_{s_j}^l$ and L means the desired MSD length. The procedure of constructing the Tanner graph is the following:

```

for  $j = 0$  to  $n - 1$  do
begin
  for  $k = 0$  to  $d_{s_j} - 1$  do
    begin
      if  $j = 0$ 
         $A_{s_j} = V_c$ 
         $A_{s_j}^l = \overline{N_{s_j}^l}$ 
         $A_{s_j}^{l+1} = \overline{N_{s_j}^{l+1}}$ 
      else
         $A_{s_j} = B_{s_j}$ 
         $A_{s_j}^l = B_{s_j}^l$ 
         $A_{s_j}^{l+1} = B_{s_j}^{l+1}$ 
      end
    if  $k = 0$ 

```

$E_{s_j}^0 \leftarrow \text{edge}(c_i, s_j)$ where $E_{s_j}^0$ is the first edge incident to s_j and c_i is one check node picked from the set A_{s_j} such that it has the lowest check degree under the current graph setting $\bigcup_{r=0}^{j-1} E_{s_r}$.

else

expanding a tree from symbol s_j up to depth l under the current graph setting such that $A_{s_j}^l \neq \emptyset$ but $A_{s_j}^{l+1} = \emptyset$, or the cardinality of $A_{s_j}^l$ stops increasing but is less than m , then $E_{s_j}^k \leftarrow \text{edge}(c_i, s_j)$, where $E_{s_j}^k$ is the k^{th} edge incident to s_j and c_i is one check node picked from the set $A_{s_j}^l$ having the lowest check-node degree

end

end

end

2.2 PEG with the MREBL constraint

The construction procedure of the Tanner graph is similar to the one described in the paragraph above, except for the meaning of some definitions: B_{s_j} represents the set of check nodes $\{c_q\}$ that can be reached from symbol node s_j such that $c_q \neq \tilde{c}$, $|B_{s_j}| = m - 1$, where $\tilde{c} \in \{c_i, i = 0, \dots, m-1\}$ is the checknode with the lowest degree under the current graph setting that includes the symbol nodes s_a , $a \in I_j^L$, where L is fixed and I_j^L is defined as the following:

$$I_j^L = \begin{cases} [0, j] & \text{if } j < L \\ [j-L, j] & \text{otherwise} \end{cases} \quad (1)$$

According to these assumptions, it is possible to define $\overline{N_{s_j}^l} \cap B_{s_j} = B_{s_j}^l$ and L as the desired MREBL length.

2.3 Parity-check matrices

In order to test different way of optimizing the PC matrix with respect to the burst error correction capabilities, we have considered the five LDPC code matrices reported in table 1. The first matrix represents the result of a mild MSD maximization which results also in a good MREBL value, the second and the third ones have been obtained by respectively optimizing the MREBL and the MSD, while the last two have been constructed with a standard PEG search algorithm

without burst error constraints. For all the matrices we have computed the MSD, MREBL and the number of maximum iterations requested for correcting a burst of length equal to MREBL; each PC matrix have $(N, rate, t, q) = (1152, 8/9, 2.88, 16)$. The $GF(16)$ q-ary codes are obtained by multiplying each non zero element of the mother PC matrix by a random number defined in the non-binary field. The error burst length correction capabilities, reported in table 1, are given in bits (i.e. MREBL $\times \log_2(q)$). For sake of reader's convenience the MREBL of the RS code adopted for the comparison is given.

Table 1: PC matrices

| PC matrix | MSD | MREBL | MREBL number of iterations |
|-----------|-----|-------|----------------------------------|
| mat1 | 20 | 100x4 | 23 |
| mat2 | 1 | 104x4 | 28 |
| mat3 | 38 | 90x4 | 14 |
| mat4 | 1 | 85x4 | 10 |
| mat5 | 1 | 94x4 | 12 |
| RS | - | 25x10 | - |

3 System model description

A non-binary LDPC coded scheme has been tested on AWGN channel and a perpendicular magnetic recording channel. The LDPC decoder runs a maximum of 25 iterations of a standard message passing (MP) algorithm ([3], [6], [7]) in both cases.

3.1 The AWGN channel

On AWGN channel we adopt a simple binary phase-shift-keying (BPSK) modulation. In order to evaluate the performance of the different codes in presence of long bursts, we randomly inject an error burst where N_b BPSK symbols are null.

3.2 The perpendicular magnetic recording channel (PMRC)

User data in a HDD channels are generally encoded with a Run-Length-Limited code (RLL) followed by a Reed-Solomon code. In the present work RLL encoding is neglected. The Reed-Solomon is defined over $GF(2^{10})$; in the simulation results the power correction is set equal to 25 symbols. User data are separated in blocks,

namely sectors, 512 bytes long. Here, the non-binary LDPC are supposed to replace the RS code. As demonstrated later, it may be beneficial to consider also an interleaver to separate channel and LDPC code. In this case, the write path is modified according to the Figure 2. User bits are first scrambled by the interleaver and then encoded. Parity bits are then evenly distributed along user bit stream. This approach preserves the RLL constraints [13], possibly applied to the bits u_k . Encoded bits are used to generate the write currents which determine, through the inductive write disk head, the magnetization of the platter magnetic media; a different magnetization is applied whenever a bit differs from the previous one.

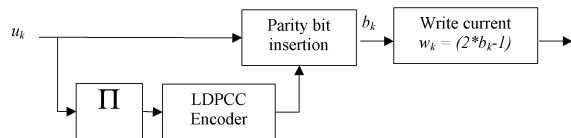


Figure 2: Write path schematic with interleaver.

The readback signal $s(t)$ of the PMR channel can be described by the following equation [14]:

$$s(t) = \frac{1}{2} \sum a_k h(t - kT + \tau_k) + n(t) \quad (2)$$

where $a_k = (w_k - w_{k-1})$, $w_k \in \{+1, -1\}$ are the recorded data and $n(t)$ is the electronic noise. $h(t)$ is the transition waveform; it corresponds to the current modification induced when the magneto-resistive read head detects a magnetic field change. The transition waveform model is:

$$\alpha \cdot \operatorname{erf} \left(\frac{2\sqrt{\ln 2}}{T \cdot CBD} t \right) + (1-\alpha) \cdot \tanh \left(\frac{\ln(3 + 2\sqrt{2})}{T \cdot CBD} t \right) \quad (3)$$

with $\alpha = 0.7$. The coded bit density CBD is defined as $PW50/T$, where $PW50$ is the width of the transition waveform derivative at half the maximum amplitude. CBD will be derived as $CBD = UBD/Rate$ where UBD is the user bit density (the uncoded bits) and $Rate$ takes into account the redundancy of Reed-Solomon or LDPC code. The jitter transition noise (or media noise) is generated by the uncertainty of the transition position τ_k which is a zero-mean Gaussian random variable. The SNR definition is given by:

$$SNR_{dB} = 10 \log_{10} \left(\frac{1}{N_0 + M_0} \right) \quad (4)$$

The noise power is split into media noise M_0 and thermal noise N_0 according to the parameter $0 \leq mix \leq 1$, so that:

$$\begin{aligned} N_0 &= mix \cdot 10^{-SNR_{dB}/10} \\ M_0 &= (1 - mix) \cdot 10^{-SNR_{dB}/10} \end{aligned} \quad (5)$$

The jitter noise power is given by:

$$\frac{M_0}{2} \cong P_t \cdot \sigma_\tau^2 \int_{-\infty}^{+\infty} \left(\frac{dh(t)}{dt} \right)^2 dt \quad (6)$$

where σ_τ is the variance of the random variable τ and $P_t = 0.5$, is the transition probability.

The read path described in Figure 3 consists of an anti-alias continuous time filter and an analog-to-digital converter (ADC) followed by a digital adaptive 10 taps finite impulse response (FIR) filter which equalizes the read signal to match a generic partial response target. A maximum likelihood (ML) sequence detection is performed with a Data-Dependent-Noise-Predictive (DDNP) [15] Soft Output Viterbi Algorithm (SOVA) [16]. The SOVA trellis has 8 states. The branch whitening filters have three taps. For the present work simulations the partial response target is [4 10 7 -2]. The LDPC decoder is possibly separated by the interleaving block. Error bursts are simulated with an attenuation of N_b consecutive read-back signal samples in a random location along the sector; thermal and media noise are added after sample attenuation.

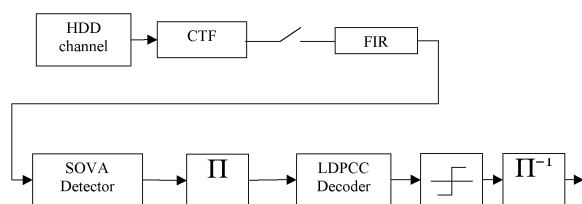


Figure 3: Read path schematic.

4 AWGN Simulation Results

Performance has been evaluated by comparing the proposed scheme with a non-binary RS coding scheme with the same redundancy, designed and shortened to work in HDD sector of 512 bytes as already mentioned in section 3.2. In Figure 4 the performance for the sector-failure rate (SFR) versus the average signal-to-noise ratio per bit (E_b/N_0), is shown in presence of only AWGN noise. According to the code parameters definition, a sector corresponds to a codeword. In

this sense a sector failure is detected when the maximum number of iterations is reached without finding a zero syndrome codeword. According to the main literature results, q-ary LDPC outperform RS of about 2 dB. Furthermore all the PC matrices considered show the same performance according to the fact that they have all been constructed by a PEG search with the same average column weight $t = 2.88$.

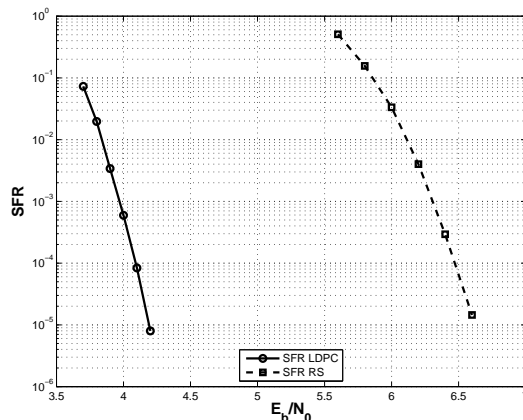


Figure 4: Performance results on AWGN channel.

Figure 5 and 6 show performance results on AWGN channel with burst erasure of length respectively of 100 and 150 bits. It seems that the PC matrix optimized for maximizing the MSD exhibits the best performance, as further shown in Figure 7, where the error burst is 175 bits long. Actually, the MREBL measure assumes an ideal erasure bursts where perfect knowledge of the other bits is given. In real burst-error channel situations, erasure bursts are observed in presence of other noise sources. Ongoing works inspired to [10] are devoted to a theoretical explanation of these results.

5 PMRC Simulation Results

Results in Figure 8 show that in absence of error burst, codes generated with (mat1) or without (mat5) MSD constraint performs similarly. When an error burst of 150 bits with attenuation 0.5 affects the sector, it is confirmed that the MSD optimization provides with better performance. Additionally, the performance of the designed code with the largest MSD (mat3) is reported. Its burst robustness is further confirmed. Performance can be compared with a RS code with same redundancy. The LDPC gain is about 1dB. However, RS slopes looks steeper compared

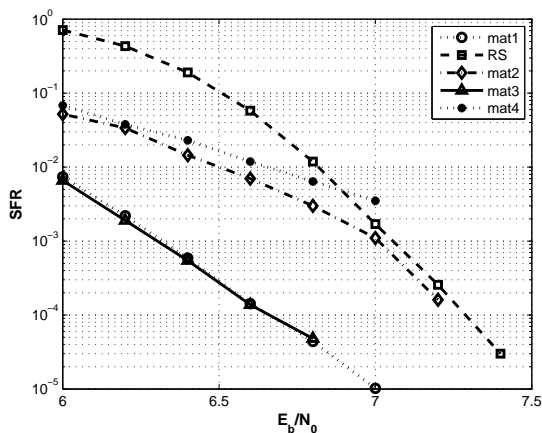


Figure 5: Performance results on AWGN channel in presence of bursts of length = 100 bits.

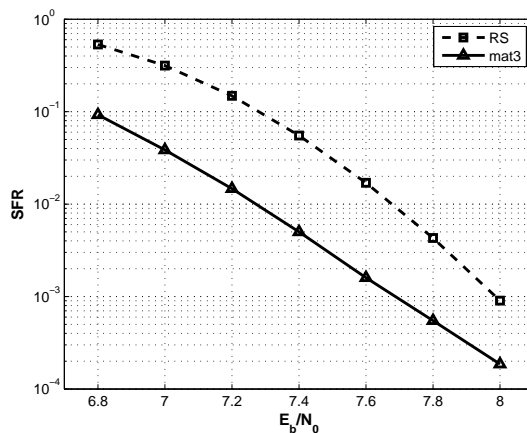


Figure 7: Performance results on AWGN channel in presence of bursts of length = 175 bits.

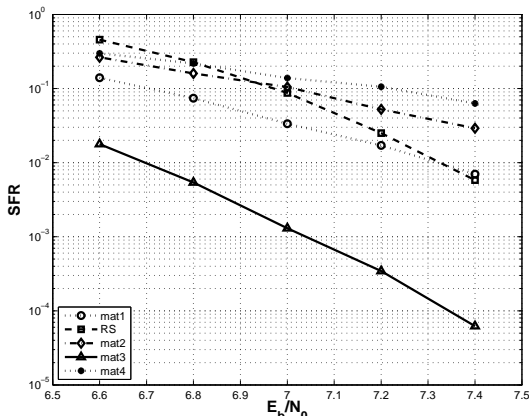


Figure 6: Performance results on AWGN channel in presence of bursts of length = 150 bits.

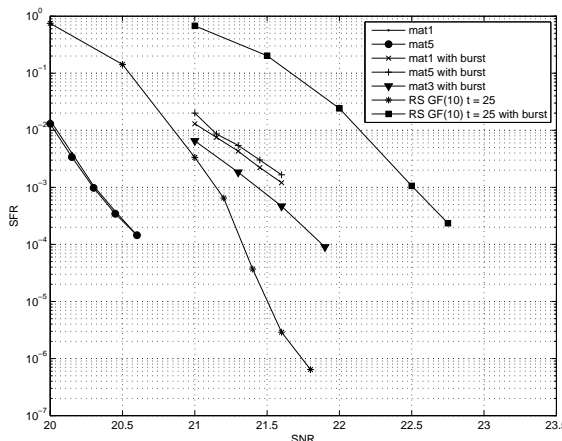


Figure 8: Performance results on PMRC channel: results without and with a bursts of length = 150 bits are compared with the RS code.

to LDPC. The present day explanation is given by the absence of an interleaver separating the ISI detector and the LDPC decoder. In fact, ISI detector generates short burst of correlated errors; most of them affect two to six consecutive bits. Correlated error can results in a suboptimal estimation of the GF symbol likelihoods which are the LDPC decoder input. To prove this point, the same LDPC codes have been simulated with a random interleaver as already described in section 3.2. It is obvious that, in this case, the optimization of the code against error burst is lost. Nevertheless, their performance are significantly improved as demonstrated in Figure 9. Burst error optimization gain is no more apparent. As a consequence, further work is necessary in order to optimize the codes separated by an interleaver.

6 Conclusion

The presented results highlight some interesting design parameters to take into account for the construction of efficient burst error correction LDPC codes. Although the MREBL is a valid performance indicator of a code under ideal noise burst, MSD optimization looks more promising when burst error correction robustness is required in random noise channels like AWGN and PMRC. Performance on AWGN channel demonstrate that non-binary LDPC codes outperform RS codes in presence of long bursts. On the other hand, on PMR channels, non-binary codes suffers from performance degradation due the ISI detector correlated outputs. Two interesting perspectives of development are both the joint optimization of the PC matrix and interleaver and a more so-

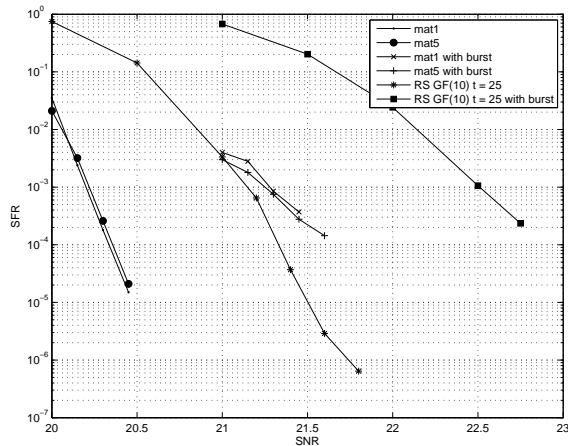


Figure 9: Performance results on PMRC channel: results without and with a bursts of length = 150 bits are compared with the RS code; an interleaver separates ISI detector and LDPC decoder.

phisticated method for computing the likelihoods of the non-binary symbols in order to match the detector soft outputs to the LDPC decoder.

References

- [1] M. Yang, W.E. Ryan, "Performance of Efficiently Encodable Low-Density-Parity-Check Codes in Noise Bursts on the EPR4 Channel," *IEEE Trans. on MAG*, vol. 40, no. 2, March, 2004.
- [2] J. Bellorado, A. Kavčić, L. Ping, "Soft-Input, Iterative, Reed-Solomon Decoding using Redundant Parity-Check Equations," in proc. *ITW 2007*
- [3] H. Song, J.R. Cruz, "Reduced-Complexity Decoding of Q-ary LDPC Codes for Magnetic Recording," *IEEE Trans. on MAG*, vol. 39, no. 2, March 2003.
- [4] J. Chen, L. Wang, Y. Li, "Performance Comparison between Non-binary LDPC Codes and Reed-Solomon Codes over Noise Bursts Channels," in proc. *International Conference on Communications, Circuits and Systems 2005 (ICCCAS 2005)*, vol. 1, May 2005.
- [5] E. Paolini, M. Chiani, "Improved LDPC Codes for Burst Erasure Channels," in proc. *IEEE International Conference on Communications 2006 (ICC 2006)*, vol. 3, pp. 1183-1188, June 2006.
- [6] M.C. Davey, D. MacKay, "Low-Density Parity Check Codes over GF(q)," *IEEE Communications Letters*, vol. 2, no. 6, June 1998.
- [7] M.C. Davey, "Error-Correction Using Low-Density-Parity-Check Codes," Ph.D. Thesis, University of Cambridge, UK, December 1999.
- [8] A. Ghaith, J.B. Boutros, Y. Yuan-Wu, "Fast and Reduced-Complexity Decoding Rule for q-ary LDPC Codes by Using the Duality properties," in proc. *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, 2006. (WiMob'2006)* June 19-21, 2006, Montreal, Que.
- [9] G. Hosoya, H. Yagi, S. Hirasawa, "Modification Methods for Construction and Performance Analysis of Low-Density Parity-Check Codes over the Markov-Modulated Channel," in proc. *International Symposium on Information Theory and its Applications, ISITA 2004*, October 10-13 2004, Parma, Italy.
- [10] E.A. Krouk, S.V. Semenov, "Low-Density Parity-Check Burst Error-Correcting Codes," in proc. *2nd International Workshop Algebraic and combinatorial coding theory*, pp. 121-4, 1990.
- [11] E.A. Krouk, S.V. Semenov, *Error Correcting Coding and Security for Data Networks*, McGraw Hill.
- [12] E. Eleftheriou, S. Özçer, "Low-Density Parity-Check Codes for Digital Subscriber Lines," in proc. *IEEE International Conference on Communications, ICC 2002*, April 28-May 2, 2002.
- [13] Jin Lu and Keith G. Boyer, "Novel RLL-ECC Concatenation Scheme for High-Density Magnetic Recording," *IEEE Trans. on Magnetics*, vol. 43, no. 6, pp. 2271-2273, June 2007.
- [14] B. Vasic (Editor), E. M. Kurtas (Editor), *Coding and Signal Processing for Magnetic Recording Systems*. CRC Press.
- [15] A. Kavčić and J. F. Moura, "The Viterbi algorithm and Markov noise memory," *IEEE Trans. Inform. Theory*, vol. 46, no. 1, pp. 291-301, Jan. 2000.

- [16] J. Hagenauer, P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," in proc. *IEEE GLOBECOM*, pp. 47.11-47.17, Dallas, TX, Nov 1989.
- [17] X.Y. Hu, E. Eleftheriou, D.M. Arnold, "Regular and Irregular Progressive Edge-Growth Tanner Graphs," *IEEE Trans. on Information Theory*, vol. 51, no.1, Jan. 2005.
- [18] B. Rong, T. Jiang, X. Li, M.R. Soleymani, "Combine LDPC Codes Over GF(q) With q-ary Modulations for Bandwidth Efficient Transmission," *IEEE Trans. on Broadcasting*, vol. 54, no. 1, March 2008.