

# SYSTEMATIC LUBY TRANSFORM CODES AND THEIR SOFT DECODING

*T. D. Nguyen, L. L. Yang and <sup>1</sup>L. Hanzo*

School of ECS, University of Southampton, SO17 1BJ, UK.

Tel: +44-23-8059 3125, Fax: +44-23-8059 4508

Email: <sup>1</sup>lh@ecs.soton.ac.uk, <http://www-mobile.ecs.soton.ac.uk>

**Abstract** – Luby Transform codes (LT) were originally designed for the Binary Erasure Channel (BEC) encountered owing to randomly dropped packets in the statistical multiplexing aided classic wireline-based Internet, where transmitted packets are not affected by the fading or noise of the propagation environment of the wireless Internet. For the sake of transmitting data over the BEC routinely encountered in statistical multiplexing aided wireless Internet - style scenarios, we applied the belief propagation algorithm for decoding LT codes and designed a novel version of LT codes, which we refer to as systematic LT codes. When using soft decoding of the proposed systematic LT code, the decoding process becomes capable of preventing the potentially avalanche-like inter-packet error propagation. For example, the systematic LT(1000,3000) code achieved a BER below  $10^{-5}$  at  $E_b/N_0 = 3.5dB$  after six decoding iterations. An even lower  $E_b/N_0$  of 2.7dB was required, when using a longer systematic LT(10000,30000) code for transmission over the AWGN channel. In the combined BEC-AWGN channel the BER recorded at the output of the systematic LT(1000,3000) code was about  $10^{-5}$  at  $E_b/N_0 = 4.5dB$ , when encounter an erasure probability of  $P_e = 0.1$ .

## 1. INTRODUCTION

LT codes [1] were originally designed for the Binary Erasure Channel (BEC) channel, where transport packets may be erased with an erasure probability of  $P_e$ . When applying LT codes for transmitting data over wireless channels contaminated by Rayleigh fading and inter-symbol interference (ISI), the packets may become contaminated, which may result in catastrophic inter-packet error propagation during LT decoding [2]. LT codes were originally designed for error-free BEC channels. For the sake of using LT codes to protect data for transmission over the wireless Internet, where fading, noise and packet erasures are encountered, numerous researchers endeavoured to improve the achievable performance [2] [3] [4]. Naturally, the complexity of these schemes

tends to be increased. For the sake of improving the error correction capability of LT codes, in this paper we propose the **soft decoding of LT codes** using the probabilistic decoding technique of Low Density Parity Check Codes (LDPC) [5]. The outline of the paper is as follows. Section 2 analyses the disadvantages of the LT codes, when applying the message passing algorithm and detail the design of **systematic LT codes**. Section 3 analyses **systematic LT codes** using extrinsic information transfer (EXIT) charts and Section 4 characterizes their performance over the wireless Internet channels. Finally, Section 5 provides our conclusions.

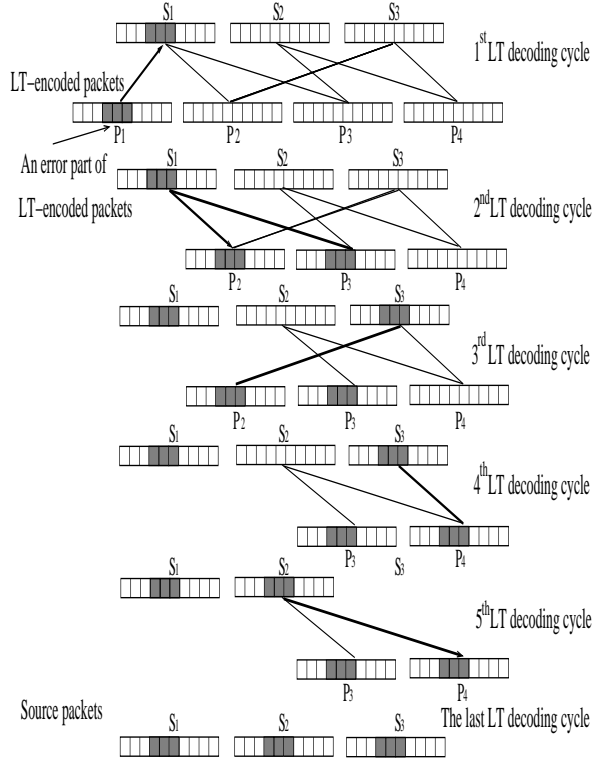
## 2. LT CODE DESIGN FOR SOFT DECODING AND THE SYSTEMATIC-LT CODE

LT codes were originally designed for hard-decoding [1] in the context of the BEC. The error propagation phenomenon of the LT decoding process is portrayed in Fig 1. For the sake of avoiding this detrimental effect, they have been combined with various forward error correcting (FEC) codes [2] [4]. These combined schemes substantially mitigated the effects of error propagation. However, the attainable performance improvement of these schemes were still limited owing to the employment of hard-decision aided LT decoding. For the sake of circumventing this deficiency, we introduce the novel concept of soft LT decoding based on the classic belief-propagation technique applying Tanner Graphs. We commence our discourse by introducing the concept of single-bit packets. The LT codes having larger packets will be considered as generalized the scenario. The soft LT decoding process is based on the classic concept of LDPC decoding. Given the generator matrix  $G$  of the LT code, we calculate the Parity Check Matrix (PCM)  $H$  of the LT code similarly to that of a classic LDPC code, namely by dividing the LT code's generator matrix into two matrices, where  $A$  and  $B$  have a size of  $(K \times K)$  and  $(K \times M)$ , respectively. We choose the non-singular matrix  $A$  based on the conventional LT decoding process. Then the PCM is calculated as

$$H_{(M \times N)} = [(B^T \cdot (A^T)^{-1})_{(M \times K)} | I_{(M \times M)}]. \quad (1)$$

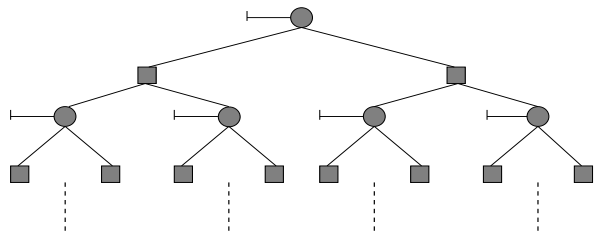
The LT decoding process may be implemented as follows. An LT PCM can be represented by a classic Tanner graph [4].

<sup>1</sup>The financial support of the EPSRC, UK and that of the EU is gratefully acknowledged



**Fig. 1.** An example of error propagation in the LT hard-decoding process, when the LT decoder receives packets affected by the channel error

To elaborate a little further, the filled circles and the filled



**Fig. 2.** A tree-based representation of LT code

squares of Fig. 2 represent the LT check nodes and the LT check nodes, respectively, while the horizontal lines connected to the variable nodes represent the intrinsic information provided by the channel's output. Let us assume that the circular node at the top of Fig. 2 represents the  $k^{th}$  variable node in the block of  $N$  number of single-bit LT encoded packets, which is also termed as the root node [6]. The root node receives information from the check nodes it is connected to at the level seen below it in Fig. 2 and those check nodes also receive information from the variable nodes they are connected to at the next level down, etc. The dotted lines in Fig. 2 indi-

cate that the above process is repeated further by expanding the tree. The number of connections associated with a variable node of the LT code - excluding the line representing the intrinsic information - indicates the column weight of this particular message node, while the number of connections associated with an LT check node represents the corresponding row weight. The column weight and row weight of the LT PCM are related to the degree distribution of LT packets.

The LT decoding process is implemented in the same way as the classic LDPC decoding procedure. Initially, the LT decoder's soft values are set to a value corresponding to the demodulator's soft output. The decoder's soft values of  $R_{i,j}^a$  and  $Q_{i,j}^a$  which denote the LLRs passed from the check nodes to the variable nodes and vice versa are then iteratively updated after each decoding iteration as follows:

$$\tanh\left(\frac{R_{i,j}}{2}\right) = \prod_{n \in \{C_i\}, n \neq i} \tanh\left(\frac{Q_{n,j}}{2}\right), \quad (2)$$

where we have

$$\tanh(x/2) = \frac{e^x - 1}{e^x + 1}. \quad (3)$$

After each iteration, the LT decoder outputs its tentative hard-decision and checks, whether the product of the corresponding codeword and the transpose of the PCM  $H$  is equal to zero i.e whether a legitimate codeword was produced. If not, the LT decoding process will be continued in an iterative fashion, until the output codeword becomes legitimate or the maximum affordable number of iterations is exhausted. For the sake of improving the performance of LT codes at high  $E_b/N_0$  values, we invoke hard-decoding after the last LT decoding cycle in order to erase the low-confidence LT packets, namely those, which have low Logarithm Likelihood Ratios (LLRs). We employ a low complexity packet-reliability evaluation technique based on the average LLR of the packet. However, the soft LT decoding process based on the LT code's PCM exhibits some deficiencies owing to the following two reasons:

- the LT code's PCM contains many zero-columns, which degrades the performance of the LT code, when using the above mentioned soft decoding process.
- The conventional non-systematic LT code will impose error propagation, when using the above-mentioned hard-decoding process for the sake of recovering the original information packets from the erroneously decoded LT-encoded packets, as seen in Fig.1.

Hence, for the sake of improving the LT code's performance in hostile wireless channels, we specifically design the LT code's degree distribution by expanding its generator matrix with the aid of a unity matrix having a size of  $(K \times K)$ , which results in a systematic LT code. The resultant relationship of the systematic LT generator matrix  $G$  and the PCM  $H$  may be stated as follows:

|                 |                                 |
|-----------------|---------------------------------|
| 1 0 0 0 0 0 0 0 | 1 0 1 1 1 1 1 1 1 0 1 1 1 1 0   |
| 0 1 0 0 0 0 0 0 | 0 0 0 0 0 1 0 0 0 0 0 1 0 1 0   |
| 0 0 1 0 0 0 0 0 | 1 1 0 1 1 1 1 1 1 1 0 1 1 1 0   |
| 0 0 0 1 0 0 0 0 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1   |
| 0 0 0 0 1 0 0 0 | 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0   |
| 0 0 0 0 0 1 0 0 | 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 |
| 0 0 0 0 0 0 1 0 | 1 1 1 1 1 1 1 1 0 0 0 0 1 1 0 0 |
| 0 0 0 0 0 0 0 1 | 1 1 0 1 1 1 1 1 1 0 0 1 1 1 1 1 |

Fig. 3. The Systematic LT generator matrix

- If we have a generator matrix  $G_{K \times N} = [I_{K \times K} / A_{K \times M}]$ , where  $I$  is an identity matrix having a size of  $[K \times K]$  and  $A$  is a non-singular matrix having a size of  $[K \times M]$ , then the PCM obeys the construction of  $H = [A^T / I']$ , where  $A^T$  is the transpose of  $A$  and  $I'$  is an identity matrix having a size of  $[M \times M]$  [7], where  $N = K + M$  is the number of rows in  $G$  and  $K$  is the number of columns in  $G$ .

### 3. ANALYSIS OF SYSTEMATIC LUBY TRANSFORM CODES BASED ON EXIT CHARTS

Extrinsic Information Transfer (EXIT) charts [8] [9] were proposed by ten Brink and his collaborators. As mentioned in Section 2, systematic LT codes may be decoded by belief propagation. Hence, we can analyse systematic LT codes based on EXIT charts [8] [10]. Since systematic LT codes have a specific PCM construction, which contains the check bits in the corresponding rows and the message bits in the corresponding columns, while the second part of the PCM is constituted by a unity matrix having a size of  $M \times M$  containing the check bits in both its rows and columns. Hence, we have to re-define the passing of the LLR messages during the systematic-LT decoding process as follows:

- The LLR messages are passed between the systematic LT message nodes and the systematic LT parity nodes.

Let  $d_m$  represents the degree of the message nodes and  $d_c$  the degree of the parity nodes. Furthermore, let  $K$  denote the number of message nodes and  $N$  the number of variable nodes. Then the number of parity nodes is given by  $M=N-K$ . The degree distribution  $D(d_c)$  of the check nodes is defined as follows:

$$D(d_c) = \begin{cases} 0 & \text{if } d_c = 1 \text{ or } d_c > \frac{M}{S}, \\ \frac{1+S+\nu}{Z^M} & \text{if } d_c = 2, \\ \sum_{t=3}^{M-1} \left( \frac{1}{t(t-1)} + \frac{S}{M} \cdot \frac{1}{t} \right) & \text{if } 3 < d_c < M, \\ \frac{S}{M} \cdot \log \frac{S}{\delta} & \text{if } d_c = \frac{M}{S}, \end{cases} \quad (4)$$

where  $S, \nu$  were defined in [2]. We assume that **the LT encoding process of [2] using the ideal uniform random degree distribution is employed**. Then the degree distribution of the message nodes  $D(d_v)$  is defined as:

$$D[d_m(\bar{d}_c)] = r, \quad (5)$$

where  $r = \frac{K}{N}$  is the code rate of the systematic LT code. Let us now consider the messages passed between the systematic LT message nodes and the check nodes, when the LLR representation of the messages. Let  $Q'$  as well as  $R'$  denote the LLR information passed from the information nodes to the check nodes and that passed from the check nodes to the information nodes, respectively. The soft bit message corresponding to  $Q'$  is repeated here again

$$Q = \tanh\left(\frac{Q'}{2}\right). \quad (6)$$

The extrinsic LLR information passing from the check nodes to the information nodes is defined by:

$$Q' = \sum_{i=0}^{d_m-1} R'_i, \quad (7)$$

where the initial soft channel-output message associated with the variable nodes is given by  $(R'_0 = \frac{4}{N_0} \cdot y)$  and  $y$  denotes the output of an AWGN channel. Furthermore,  $R'_i, i = 1 \dots d_v - 1$  represents the LLR information arriving from the check nodes to the information nodes, except from that particular check node to which the LLR information message  $Q'$  was sent. The extrinsic LLR information  $R'$  passed from the check nodes to the information nodes is defined as:

$$\tanh\left(\frac{R'}{2}\right) = \prod_{i=1}^{d_c-1} \tanh\left(\frac{Q'_i}{2}\right) \iff R = \prod_{i=1}^{d-1} R_i, \quad (8)$$

where  $Q'_i, i=1, \dots, d_c-1$  represents the LLR information arriving from the information nodes, except from that particular information node to which the LLR information message  $R'$  was sent. Let  $m_R, m_{R_0}$  and  $m_Q$  denote the mean of  $R', R'_0$  and  $Q'$ . Then, from Equation (7) we have:

$$m_Q^{(l)} = m_{R_0} + (d_m - 1) \cdot m_R^{(l-1)}, \quad (9)$$

where  $l$  is the  $l$ 'th iteration and  $m_{R_0} = 4 \cdot \frac{E}{N_0}$ , while  $E$  is the transmitted bit energy, while  $\frac{N_0}{2}$  is the one-side power spectral density of the noise. We assume that  $R'$  is Gaussian distributed. Hence,  $m_R$  is can be updated according to:

$$m_R^{(l)} = J^{-1}(I(X; R^{(l)})), \quad (10)$$

where  $J(m_R)$  is defined as follows:

$$J(m_R) = I(X; R^l) = \int \frac{1}{\sqrt{4\pi m_R}} e^{-\frac{(l-m_R)}{4m_R}} (1 - \log_2(1 + e^{-l})) dl. \quad (11)$$

From the equation used for calculating the mutual information in [11] we have:

$$I(X; R') = \frac{1}{\ln 2} \sum_{i=1}^{\infty} i = 1 \frac{1}{2i(2i-1)} [E(T_Q^{2i})]^{d_c-1}, \quad (12)$$

where:

$$\begin{aligned} \phi_i(m_Q) &= E(T_R^{2i}) \\ &= \int_{-1}^{+1} \frac{2t^{(2i)}}{(1-t^2)\sqrt{4\pi m_Q}} e^{-\frac{(\ln(\frac{1+t}{1-t}) - m_Q)^2}{4m_Q}} dt. \end{aligned} \quad (13)$$

Finally, we arrive at the required update formulae for the means  $m_Q$  and  $m_R$  as follows:

$$m_Q^{(l)} = m_{R_0} + (d_m - 1)m_R^{(l-1)} \quad (14)$$

$$m_R^{(l)} = J^{-1}\left(\frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{2i(2i-1)} [E(T_{m_Q}^{2i})]^{d_c-1}\right). \quad (15)$$

Based on Equation (14) we arrive at the variable node's EXIT function:

$$\begin{aligned} I_{E_m} &= I(X; Q') = I(X; R'_0, R'_1, \dots, R'_{d_m-1}) \\ &= f(I(X; R'_0), I(X; R')) = f(I_{ch}, I_{A_m}). \end{aligned} \quad (16)$$

Similarly, from Equation (15) we derive the check node's EXIT function as follows:

$$\begin{aligned} I_{E_c} &= I(X; R') = I(X; Q'_1, \dots, Q'_{d_c-1}) \\ &= f(I(X; Q')) = f(I_{A_c}) \\ &= \frac{1}{\ln 2} \sum_{i=1}^{\infty} \frac{1}{(2i-1)(2i)} [\phi_i(J^{-1}(I_{A_c}))]^{d_c-1}, \end{aligned} \quad (17)$$

where  $I_{ch} = I(X; R'_0)$  is the average channel output information,  $I_{A_m} = I(X; R')$  is the average a priori information at the input of the message node decoder and  $I_{E_m}$  is the average extrinsic information at the output of the message node decoder.

The systematic LT code has the degree distributions  $D(d_v)$  and  $D(d_c)$  formulated in Equations (4) and (5). Therefore, the EXIT functions of the message and check nodes are given in Equations (16) and (17), which depend on the degrees of the message nodes and check nodes, respectively. Again, we assumed that **the LT encoding process of [2] using the ideal uniform random degree distribution** is employed. Hence, all the LT message nodes have the same degree, namely a degree of  $\bar{d}_c - 1$ . Finally, we arrive at the EXIT function of the message nodes and the check nodes for the systematic LT code are expressed in the following form:

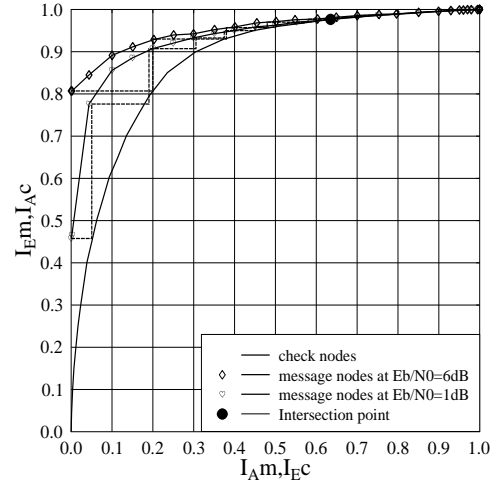
- Message node EXIT function:

$$\begin{aligned} I_{E_m} &= f(I_{ch}, I_{A_m}) \\ &= J(J^{-1}(I_{ch} + (j-1)J^{-1}(I_{A_m}))); \end{aligned} \quad (18)$$

- Check node EXIT function:

$$\begin{aligned} I_{E_c} &= f(I_{A_c}) = \sum_{j=2}^{d_c} d_{c_j} I_{E_{c_j}} \\ &= \sum_{j=2}^{d_{c_{max}}} d_c(j) \cdot \frac{1}{\ln 2} \sum_{i=1}^{\infty} \left(\frac{1}{2i(2i-1)}\right) [\phi_i(J^{-1}(I_{A_c}))]^{j-1} \end{aligned} \quad (19)$$

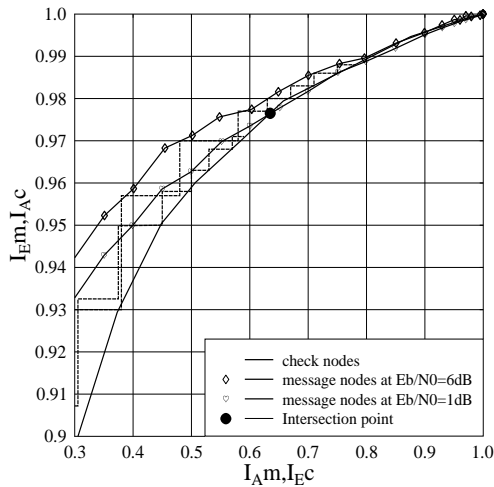
The EXIT chart of the systematic LT code is portrayed in Fig 4. The associated parameters are:



**Fig. 4.** EXIT chart of the systematic LT(1000,3000) code

- The check bits obey the same degree distribution as the LT packets, namely the Full Robust Soliton Distribution (RSD) proposed in [2].
- All of the message bits have a degree of one.
- A total of 500 blocks are transmitted and each of them has 1000 bits. The rate of the systematic LT code was set to  $r = \frac{1}{3}$ .

The systematic-LT code's message-node-related EXIT curve recorded at  $E_b/N_0 = 1dB$  intersects the check-node-related curve relatively close to the point of perfect convergence, namely the (1,1) point of Fig 4, which was made more visible by expanding the top right corner of Fig 4 in Fig 5. The point of intersection is marked by a bold dot in Fig. 5. Owing to the

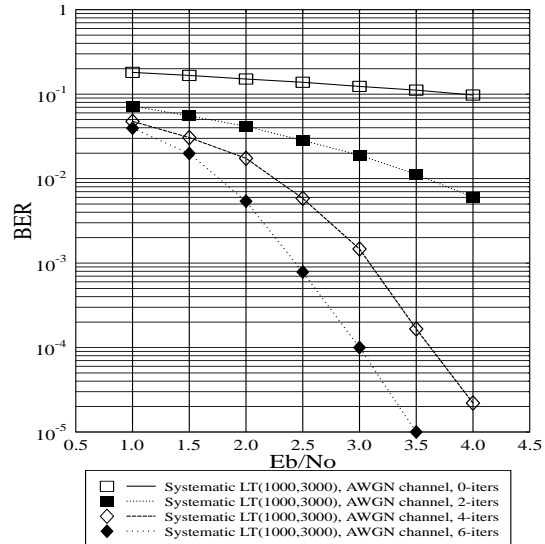


**Fig. 5.** Expanded view of the Systematic-LT(1000,3000) code's EXIT chart

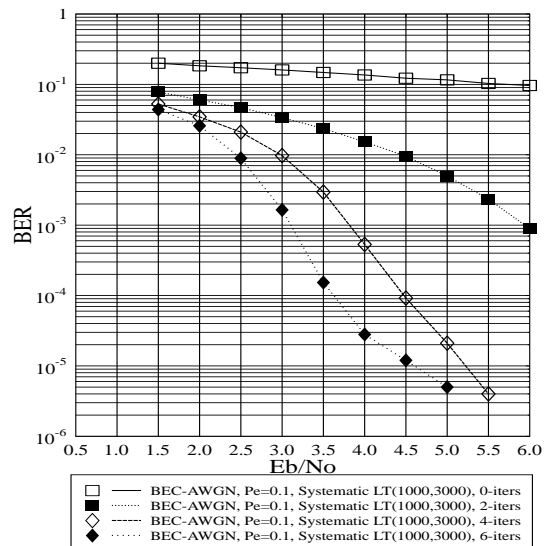
intersection between the message-node's curve and the check-node's curve, the systematic-LT decoder is unable to reach an infinitesimally low BER at  $E_b/N_0 = 1dB$ . By contrast, observe in Fig. 5 that the systematic-LT decoder becomes capable of converging to the (1,1) point at  $E_b/N_0 = 6dB$  after 10 iterations.

#### 4. SIMULATION RESULTS

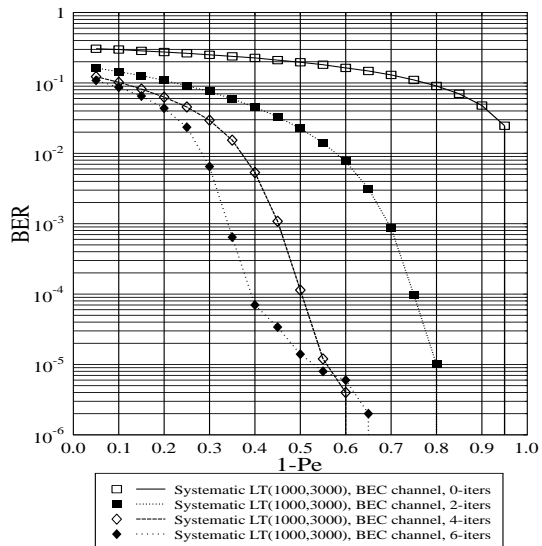
Fig. 6 shows the attainable performance of the systematic LT(1000,3000) code using the message-passing decoding technique, when communicating over the AWGN channel, having an erasure probability of  $P_e=0$ . Fig 6 may be contrasted to Fig 7, where transmission over the Wireless Internet associated with different  $E_b/N_0$  values and  $P_e = 0.1$  is considered. Fig. 8 shows the performance of the systematic LT code for transmission over the BEC having different erasure probabilities  $P_e$ , where the abscissa axis was scaled in terms of the values of  $(1 - P_e)$ . Again, the Robust Soliton degree distribution was used for the variable nodes of the systematic LT(1000,3000) code characterized in Fig. 6, Fig. 7 and Fig. 8. The degree distribution parameters of  $c=0.1$  and  $\delta = 0.5$  were used as defined in [2]. We can see in Fig. 6 that for  $P_e=0$  the BER becomes as low as  $10^{-5}$  at  $E_b/N_0 = 3.5dB$ . By contrast, when the erasure probability is  $P_e=0.1$ , the decoder reaches  $BER= 10^{-5}$  at  $E_b/N_0 = 4.5dB$ . The BER curves of the systematic LT code are significantly more smooth than for other combinations of LT codes and FEC codes [2]. The proposed scheme has the potential of maintaining a lower BER at  $E_b/N_0$  and that of tolerating higher  $P_e$  values than the system of [2]. When we increase the length of the systematic-LT



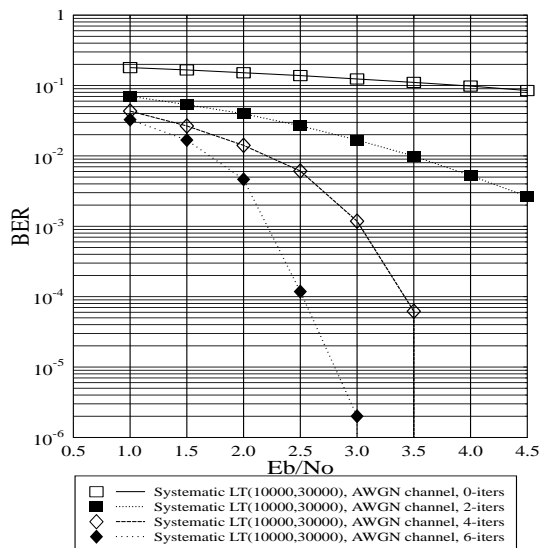
**Fig. 6.** BER versus  $E_b/N_0$  performance of the systematic LT(1000,3000) code in AWGN channels using BPSK modulation and no erasures



**Fig. 7.** BER versus  $E_b/N_0$  performance of the systematic LT(1000,3000) for transmission over the AWGN-contaminated BEC associated with  $P_e = 0.1$



**Fig. 8.** BER versus  $1 - P_e$  performance of the systematic LT(1000,3000) code in the noiseless BEC channel



**Fig. 9.** BER versus  $E_b/N_0$  performance of the systematic LT(10000,30000) for transmission over the AWGN-contaminated BEC associated with  $P_e = 0$

codes, their performance is improved, as seen in Fig. 9. More specifically, a BER of  $10^{-5}$  may be attained at  $E_b/N_0 = 3.0dB$  after 6 iterations of the systematic-LT decoder. Finally, observe in Fig 8 that the systematic LT decoder is capable of maintaining an infinitesimally low BER for erasure probabilities as high as  $P_e = 0.4$  in the BEC.

## 5. CONCLUSIONS

The concept of systematic LT codes was introduced with the aid of appropriately modifying the robust soliton degree distribution of [2]. The systematic LT coding concept facilitated the employment of the classic belief-propagation based soft decoding of the resultant systematic LT codes. The main benefit of the proposed scheme is that it is capable of mitigating the effects of catastrophic error propagation across the LT-encoded packets. Our future work will consider the employment of diverse classic FEC codes in the context of soft-decoded LT codes for the sake of improving the associated performance versus complexity as well as versus code-rate trade-offs.

## 6. REFERENCES

- [1] M. Luby, "LT codes," in *Proceeding of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, November 2002, pp. 271–282.
- [2] R. Tee, T.D. Nguyen, L-L. Yang and L. Hanzo., "Serially Concatenated Luby Transform Coding and Bit-Interleaved Coded Modulation Using Iterative Decoding for the Wireless Internet," *Proceedings of VTC 2006 Spring, Melbourne, CD ROM.*, vol. 138, no. 4, pp. 177–182, May 2006.
- [3] T. Stockhammer H. Jenkac, T. Mayer and W. Xu, "Soft decoding of lt-codes for wireless broadcast," in *Proc. IST Mobile, Dresden, Germany*, 2005.
- [4] R. Palanki, J.S. Yedidia., "Rateless codes on noisy channels," *ISIT 2004. Proceedings*, p. 37, June 2004.
- [5] T. J. Richardson and R. L. Urbanke, "The capacity of low density parity check codes under message-passing decoding," *IEEE Transactions on Information Theory*, pp. 599–618, Feb 2001.
- [6] S. Y. Chung, "On the construction of some capacity-approaching coding schemes," *Ph.D thesis, MIT, USA*, 2000.
- [7] R. Gallager, "Low Density Parity Check Codes," *IRE Transactions On Information Theory*, 1962.
- [8] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *Communications, IEEE Transactions*, 2001.
- [9] G. Kramer A. Ashikhmin and S. ten Brink, "Extrinsic information transfer functions: model and erasure channel properties," *Information Theory, IEEE Transactions*, 2004.
- [10] E. Sharon, "Analysis of belief-propagation decoding of ldpc codes over the biawgn channel using improved gaussian approximation based on the mutual information measure," *Electrical and Electronics Engineers in Israel. The 22nd Convention of*, 2002.
- [11] A. Ashikhmin E. Sharon and S. Litsyn, "Analysis of low-density parity-check codes based on exit functions," *Communications, IEEE Transactions*, 2006.