2017 4th International Conference on Electrical and Electronics Engineering

A Sequential Coding Approach for Short Length LT Codes over AWGN Channel

Nadhir Ibrahim Abdulkhaleq
Department of Electronic and Communication Engineering, Cankaya University
Ankara, Turkey
Department of Electronic Techniques, Institute of Technology, MTU Baghdad, Iraq
e-mail: c1282650@student.cankaya.edu.tr

Orhan Gazi
Department of Electronic and Communication Engineering. Cankaya University
Ankara, Turkey
e-mail: o.gazi@cankaya.edu.tr

***Abstract*—Luby Transform (LT) codes which can be considered as one of the first type of rateless codes are introduced to be an efficient erasure code. In such codes a fountain of encoded symbols is continuously transmitted until the successfully delivery of all the data packets. Whenever the receiver collects N noisy encoded symbols the decoding operation starts. Usually belief propagation (BP) algorithm is used to decipher the code and extract the k data symbols. In this paper, short length LT codes are generated using efficient sequential encoding approach (SEA) and performance measurements have been done over additive white Gaussian (AWGN) channel. This coding technique generates its degree in a repeated sequential manner which yields a mutual relation between adjacent codes that will be used in the decoding part. The decoding complexity of the proposed structure is similar to that of the Raptor codes. The simulation results show that the proposed approach has better performance in terms of error floor and successful decoding ratio when compared to LT codes using robust soliton distribution (RSD) and memory-based-RSD (MBRSD) even when supported by belief propagation-pattern recognition (BP-PR) technique.**

Keywords-rateless codes; short length LT codes; sequential encoding approach; AWGN channel; pattern recognition (BP-PR)

I. Introduction

In his PhD thesis [1] Luby introduced a class of rateless codes named as Luby Transform (LT) codes which later on became a pioneer code for erasure channels. In these codes a sparse random generating matrix is used to combine k data symbols in a linear form. Each code symbol is generated by combining a random number, called degree, of randomly selected data symbols. Designing good degree distributions is critical for the performance of rateless type codes. For the decoding operation of rateless codes belief propagation like algorithms are applied at the receiver side to process the noisy code symbols in order to extract the data symbols. To extract the k data symbols N noisy received symbols are needed. The rate of transmission is defined as $R = k / N$. The performance of LT [1] and Raptor [2] codes on noisy channels such as binary symmetric channel (BSC) and AWGN are evaluated in [3]. It was found that even with large data frames such as $k = 1000$ and $k = 10000$ these codes suffer from high error floor.

Inspecting the decoding matrix of LT codes inspired many of the early efforts whom are testing these codes over AWGN to adjust this matrix to be closed to the well-known low density parity check code (LDPC) matrix. This adjustment allows the decoder to use the iterative soft decoding of such codes [4]. Another idea for the generator matrix adjustment has been presented as systematic LT code [5]. The generator matrix of the received code symbols for such systematic LT codes become more suitable to use the standard belief-propagation soft decoding algorithm. In [6] a rate-compatible degree distribution of LT code for AWGN channels was proposed. In this study, several criteria, such as maximal rate or low complexity or even the constrained-iteration performance are taken in consideration for the optimization of the degree distribution. Similar to Shokrollahi method [2], Chen et al [7] proposed a new type of rateless code suitable for noisy channels and they had obtained their code by concatenating low density generator matrix (LDGM) rateless codes with simple post-codes, and this code structure becomes similar to that of Raptor codes. A Gaussian Jordon elimination (GJE) method has been used to overcome the problem of missing degree-one coded symbol for short length LT codes over AWGN channel has been presented in [8]. The received generator matrix $G(N\ x\ k\ )$ of the coded symbols has been used to construct the a parity check matrix $H = [G\ , I\ ]$ where $I = (N\ x\ N\ )$ identity matrix, this I matrix has been added to make check matrix suitable to use the BP decoding algorithm used in the decoding of LDPC codes [9]. Another approach has been presented in [10], where a group weight distribution of an LT code has been determined and a refined union bound used to evaluate the act of LT codes under the maximum likelihood decoding before the application of soft decoding algorithm.

In all the studies [4]-[10] additional check nodes have been introduced in the generating matrix of the received LT codes to approach the compatibility of the soft decoding algorithms. This treatment imposes significant decoding complexity and increases the time required to extract the required information symbols especially when increasing the iteration number for the soft decoding to enhance the performance.

In this paper, a less complex approach will be applied. We intend to use the exact generator matrix of the received coded symbols without any alteration. The decoding will be achieved into two steps: first, a soft demodulation will be done to estimate the coded symbols values. Second three deferent hard decoding techniques will be applied named as regular BP, BP with pattern recognition BP-PR [11] and sequential decoding (SD). The complexity of the iterative soft decoding which have been applied in [4]-[10] will be decreased using this soft-demodulation-hard -decoding (SDHD) technique. In Section II three deferent degree generation of LT codes will be presented. In Section III the system model for (SDHD) will be discussed. A short length LT code simulation results for the error rate and successful decoding ratio performance of the three degree generators will be shown in Section IV. Finally, conclusions will be listed in Section V.

II. Degree Generation Methods

Suppose that a message file of length k has been sent over an AWGN channel and LT code will be used to support the reliability of the transmission. For the LT encoding operation, first the file will be sliced into packets, these packets may have any length; (however in this paper we deal with the packet as a single bit). Next a degree distribution has to be used to generate a number called degree (d). For each code symbol, d random data symbols will be chosen uniformly and combined using binary XOR operations to form this code symbol. These three steps will be repeated until an acknowledgement signal from the destination has been received. In this paper, three deferent degree distributions will be applied named as RSD, MBRSD and SEA. The following sections point out the main features for these encoding approaches.

A. Robust Soliton Degree Distribution (RSD)

The robust soliton distribution is given by:

$$\Omega(d) = \frac{\rho(d)+\tau(d)}{\beta} \quad (1)$$

where $\rho(d)$ is the ideal soliton distributions (ISD) [1] which is defined as:

$$\rho(d) = \begin{cases} \frac{1}{k} & d = 1 \\ \frac{1}{d(d-1)} & d = 2,3 \cdots, N \end{cases} \quad (2)$$

and the probability function $\tau(d)$ will be equal to:

$$\tau(d) = \begin{cases} \frac{S}{k}\frac{1}{d} & d = 1,2,\cdots,\frac{k}{S}-1 \\ \frac{S}{k}\ln\left(\frac{S}{\delta}\right) & d = \frac{k}{S} \\ 0 & d > \frac{k}{S} \end{cases} \quad (3)$$

where S denotes the expected number of degree-one coded symbols and it will be calculated as:

$$S = c\,.\ln\left(\frac{k}{\delta}\right)\sqrt{k} \quad (4)$$

where $c \in (0,1)$ and δ is the probability of the decoding failure and $\beta = \sum_d \rho(d) + \tau(d)$. Robust soliton distribution is mainly suitable for the transmission of bulk data files.

B. Memory Based RSD (MBRSD)

In memory-based RSD the data symbol with highest degree is tracked and whenever a degree one code symbol is to be generated it is connected to the data symbol with highest degree. The algorithm for this distribution splits the degrees of the RSD into two groups, degree one and the other degrees [12]. In both RSD and MBRSD the degrees are generated in a random manner while the selection for the data symbols for each degree is uniformly at random for RSD only. The attractive advantage of the MBRSD appears in the decoding complexity. Since assigning degree one coded symbol to the highest degree data symbol will increase the number of released edges which fastens the decoding speed and increase the probability of producing new degree one in the ripple box.

C. Sequential Encoding Approach (SEA)

Many studies have been done on LT codes in the literature focused on improving the RSD and majority of them uses large file sizes. In this paper, we propose SEA for short file sizes such as $k = 32$ or $k = 16$. Differ from the work in [13]-[16] which focuses on optimizing the random RSD for small message sizes, our method involves a deterministic way for the degree generation and sequential selection of data symbols for the generation of the code symbol. The degree is generated according to:

$$deg(C_n) = \begin{cases} n, & n = 1,2,\cdots,R_p \\ n - m \cdot R_p, & n = R_p + 1,\cdots,k \end{cases} \quad (5)$$

where R_p is the repetition period value chosen to satisfy

$$3 \le R_p \le k/2 \quad (6)$$

and m is an integer denoting the number of the repeated degree group which can have a value from the integer set $1,2,\cdots,k/R_p - 1$. For instance, let $k = 8$ be the number of data symbols, and the data symbols are $u_1, u_2, u_3\ ,\cdots, u_8$, and let $N = 10$, represents the length of the code symbols has to be generated. Hence, using (5) the generator matrix for such LT code with $R_p = 4$ can be constructed as:

$$G = \begin{bmatrix} u_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ u_1 & u_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ u_1 & u_2 & u_3 & 0 & 0 & 0 & 0 & 0 \\ u_1 & u_2 & u_3 & u_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & u_5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & u_5 & u_6 & 0 & 0 \\ 0 & 0 & 0 & 0 & u_5 & u_6 & u_7 & 0 \\ 0 & 0 & 0 & 0 & u_5 & u_6 & u_7 & u_8 \\ 0 & 0 & 0 & 0 & u_5 & u_6 & u_7 & 0 \\ u_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (7)$$

The matrix in (7) has the following properties.

- The degrees for the coded symbol from C_1 to C_k are generated using (5).
- $N-k$ extra code symbols are randomly chosen from those of the previously generated code symbols.
- For each code symbol, the decoder is informed about the formation of code symbol by providing the degree and the repetition value m, R_p.
- Changing $R_p$affects the number of degree one coded symbols as well as the encoding complexity, i.e., average degree.

Sequential encoding method will be generated as in figure below:

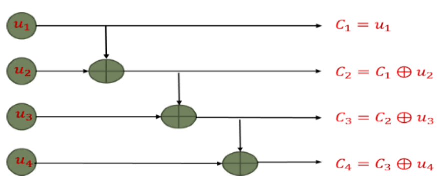

Figure 1. Sequential encoding approach for the first repetition period with $= 8, R_p = 4$.

The encoding structure of the SEA produces successive coded symbols which make the decoding process much easier and less complex.

III. System Model

Fig. 2 shows the scenario used in our simulations. A file of length k will be represented by a set of random binary symbols, $u = (u_1\ u_2\ u_3 \cdots, u_k)$. An LT codeword of length N is produced by multiplying this data stream by the generator matrix G which can be generated using one of the degree distributions listed in Section II. The codeword $C = (C_1\ C_2\ C_3 \cdots C_N\ )$ is modulated using Binary phase shift keying (BPSK) and producing the modulated signal V which will be transmitted through an additive white Gaussian noise (AWGN) channel with noise power σ^2. The received signal equals to $Y = V + n$. The decoder has two sections:

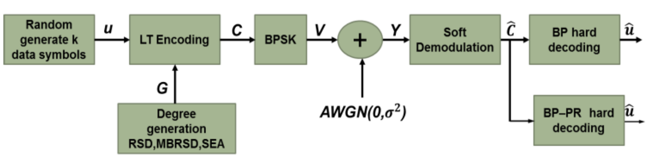

Figure 2. System model for soft-demodulation-hard-decoding approach.

A. Soft Demodulation

Let the received noisy codeword represented by $Y = (y_1, y_2, y_2, \cdots y_n)$, then the probabilities $Pr(C_i = 1\ /y_i)$ and $Pr(C_i = 0\ /y_i)$ can be calculated as:

$$Pr(C_i = 1\ /y_i) = (1 + e^{-\frac{y_i}{\sigma^2}})^{-1} \tag{8}$$

and,

$$Pr(C_i = 0\ /y_i) = 1 - Pr(C_i = 1\ /y_i) \tag{9}$$

Comparing (8) and (9), the value of the codeword bit C_i be estimated based on the maximum likelihood probability.

B. Hard Decoding

The estimated coded symbols will be sent to the belief propagation (BP) hard decoder. Three approaches will be used in this hard decoding, the regular BP [1] and the modified belief propagation with pattern recognition BP-PR [11] will be utilized for the LT codes generated with RSD and MBRSD. While the code with SEA generator matrix will be decoded using sequential decoding (SD) method. The decoder knows the degree for each coded symbol as well as the set of connected data symbols. A brief description for the three decoding approaches will be given in the following sections.

1) Regular BP

The bipartite representation of LT code shown in Fig. 3 illustrates the three main decoding steps. In this figure the gray circles represent unrecovered data symbols while the red rectangulars represent the estimated code values.

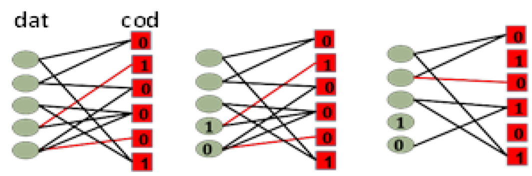

Figure 3. Regular BP decoding steps: (a) copy of degree-one coded symbol (b) update the neighbors (c) release the involved edges.

The decoding steps can be summarized as follows.

Step-1: The decoder finds a code symbol with degree one, (if not found the decoder declares failure), and the corresponding data symbol will be recovered, as in Fig. 3(a)

Step-2: The neighbor coded symbols of the recovered data symbol will be updated using mod-2 addition operation as in Fig. 3(b).

Step-3: All the connections to the recovered data symbol will be released. And a new degree one code symbol has to be searched again to continue recovering the rest of data symbols, Fig. 3(c).

2) BP with pattern recognition (BP-PR)

BP decoding is an iterative process, if there is no degree one coded symbol in the ripple box at any stage of the decoding operation, the decoder halts and the rest of unrecovered data symbols are recorded as error. To continue the decoding operation in the absence of degree-one code symbols BP-PR method in [11] has been proposed. In this method $rs(\boldsymbol{c_i}) = \{u_l, u_s, \cdots\}$ represents the run set of $\boldsymbol{c_i}$ which is the set of data symbols used while generating $\boldsymbol{c_i}$ and for any two coded symbols C_i and C_j if:

$$rs(C_i) \subset rs(C_j) \tag{10}$$

Then by row elimination in an iterative manner we could have either:

- A new degree one coded symbol will be found if:

$$d(C_i) - d(C_j) = 1 \tag{11}$$

- Or if no degree one coded symbol has found then try to reduce the degree for the coded symbol whom satisfying equation (10) i.e.;

$$d(C_i) < d\left(d(C_i) - d(C_j)\right) \tag{12}$$

These extended operations enhance the error performance with a little price of extra symbol operations. The details of this method can be found in [11].

3) Sequential decoding approach

The sequential encoding approach proposed in Section-II enables the successive decoding of the adjacent coded symbols. Assume that the generator matrix in (7) will be used for the encoding operation. Then for the set of estimated symbols $\hat{C} = (\hat{c}_1, \hat{c}_2, \hat{c}_3, \cdots \hat{c}_N)$, the coded symbol which connected to a single data symbol can be used recover its connected data symbol as:

$$u_i = \widehat{c_i} \tag{13}$$

while for any adjacent coded symbols having degree greater than one the decoding can be achieved using:

$$u_j = \hat{c}_j \oplus \hat{c}_{j-1} \quad for \quad j = 2,3,4 \tag{14}$$

So, it requires only $O(k)$ symbol operations to recover k data symbols from any k coded symbols rather than an average of $O(klog(k))$ symbols for the regular BP with $N = k + \varepsilon$ received coded symbols [1].

IV. Simulation Results

In this section, simulation results will be presented to show the decoding efficiency of LT codes using our proposed sequential encoding approach (SEA) with sequential decoding (SD) and compared its performance to that of LT codes employing RSD and MBRSD with BP decoding algorithm. The bit error rate (BER) performances and successful decoding ratio for short length messages with $k = 32$ and 16 symbols have been examined with a wide range of signal-to-noise ratio (SNR) through an AWGN channel. For the RSD and MBRSD the parameters $c = 0.02, \delta = 0.1$ are used while the SEA generates its code using $R_p = k/8$. BER has been recorded for each point on the graph after receiving 100 erroneous frames i.e., the number of processed symbols is not fixed for all SNR points on the graph. The decoder collects $N = 2k$ symbols with $R = k/N$ equal to 0.5 for RSD and MBRSD.

The overhead of LT codes can be as large as 100% when the number of input symbols is small [8]. The comparison allows the LT codes using RSD and MBRSD to get this full assistance, i.e. the decoder will start its mission after receiving a code with length of twice the data length. On the other hand, the proposed sequential decoding method achieves the decoding operation by processing the first k received coded symbols.

As shown in Fig. 4 the BER performance of LT code with SEA using SD and with no overhead symbols, i.e. $R = 1$, outperforms the code generated with RSD and MBRSD with $R = 0.5$ and utilized the conventional BP decoding technique. It is clear that SEA with its sequential decoding (SD) gets the lower BER floor for all the SNR values and starts its low records even with low SNRs.

In order to alleviate the problem of losing degree-one code symbol which frequently occurs with such small message sizes, we apply our treatment for such limitations which presented in [11] and called BP-PR to enhance the performance of RSD and MBRSD. Fig. 5 shows a noticeable enhancement for the error floor of both LT codes with RSD and MBRSD with the same code rate of $R = 0.5$ and supported by the BP-PR. When the decoder collects more decoded symbols for the code with SEA and $R = 0.5$, it is obvious that sequential coding records the best error floor and the extra received coded symbols decreasing the error floor at SNR of 7dB from 10^{-4} to be 4.522×10^{-7} with a clear enhancement for all the SNRs.

In Fig. 6, the successful decoding ratio which is a measure for the number of full recovery frames to the total received frames by the decoder. An average records for transmitting 500 frames of length $k = 32$ and $k = 16$ symbols were used for LT code with SEA using SD and of length $k = 32$ for that with RSD using BP-PR. The decoder allows to collect $N = 2k$ for the RSD case while it starts decoding with $N = k$ for the case of SEA. Even with very low SNR, the SEA code with SD successes in full recovery of about 70% of all the transmitted frames with the extreme short length of $= 16$. furthermore, our approach achieves the full success at SNR equal to 6dB for the case of $k = 16$ and of SNR equal to 7dB for $k = 32$ while the code with RSD using BP-PR requires a signal with SNR equal to 10dB to achieve almost the its full data recovery. The interested point that this SEA with its SD achieves these performances with less overhead and less decoding complexity since it needs only k received coded symbols to recover the k data symbols compared to that of the RSD with BP-PR which needs to receive $N = k + \varepsilon$ to recover the information frame of length k.

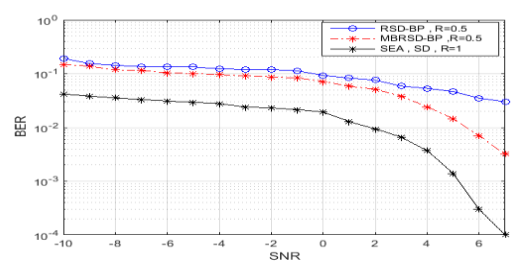

Figure 4. LT code with $k = 32$ for SEA, RSD, and MBRSD using regular BP

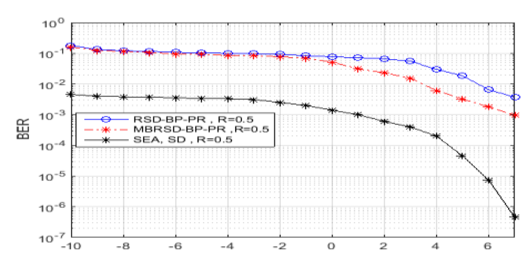

Figure 5. LT code with for SEA, RSD, and MBRSD using BP-PR

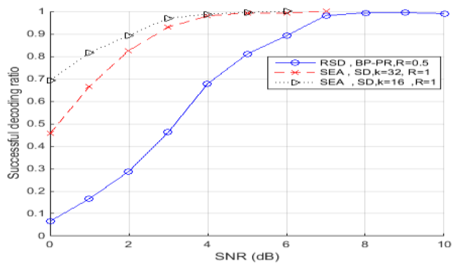

Figure 6. Successful decoding ratio for LT code with RSD using BP-PR for k=32 and LT code with SEA for k=32 and 16 using sequential decoding

V. CONCLUSION

A sequential encoding-decoding method for short length LT codes over an AWGN channel has been proposed. This coding method leads to decrease the error floor and decoding complexity. The decoding technique based on estimating the code values from the noisy received code by a soft demodulation using maximum likelihood probabilities, then a hard BP is used to extract the information data symbols. Three deferent hard decoding techniques has been applied named as (regular BP, BP-PR and sequential decoding). The simulation results show that the proposed method outperforms the LT code generated using RSD or MBRSD. The code has been generated for short length data sizes and R_p has been adjusted to maintain minimum encoding complexity.

REFERENCES

[1] M. Luby, "LT Codes," in Proc. of the 43rd Symposium on Foundations of Computer Science, pp. 271–280, 2002.

[2] A. Shokrollahi, "Raptor codes," IEEE Transactions on Information Theory, vol. 52, no. 6, pp. 2551–2567, 2006.

[3] R. Palanki, J.S. Yedidia, "Rateless codes on noisy channels,"in Proc. of International Symposium on Information Theory, p. 37, 2004.

[4] 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.

[5] T.D. Nguyen, L. L. Y, L. Hanzo, "Systematic Luby transform codes and their soft decoding," in Proc. of IEEE Workshop on Signal Processing Systems. pp. 67-72, 2007.

[6] I. Hussain, I. Land, T. H. Chan, M. Xiao, L. K. Rasmussen, "A new design framework for LT codes over noisy channels," in Proc., of International Symposium on Information Theory, pp. 2162-2166,2014.

[7] S. Chen, Z. Zhang, L. Zhu, K. Wu, X. Chen, "Accumulate rateless codes and their performances over additive white Gaussian noise channel," IET Communications J., vol. 7, no. 4, pp. 372-381,2013.

[8] A. Kharel, L. Cao. "Decoding of short LT codes over BIAWGN channels with Gauss-Jordan elimination-assisted belief propagation method," in Proc. of Wireless Telecommunications Symposium (WTS),pp. 1-6, 2015.

[9] A. Shokrollahi, "LDPC codes: An introduction," Digital Fountain Inc., Tech. Rep,2003.

[10] X. Ma, C. Li, B. Bai, "Maximum likelihood decoding analysis of LT codes over AWGN channels," Proc. of 6th IEEE International Symposium on Turbo Codes and Iterative Information, pp. 285-288, 2010.

[11] N. I. Abdulkhaleq, O. Gazi, "Decoding of LT-Like Codes in the Absence of Degree-One Code Symbols," ETRI J., vol. 38, no. 5, pp. 896-902, 2016.

[12] K. F. Hayajneh, S. Yousefi, M. Valipour, "Improved finite-length Luby-Transform codes in the binary erasure channel," IET Communications J., vol. 9, no. 8, pp. 1122–1130, 2015.

[13] E. Hyytia, T. Tirronen, J. Virtamo, "Optimal degree distribution for LT codes with small message length," in the 26th IEEE INFOCOM, pp. 2576–2580, 2007.

[14] E. A. Bodine, M. K. Cheng, "Characterization of Luby Transform codes with small message size for low-latency decoding," IEEE International Conference in Communications, pp. 1195-1199,2008.

[15] C. M. Chen, Y. P. Chen, T. C. Shen, J. K. Zao, "On the optimization of degree distributions in LT codes with covariance matrix adaptation evolution strategy," IEEE Congress on Evolutionary Computation, pp. 1-8, 2010.

[16] E. Hyytia, T. Tirronen, J. Virtamo, "Optimizing the degree distribution of LT codes with an importance sampling approach", In RESIM, 6th International Workshop on Rare Event Simulation. 2006.