

# Rate adaptive binary erasure quantization with dual fountain codes

Dino Sejdinović, Robert J. Piechocki and Angela Doufexi  
Centre for Communications Research  
Department of Electrical & Electronic Engineering  
University of Bristol, Bristol, UK  
Email: {d.sejdinovic, r.j.piechocki, a.doufexi}@bristol.ac.uk

Mohamed Ismail  
Toshiba Research Europe Ltd  
Telecommunications Research Laboratory  
32 Queen Square, Bristol, UK  
Email: mohamed.ismail@toshiba-trel.com

**Abstract**—In this contribution, duals of fountain codes are introduced and their use for lossy source compression is investigated. It is shown both theoretically and experimentally that the source coding dual of the binary erasure channel coding problem, binary erasure quantization, is solved at a nearly optimal rate with application of duals of LT and raptor codes by a belief propagation-like algorithm which amounts to a graph pruning procedure. Furthermore, this quantizing scheme is rate adaptive, i.e., its rate can be modified on-the-fly in order to adapt to the source distribution, very much like LT and raptor codes are able to adapt their rate to the erasure probability of a channel.

## I. INTRODUCTION

Whereas the use of sparse graph codes equipped with a belief propagation decoding apparatus in the pure channel coding problem is well understood and characterized [1], [2], their assumed potential for lossy source coding has not yet been fully utilized. The structure of the posterior distribution in a lossy compression problem presents substantial challenges for the message-passing sum-product algorithms used in the pure channel coding problem [3], which suggests that novel algorithmic approaches may be required. Advances in the statistical physics community and their work on efficient algorithms for  $k$ -SAT problems [4],[5],[6], namely the survey propagation (SP) algorithm, inspired investigation of SP-based algorithms for lossy source compression with LDGM (low density generator matrix) codes [7], which arise naturally as duals of LDPC (low density parity check) codes. The use of LDGM codes was instigated by their rate-distortion bound saturation for the binary erasure quantization problem, the source coding dual of the binary erasure channel coding problem, which was noted in [8]. Our proposal throughout this paper is to extend a sparse graph lossy source coding framework to the fountain coding paradigm [9], [10], [11] and duals of fountain codes. Fountain codes disregard conventional code modelling with a fixed code rate. Instead, their theory provides the framework for encoder to spray a potentially infinite stream of encoding symbols across the channel by continually generating random equally important descriptions of the source. These codes lend themselves to theoretical study and show high potential for practical realizations in broadcasting applications. The property of ratelessness, i.e., to adapt their code rate on the fly, makes them inherently adjustable to different and varying channel conditions, and thus

an attractive solution for wireless and mobile communication systems. By applying duals of fountain codes instead of duals of LDPC codes for lossy source coding problem, possibility to produce rate adaptive lossy compression schemes seems plausible. A number of communication systems which utilize lossy compression techniques may benefit from such a coding apparatus able to adapt in order to reach desired distortion at a near optimal rate. Our preliminary investigations, presented here, did confirm that a near optimal rate adaptive binary erasure quantization is possible with duals of fountain codes.

## II. DUALITIES IN SPARSE GRAPHS

Both the channel decoding and source encoding, i.e., quantization, are essentially a search process over an exponentially large set of codewords. Thus, we must capitalize on a special structure that facilitates channel decoding and source encoding operations if we are to devise practical corresponding techniques. A conceptually simple rationale motivates as to study sparse graphical modelling of codes, which may be able to reduce computational complexity associated with these encoding/decoding methods. The code structure of a code  $\mathcal{C}$  is conveniently captured by a bipartite graph  $\mathcal{G}_H$  based on its parity check matrix  $H$ . It is a graph with two classes of nodes, one of them representing projections of codeword  $\mathbf{c} \in \mathcal{C}$  to its coordinates, i.e., the bits of the codeword, and the other one representing parity check equations that codewords satisfy. This is a special type of a factor graph [12], which is commonly used in mathematics to represent a factorization of a function such that each factor is dependant only on the subset of the set of possible function arguments. In that way, the computation of the factorized function may be performed more efficiently by exploiting the distributive law. Alternatively, code structure may also be captured by another bipartite graph  $\mathcal{G}_G$ , based on the code's generator matrix  $G$ . This is a direct representation of the specific mapping

$$\phi : \mathbb{F}_2^k \rightarrow \mathbb{F}_2^n, \quad (1)$$

which describes an actual encoding procedure, i.e., an isomorphic assignment of a codeword  $\mathbf{c} = \phi(\mathbf{x}) \in \mathcal{C}$  to the message block - vector  $\mathbf{x} \in \mathbb{F}_2^k$  - which is assumed to represent information content. Two classes of nodes in a  $G$ -representation are (input) bits of the message block and

(coded) bits of the codeword. Each coded bit is a linear combination of a number of input bits, and this is thus yet another realization of a factor graph. We take special interest in those graphical models whose graphical representations are sparse and thus their sparsity may be capitalized in the design of efficient decoding algorithms. These are low-density constructions - codes associated with sparse  $H$ -representations are known as low-density parity-check (LDPC) codes and those associated with sparse  $G$ -representations are known as low-density generator-matrix (LDGM) codes.

For each binary linear code  $\mathcal{C}$ , as it is a subspace of vector space  $\mathbb{F}_2^n$ , we may define its dual code  $\mathcal{C}'$  as a dual subspace of  $\mathbb{F}_2^n$ , or

$$\mathcal{C}' = \{\mathbf{y} \in \mathbb{F}_2^n : \mathbf{y} \cdot \mathbf{z} = 0 \forall \mathbf{z} \in \mathcal{C}\}. \quad (2)$$

If  $\mathcal{C}$  has dimension  $k$ , then its dual code  $\mathcal{C}'$  has dimension  $n-k$ . The generator matrix of  $\mathcal{C}$  is equal to the transpose of parity check matrix of  $\mathcal{C}'$  and vice versa. Thus, if  $\mathcal{C}$  allows sparse  $H$ -representation (as an LDPC code), it is clear that its dual code  $\mathcal{C}'$  allows sparse  $G$ -representation (as an LDGM code) and that  $\mathcal{G}_G(\mathcal{C}') = \mathcal{G}_H(\mathcal{C})$ . These graphs are equal, however their interpretation as factor graphs slightly differs. The nodes which are parity check nodes in  $\mathcal{G}_H(\mathcal{C})$  are input nodes in  $\mathcal{G}_G(\mathcal{C}')$ , and thus the role of the factor and variable nodes in a graph is exchanged between two classes of nodes in a factor graph representation of the dual code.

It has been well publicized that source and channel coding are dual problems (cf. [13] and references therein), both in the classical and side information cases. Informally, a source encoder and a channel decoder have a similar role of “removing entropy” in these problems. A channel decoder tries to map a received channel output to a nearest codeword and should succeed with high probability whenever a received channel output is within a certain distance  $D$  of some codeword  $c \in \mathcal{C}$  with respect to some metric on a channel output alphabet. On the other hand, a source encoder *quantizes* the source to a nearest codeword  $c \in \mathcal{C}$  and the distance from source to codeword should with high probability be lower than some prescribed  $D$  for an arbitrary source. In the channel encoding case,  $D$  is the noise induced by the channel and in the source encoding case,  $D$  is the distortion induced by quantizing the source. Thus it is natural to investigate whether channel decoders would produce good source encoders. However, this approach generally fails, and dualizing code structure seems as the next plausible solution [8].

### III. BINARY ERASURE QUANTIZATION AND FOUNTAIN CODES

One of the simplest channel coding problems is data transmission over a binary erasure channel. A uniformly chosen message block  $\mathbf{x} \in \mathbb{F}_2^k$  - a symmetric Bernoulli source - is encoded by a code which eventually leads to transmission of  $n$  coded bits over binary erasure channel of erasure probability  $e$ . This may be achieved either by using a fixed-rate  $(n, k)$ -binary code, or by puncturing a higher-rate or a rateless code (e.g., a

fountain code). Let us describe the channel output alphabet. It is an  $n$ -word over alphabet  $\mathcal{A} = \{0, 1, *\}$  where  $*$  stands for “erasure”. Each output bit  $z$  is a realization of random variable  $Z$  on  $\mathcal{A}$  with  $p_Z(0) = p_Z(1) = (1-e)/2$  and  $p_Z(*) = e$ .

We note that  $\mathcal{C} \subset \mathbb{F}_2^n \subset \mathcal{A}^n$  and we introduce a standard metric on  $\mathcal{A}^n$ , with

$$d(\mathbf{a}, \mathbf{b}) = \left| \left\{ i \in \{1, 2, \dots, n\} : a_i \neq b_i \right\} \right|, \quad \mathbf{a}, \mathbf{b} \in \mathcal{A}^n. \quad (3)$$

Since a binary erasure channel does not introduce any bit-flips, the distance between the channel output and the transmitted codeword is precisely the number of erasures induced by the channel.

LDPC codes may be applied to this problem; good LDPC codes should provide reliable source recovery at a fixed rate slightly below capacity, i.e.,

$$1 - e > R = k/n > 1 - e - \epsilon, \quad (4)$$

in large block lengths  $k$ , where  $\epsilon$  is a small gap to capacity.

The dual quantization problem, binary erasure quantization (BEQ) is the one with source  $Z$  on  $\mathcal{A}$  distributed the same as the output in the channel coding problem. The  $n$ -word should be mapped into a nearest codeword  $\mathbf{y}$  (code used has dimension  $k$ ) with respect to metric  $d$ . If we wish to obtain the minimum distortion of  $D = e$ , the random variable  $Y$  describing the coded bits needs to satisfy

$$p_{Y|Z}(y|z) = \delta(y - z), \quad z \in \{0, 1\}, \quad (5)$$

where  $\delta$  denotes the Dirac  $\delta$ -function. The rate-distortion function can thus be calculated by

$$\begin{aligned} R(e) &= \min_{p_{Y|Z}} I(Y; Z) \\ &= \min_{p_{Y|Z}} \left( 1 - e + e \sum_{y \in \{0, 1\}} p_{Y|Z}(y|*) \log_2 \frac{p_{Y|Z}(y|*)}{p_Y(y)} \right) \\ &= 1 - e, \end{aligned} \quad (6)$$

and is achieved by  $p_{Y|Z}(y|*) = 1/2$ , for  $y \in \{0, 1\}$ . A good “quantizer” in a similar sense as for the channel codes would be the one that compresses slightly above rate-distortion function, i.e.,

$$1 - e < R = k/n < 1 - e + \epsilon, \quad (7)$$

in large block lengths  $k$ .

One may quickly realize that the attempt to use the same code structure as for the channel decoding generally fails. The codewords should be within distance  $e$  from any source and the source will have  $e$  erasures on average, with no guarantee that the non-erased part of the source constitutes part of the valid codeword. This guarantee can be achieved only through making the factor graph more dense, i.e., by guaranteeing that each parity check node would be connected to at least one erased symbol. This implies that the lower bound on the average degree of parity node check degrees would in fact be logarithmic in block length by a simple “balls-in-bins” argument. This argument supported the claim in [8] that LDPC codes are generally bad quantizers and

introduced the use of their duals, LDGM codes, for the source quantization problems. The authors of [8] have shown that duals of capacity approaching LDPC channel codes for the BEC yield minimum rate approaching LDGM codes for the BEQ problem. Although the practical significance of the BEQ problem is questionable, this work provided two important insights into the area, that (1) graphical models may yield near optimal codes in lossy compression and (2) there may exist efficient iterative decoding algorithms related to belief propagation (sum-product) for other, more practicable, quantization problems.

At this point we should mention an existing analogy with the logarithmic lower bound on the average degree of output symbols in LT codes [10], derived also by the same “balls-in-bins” argument. However, although the logarithmic lower bound is *required* for a reliable LT code, it was shown in [10] that there exist output node degree distributions which meet this lower bound and also approach capacity in large block-lengths at any erasure probability. These distributions were named Robust Soliton distributions. Furthermore, in [11], LT codes were modified by precoding which allowed relaxing the lower bound on the average degree of output symbols: raptor codes were born, today’s state-of-the-art rateless channel codes with a linear encoding/decoding complexity and excellent performance in practice. Thus, dismissal of LDPC codes for source quantization problems may be premature. Namely, LDPC codes would arise naturally as duals of “truncated” LT codes. Indeed, by setting the distribution of bit nodes in an LDPC code for erasure quantization to a Robust Soliton distribution we can guarantee that each parity check node is connected to at least one bit node carrying an erased symbol  $*$  and thus the non-erased coded bits constitute a valid part of some codeword. The next step would be to translate the graph pruning procedure arising from the belief propagation algorithm for LT codes over erasure channels to this new dual setting. Furthermore, we will show that this is possible and in fact the dual algorithm is of the same complexity and fails/succeeds together with the original graph pruning procedure. Moreover, one can fix the number of checks (like fixing the block length in fountain codes) and attempt quantization on a source of increasing length (like an output stream of increasing length) at equal intervals - larger the source, larger the compression rate - and once the required rate is reached, quantization succeeds.

The rationale lies within the duality of LDPC and LDGM codes we have discussed above. The decoding graph of a robust soliton LT code is its sparse  $G$ -representation with a variable number of coded bit nodes. For a fixed erasure probability  $e$ , the number  $n = n(e)$  of coded bit nodes necessary for successful decoding is slightly above the optimal  $k/(1 - e)$ . The decoding graph of its dual is an LDPC decoding graph with a variable number of bit nodes. The rate duality implies that the BEQ source should not be the same as the BEC output, but the dual code applies to the distortion  $e' = 1 - e$ . This is just another way of stating the noise/distortion parameters at which the pair of channel code

and its dual quantization code are both good. We have

$$1 - e > k/n > 1 - e - \epsilon \equiv 1 - e' < (n - k)/n < 1 - e' + \epsilon. \quad (8)$$

Thus, these “dual” LT codes can work on any source  $Z$  with variable  $e'$  and approach the rate-distortion function in high block lengths with quantization time that scales as  $O(k \log k)$ . Furthermore, one may apply a raptor-like scenario and introduce precoding to remove the error floor when a constant average degree Soliton-like distribution is used. Precoding simply increases the number of erased symbols - as it would increase the number of known relations within our input symbols in a raptor code.

#### A. Algorithms for LT decoding over a BEC and dual LT encoding for BEQ

Both LT decoding over a BEC and dual LT encoding for BEQ are simple graph pruning procedures and can be implemented such that a number of operations scales linearly with the number of edges on the decoding graph. Furthermore, when failure occurs, neither the LT decoder nor the dual LT encoder needs to start over from the full decoding graph with a larger number of coded bit or source nodes, but may just embed the additional nodes into the pruned version of the graph. If a Robust Soliton distribution is used, the computational cost amounts to  $O(k \log k)$  and if a constant average degree output symbol degree distribution is used, as in raptor-like scenarios, the computational cost is  $O(k)$ . In [8], it was shown that the algorithms for LDPC decoding over BEC( $e$ ) and LDGM encoding for BEQ( $1 - e$ ) concurrently fail or succeed when the codes used for both problems are dual. The same clearly holds for the algorithms 1 and 2 presented here.

---

#### Algorithm 1 LT decoding over BEC( $e$ )

---

Input: channel output  $\mathbf{z} \in \mathcal{A}^n$ , decoding graph  $\mathcal{G}_{GLT[1:n]}$  representing the first  $n$  columns of the LT generator matrix.  
Output: information sequence  $\mathbf{x} \in \{0, 1\}^k$

```

Remove erased output nodes from the graph. Assign an all-erasures vector  $\mathbf{x}$  to input nodes.
while  $\mathbf{x}$  has at least one erased sample do
  Find unerased output node  $a$  connected to exactly one erased input node  $i$ .
  if there is no such output node then
    return FAILURE.
  else
    set  $x_i = z_a$ , and add  $x_i$  to all unerased outputs  $z_b$ , with  $b \neq a$  connected to  $i$ .
  end if
end while
return  $\mathbf{x}$ 

```

---

#### B. Asymptotic rates

In [10], Luby presented properties of the Robust Soliton distribution. He proved that the number of received (unerased)

---

**Algorithm 2** Dual LT encoding for BEQ( $e$ )

---

Input: source  $\mathbf{z} \in \mathcal{A}^n$ , decoding graph  $\mathcal{G}_{H=G_{LT}[1:n]^T}$ .  
Output: codeword  $\mathbf{y} \in \{0, 1\}^n$ .

Remove unerased bit nodes from the graph. Assign a vector  $\mathbf{x}$  to check nodes, obtained as a summation of the neighbouring unerased bit variables.

**while**  $\mathbf{x}$  has at least one unerased sample **do**

Find erased source node  $i$  connected to exactly one unerased check node  $a$

**if** there is no such source node **then**

**return** FAILURE.

**else**

reserve source variable  $z_i$  to later satisfy check  $x_a$ , and erase check  $x_a$ .

**end if**

**end while**

Set unreserved source variables to an arbitrary binary sequence. Work backward through reserved source variables and set them to satisfy corresponding checks.

**return**  $\mathbf{z}$

---

coded symbols sufficient to decode an information sequence of length  $k$  is:

$$k' = k + O(\sqrt{k} \ln^2(k/\delta)), \quad (9)$$

where  $\delta$  is allowed failure probability of the decoder. This means that the achieved code rate when transmitting over a BEC( $e$ ) is at least:

$$\frac{k}{n} = \frac{k(1-e)}{k'} = \frac{k(1-e)}{k + O(\sqrt{k} \ln^2(k/\delta))}, \quad (10)$$

which approaches capacity  $1 - e$ , when  $k$  grows large. Algorithms 1 and 2 both fail or succeed together on the same realization of the factor graph with exchanged erased and unerased output (source) symbols. Thus, the achieved compression rate for the BEQ( $e' = 1 - e$ ) problem is at most

$$\frac{n-k}{n} = \frac{k(1-e') + O(\sqrt{k} \ln^2(k/\delta))}{k + O(\sqrt{k} \ln^2(k/\delta))}, \quad (11)$$

which also approaches the optimal rate of  $1 - e'$ . This way, we have proved that dual LT codes with a Robust Soliton distribution at source nodes approach the optimal rate for any  $e'$ .

### C. Raptor-like scenario

In order to allow for degree distributions with constant (independent of  $k$ ) average degree, one may introduce some kind of precoding, as in raptor codes. This is done through introducing additional parity checks, as well as additional “deterministic source nodes” all equal to a  $*$ -value, to the decoding graph. The outer LDPC code as a precode for a raptor code in the dual version becomes an outer LDGM code. Thus, the precode-portion of the decoding graph corresponds to the dual code of the precode of a raptor code. Note that there is a simple interpretation of this procedure in terms of the

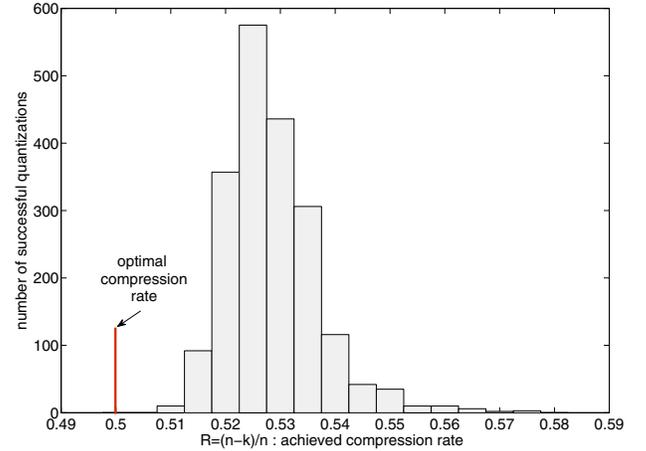


Fig. 1. Histogram showing the achieved compression rate of the dual LT code binary erasure quantization.

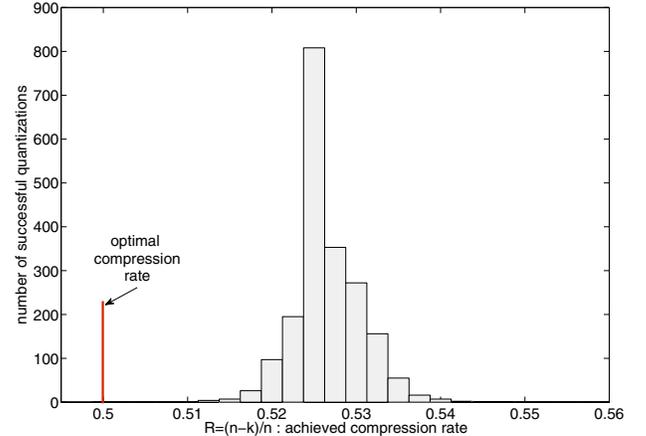


Fig. 2. Histogram showing the achieved compression rate of the dual raptor code binary erasure quantization.

forementioned arguments on the lower bound on the average degree - additional source nodes deterministically set to  $*$ -value instigate higher probability that every check is connected to at least one erased source symbol.

## IV. SIMULATION RESULTS

We have performed quantization of source  $X$  on  $\mathcal{A}$ , distributed with  $e = 0.5$ , with the dual LT code with the number of checks  $k = 10000$ . The length and rate of the dual LT code were variable, starting from the optimal  $n_{min} = 20000$  and were increased at equal steps of  $\Delta n = 100$  up to the length (rate) where quantization was successful, i.e., all the checks were satisfied. Fig. 1 shows the histogram of the achieved compression rate. We used a dual LT code with robust soliton source symbol degree distribution  $\Omega_{RS}(k = 10000, 0.03, 0.5)$ , and performed 2000 trials. Average achieved rate was  $R_{AV} = 0.5285$ .

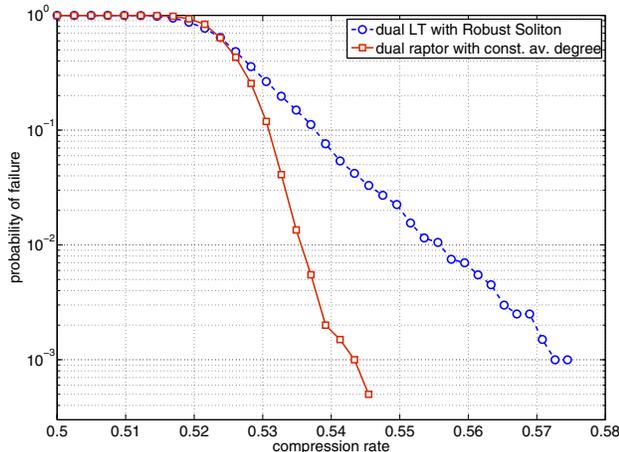


Fig. 3. Quantization failure probability as it decreases with the increase of compression rate.

Also, we have performed quantization of the same source  $X$ , with dual raptor code. Starting number of checks was  $k = 10000$ , and precode was the dual of an LDPC/Half hybrid high rate code [14]. Source symbol degree distribution was set to a constant average distribution

$$\begin{aligned} \Omega(x) = & 0.007969x + 0.493570x^2 + 0.166220x^3 + \\ & 0.072646x^4 + 0.082558x^5 + 0.056058x^8 + \\ & 0.037229x^9 + 0.055590x^{19} + 0.025023x^{64} + \\ & 0.003135x^{66}, \end{aligned} \quad (12)$$

used for generation of LT encoded symbols in raptor codes [11].

Fig. 2 shows the histogram of the achieved compression rates using the dual raptor code. Although the average achieved rate  $R_{AV} = 0.5266$  was only slightly better than that of dual LT codes, use of dual raptor codes does lead to two major advantages. Namely, as was discussed above, computational cost of dual raptor quantization is significantly lower than that of dual robust soliton LT quantization as quantization complexity is linear in  $k$ . Furthermore, the probability of quantization failure diminishes much faster. The latter is illustrated in Fig. 3.

## V. CONCLUSION

By exploiting the concept of ratelessness, fountain codes have initiated great enthusiasm in the channel coding community and shifted the channel coding paradigm over the last few years. At the time when sparse graph codes are sought to be employed for lossy source coding and algorithmic obstacles for doing so are being confronted, the question of applying ideas of fountain codes in order to produce rate adaptive lossy source coding methods arises naturally. Whereas the majority of work investigating lossy source coding and sparse graph codes to date was focused on LDGM codes, as duals of good LDPC codes, this contribution argues that by dualizing

fountain codes good LDPC codes for lossy source coding may be constructed as well. Furthermore, these dual fountain codes exhibit the sought after property to adapt the rate on the fly.

## REFERENCES

- [1] T.J. Richardson, R. Urbanke, "The Capacity of Low-Density Parity-Check Codes Under Message-Passing Decoding", *IEEE Trans. Info. Theory*, vol. 47, No. 2, 2001.
- [2] T.J. Richardson, A. Shokrollahi, R. Urbanke, "Design of Capacity-Approaching Irregular Low-Density Parity-Check Codes", *IEEE Trans. Info. Theory*, vol. 47, No. 2, 2001.
- [3] M.J. Wainwright, "Sparse graph codes for side information and binning", *IEEE Signal Processing Mag.*, vol. 24, no. 7, pp.47–57, Sept. 2007.
- [4] M. Mézard, R. Zecchina, "Random  $k$ -satisfiability: from an analytic solution to an efficient algorithm", *Phys. Rev. E*, vol. 66, no. 1, 2002, pp. 2001–2027.
- [5] M. Mézard, G. Parisi, R. Zecchina, "Analytic and algorithmic solution of random satisfiability problems", *Science*, vol. 297, 812, 2002.
- [6] A. Braunstein, M. Mézard, R. Zecchina, "Survey Propagation: an algorithm for satisfiability", *Random Structures and Algorithms* 27, 201–226, 2005.
- [7] M.J. Wainwright, E. Maneva, "Lossy source encoding via message-passing and decimation over generalized codewords of LDGM codes", in *Proc. of IEEE International Symposium on Information Theory*, pp.1439–1497, Adelaide, Australia, 2005.
- [8] E. Martinian, J. S. Yedidia, "Iterative quantization using codes on graphs", In *Proc. Allerton Conf. Comm., Control, and Computing*, Monticello, IL, USA, Oct. 2003.
- [9] J. Byers, M. Luby, M. Mitzenmacher and A. Rege, "A digital fountain approach to reliable distribution of bulk data," *Proc. of the IEEE SIGCOMM*, 1998.
- [10] M. Luby, "LT Codes," *Proc. of the 43rd Annual IEEE Symp. Foundations of Computer Science (FOCS)*, Vancouver, Canada, Nov. 2002.
- [11] A. Shokrollahi, "Raptor Codes," *IEEE Trans. Info. Theory*, vol. 52, No. 6, pp. 2551–2567, 2006.
- [12] F.R. Kschischang, B.J. Frey, H.A. Loeliger, "Factor Graphs and the Sum-Product Algorithm", *IEEE Trans. Info. Theory*, vol. 47, No. 2, 2001.
- [13] R.J. Barron, B. Chen, G.W. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Info. Theory*, vol. 49, No. 5, pp. 1159–1180, 2003.
- [14] 3GPP TS 26.346 V7.0.0. Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service; Protocols and Codecs, Dec. 2005.