Sejdinovic, D., Piechocki, R. J., Doufexi, A., & Ismail, M. R. (2008). Rate adaptive binary erasure quantization with dual fountain codes. In Globecom 2008, New Orleans. (pp. 1 - 5). Institute of Electrical and Electronics Engineers (IEEE). 10.1109/GLOCOM.2008.ECP.238

RATE ADAPTIVE BINARY ERASURE QUANTIZATION WITH DUAL FOUNTAIN CODES

Dino Sejdinović[1], Robert J. Piechocki[1], Angela Doufexi[1] and Mohamed Ismail[2]

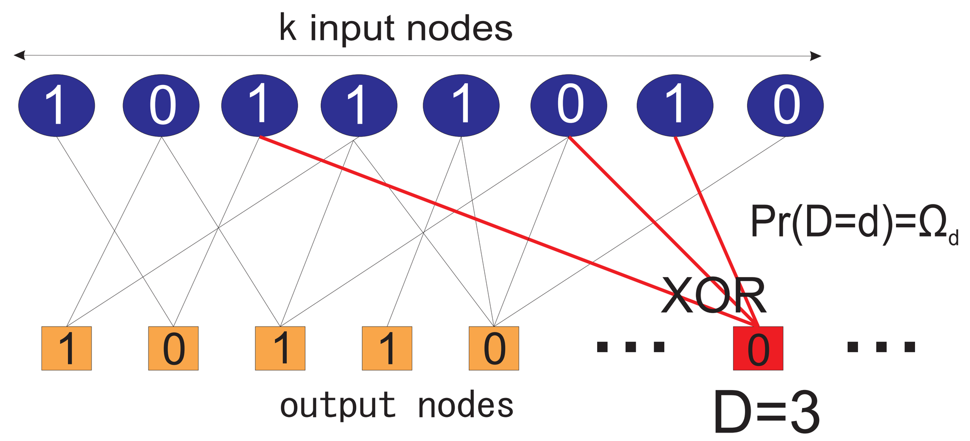
[1]Centre for Communications Research, Dept. of Electrical & Electronic Engineering, University of Bristol, Bristol, UK, Email: {d.sejdinovic, r.j.piechocki, a.doufexi}@bristol.ac.uk

[2]Toshiba Research Europe Ltd, Telecommunications Research Laboratory, Bristol, UK, Email: mohamed@toshiba-trel.com

Contribution: We formulate an asymptotically optimal rate adaptive quantizing scheme for the source coding dual of the binary erasure channel coding problem, namely the problem of binary erasure quantization. Our solution is based on dual fountain codes, i.e., dual LT (Luby Transform) codes and dual Raptor codes. This illustrates that good LDPC (Low Density Parity Check) codes for lossy source compression can be constructed, whereas majority of work so far was focused on LDGM (Low Density Generator Matrix) codes.

Fountain codes

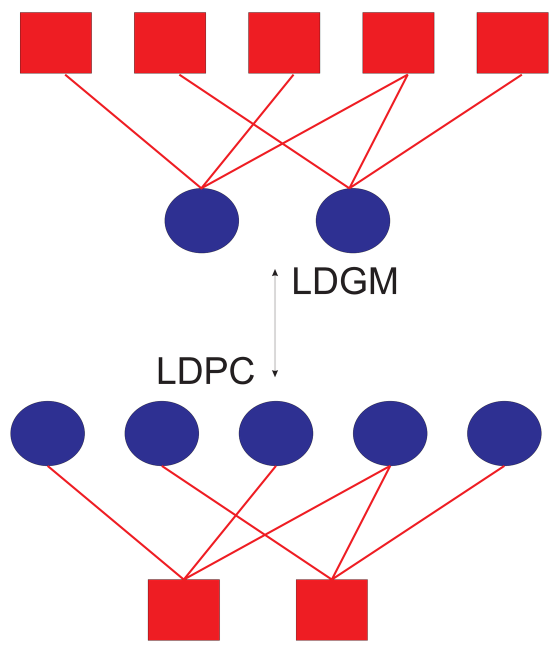
- Fountain codes [1], [2] are an attractive capacity approaching low complexity forward error correction solution for multicasting / broadcasting data over erasure channels
- LT codes [1] are a simple form of fountain codes, whose performance depends on the choice of output symbol degree distribution $\Omega(x)$; capacity achieving LT codes require computational encoding/decoding complexity of $O(k \log k)$ - Robust Soliton distribution.
- Raptor codes [2] are a concatenation of an LT code to a high-rate LDPC precode. Precode enables reduced computational complexity of $O(k)$, by allowing the use of output symbol degree distributions of constant average.



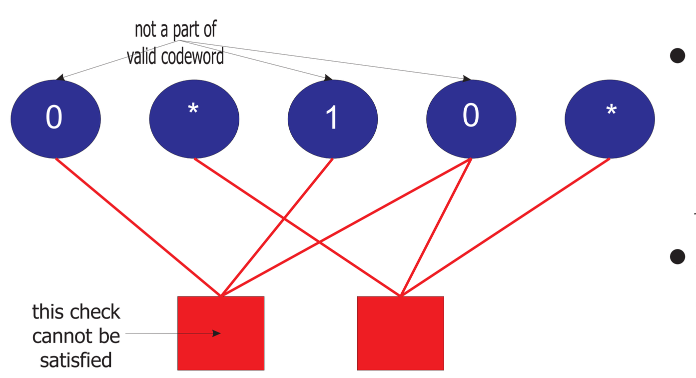
Generation of output symbols of an LT code

Dual codes

- Linear (n, k) code $\mathcal{C}$ defined by its generator matrix $G_{n\times k}$ is
$$\mathcal{C} = \{Gx^{\mathrm{T}} : x \in \mathbf{F}_2^k\} \quad (1)$$
- Its dual code $\mathcal{C}'$ is a linear $(n, n-k)$ code with parity check matrix G^{T}, i.e.,
$$\mathcal{C}' = \{y \in \mathbf{F}_2^n : G^{\mathrm{T}}y^{\mathrm{T}} = 0\}. \quad (2)$$
- Dual code of an LDGM code is an LDPC code and vice versa.



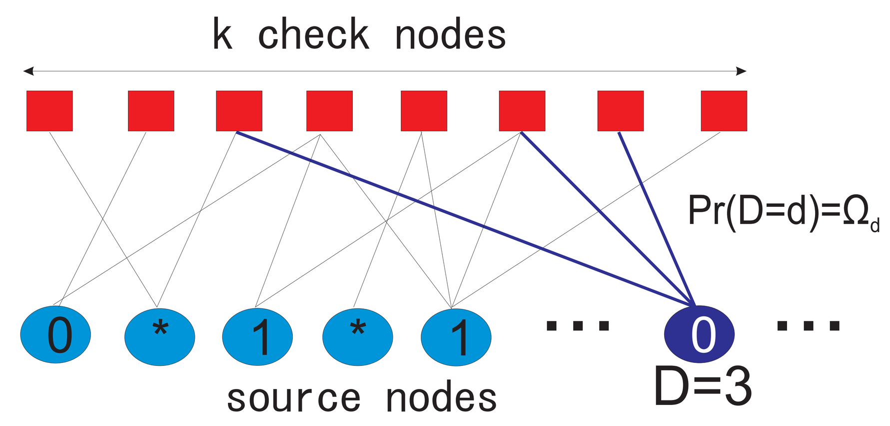
Binary erasure quantization (BEQ)



- BEQ is a source coding dual of binary erasure channel (BEC) coding, with source as a random variable Z on $\mathcal{A} = \{0, 1, *\}$, where "*" signifies *erasure*. with $p_Z(0) = p_Z(1) = (1 - e)/2$ and $p_Z(*) = e$.
- LDPC codes are bad quantizers (the curse of the coupon collector) [3] since degree of check nodes needs to grow logarithmically (computationally expensive).
- Duals of capacity approaching LDPC codes for BEC, LDGM codes, approach optimal compression rate for BEQ [3].
- Papers on LDGM codes for lossy source compression (cf. [4]) cite this result as motivation for such choice of coding scheme.

Dual LT code construction:

- Fix the number of check equations k.
- For each source symbol z_i, choose its degree $d_i \in \{1, 2, \ldots, k\}$ according to the Robust Soliton distribution and require that z_i satisfies d_i uniformly chosen check equations.
- The procedure is incremental: the number n of source symbols is determined on-the-fly.
- Resulting codewords are determined by parity check matrix G_{LT}^{T} with column weight distributed according to a Robust Soliton distribution. Compression rate $R = \frac{n-k}{n}$.



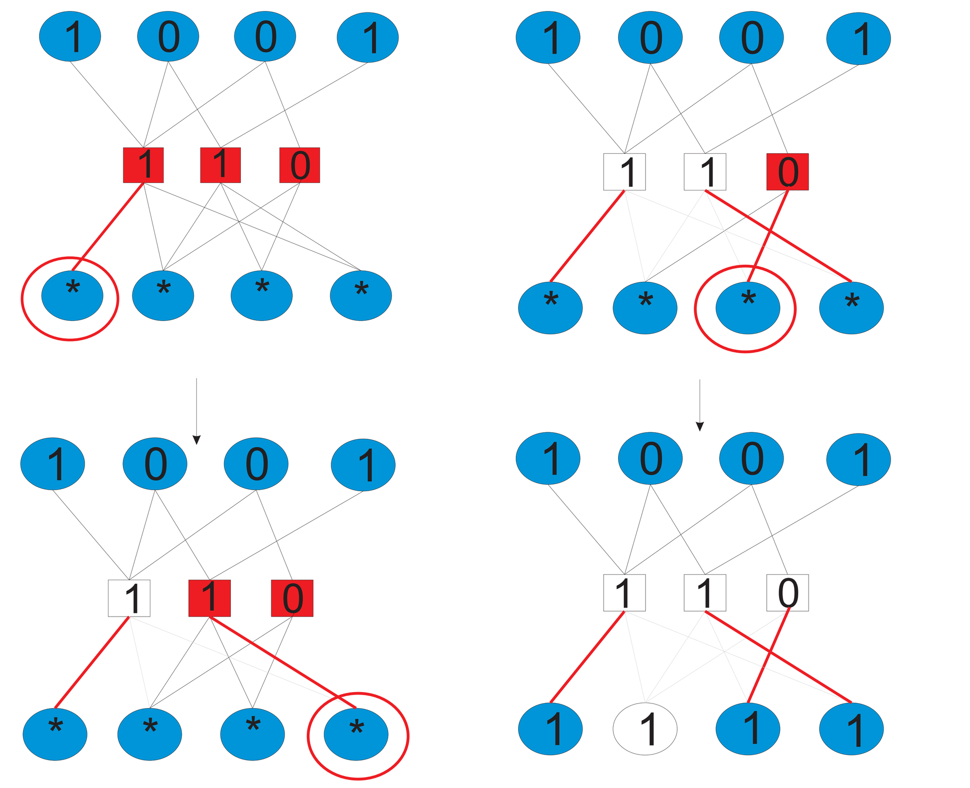
Formation of parity check relationships in a dual LT code

Dual LT quantization algorithm for BEQ(e)

Input: source $\mathbf{z} \in \mathcal{A}^n$, parity check matrix $G_{LT}[1:n]^{\mathbf{T}}$, **Output**: codeword $\mathbf{y} \in \{0,1\}^n$.

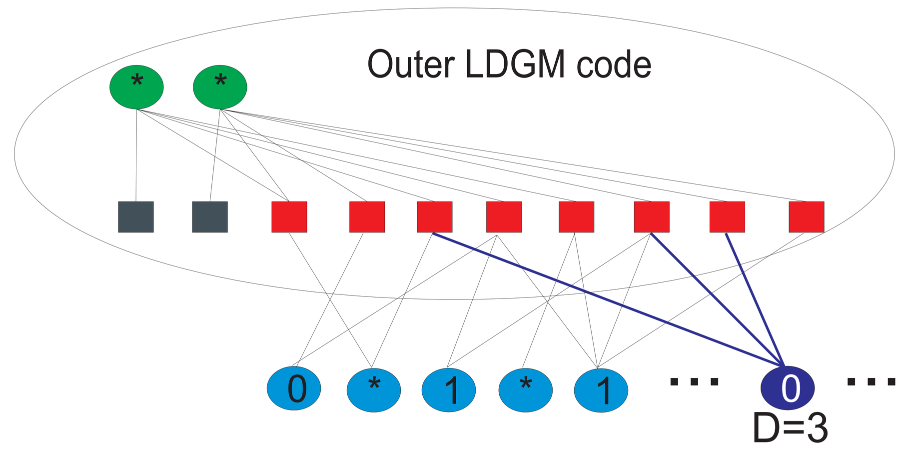
Assign a vector $\mathbf{x}$ to check nodes, with $x_a = \sum_{i\in\mathcal{N}(a), z_i \neq *} z_i$. Set $y_i = z_i$, where $z_i \neq *$.
while there is at least one unmarked check node **do**
  Find erased source node i connected to exactly one unmarked check node a
  if there is no such source node **then**
    return FAILURE
  else
    reserve source node i for check x_a, and mark check node a as satisfied.
  end if
end while
Set codeword bits $\{y_j : j \text{ unreserved}\}$ to an arbitrary binary sequence. Assign values to the codeword bits $\{y_j : j \text{ reserved}\}$ working backwards through the list of reserved nodes, such that they satisfy corresponding checks.
return $\mathbf{y}$

- **Lemma 1**. Dual LT quantization for BEQ source $\mathbf{z} \in \mathcal{A}^n$ with erasure pattern $\mathcal{E}$ succeeds if and only if LT decoding algorithm *on the same graph* succeeds for BEC output $\mathbf{z}' \in \mathcal{A}^n$ with erasure pattern $\mathcal{E}' = \{1, 2, \ldots, n\} - \mathcal{E}$.
- **Corollary 1**. Dual ensemble of an ensemble of capacity-approaching LT codes approaches the optimal compression rate for BEQ(e), independently of e.



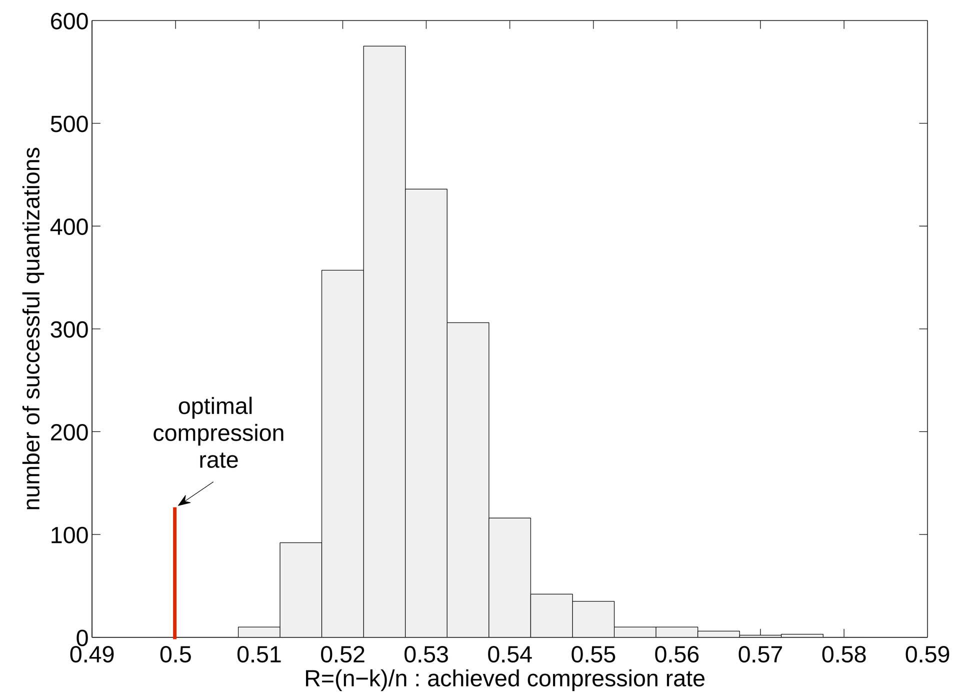
Raptor-like scenario

- Combining Lemma 1 and results of Luby [1], Dual Robust Soliton distributed LT codes approach optimal compression rate when number of parity checks k grows large with computational cost of $O(k \log k)$, for any erasure rate in the source.
- Dualizing Raptor codes [2] allows compression schemes with computational cost of $O(k)$ and the same encoding algorithm. Outer high-rate LDPC code becomes an outer low-rate LDGM code.

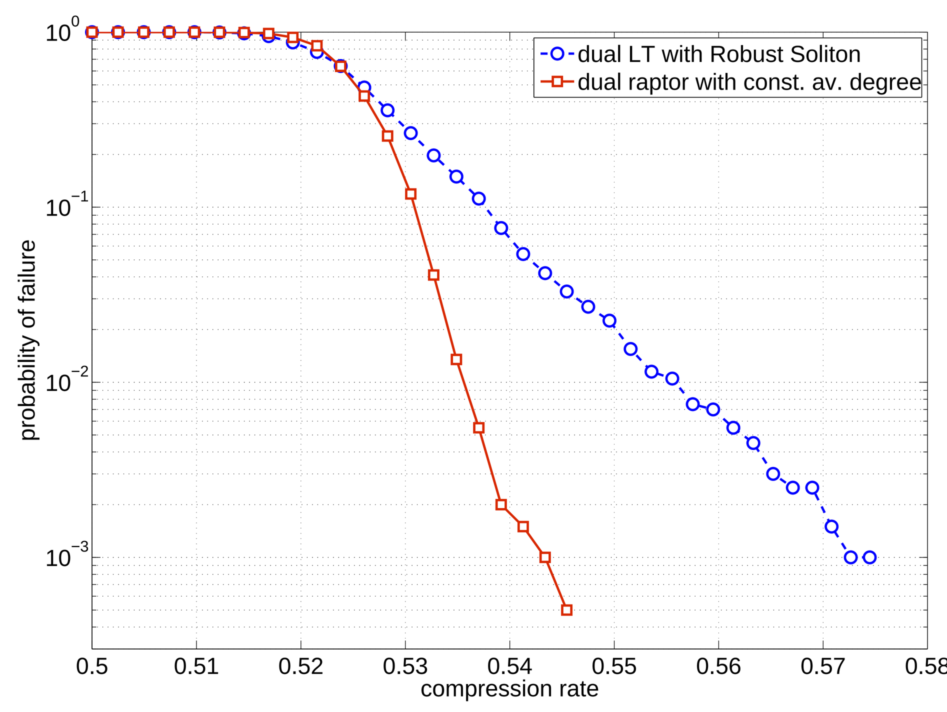


Conclusions:

- The majority of work on lossy compression and sparse graph codes is focused on LDGM codes as duals of good LDPC codes, and this focus was instigated by results in [3], based on LDGM codes as optimal solution for BEQ problem.
- By dualizing fountain (LT and Raptor) codes, we show how optimal rate approaching LDPC codes for BEQ are constructed.
- The newly constructed codes are able to adapt the rate on-the-fly and thereby approach optimal compression rate for BEQ(e), independently of e.



Histogram showing the achieved compression rate of the dual LT code binary erasure quantization, $k = 10^4$



Quantization failure probability versus compression rate

References:

1. M. Luby, "LT Codes," *Proc. of the 43rd Annual IEEE Symp. Foundations of Computer Science (FOCS)*, pp. 271–282, Vancouver, Canada, November 2002.
2. A. Shokrollahi, "Raptor Codes," *IEEE Trans. on Information Theory*, vol. 52, No. 6, pp. 2551–2567, June 2006.
3. E. Martinian, J. S. Yedidia, "Iterative quantization using codes on graphs", In *Proc. Allerton Conf. Comm., Control, and Computing*, Monticello, IL, USA, Oct. 2003.
4. 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.