# Rate Distortion Bounds for Binary Erasure Source Using Sparse Graph Codes

Grégory Demay[1], Vishwambhar Rathi[1,2], Lars K. Rasmussen[1,2]
{gdem, vish, lkra}@kth.se
[1] School of Electrical Engineering, [2] KTH Linnaeus Center ACCESS,
KTH-Royal Institute of Technology, Stockholm, Sweden

**Abstract**

We consider lower bounds on the rate-distortion performance for the binary erasure source (BES) introduced by Martinian and Yedidia, using sparse graph codes for compression. Our approach follows that of Kudekar and Urbanke, where lower bounds on the rate distortion performance of low-density generator matrix (LDGM) codes for the binary symmetric source (BSS) are derived. They introduced two methods for deriving lower bounds, namely the counting method and the test channel method. Based on numerical results they observed that the two methods lead to the same bound. We generalize these two methods for the BES and prove that indeed both methods lead to identical rate-distortion bounds for the BES and hence, also for the BSS.

## I. Introduction

Following the remarkable success of sparse graph codes for the channel coding problem, a natural progression is to explore the capabilities of such codes for the source coding problem. One of the first contributions in this direction was made in [1], where Martinian and Yedidia introduced the binary erasure source (BES), and showed that duals of good sparse-graph channel codes for the binary erasure channel (BEC) are good sparse-graph compression codes for the BES.

Ciliberti, Mezard, and Zecchina used the statistical-physics-based replica method to show in [2] that a low-density generator matrix (LDGM) code with a Poisson generator degree distribution can achieve the Shannon rate-distortion function of the binary symmetric source (BSS) as the average degree increases. Based on this method, they also designed a message-passing encoding algorithm termed *Survey Propagation* (SP). It was later shown by Wainwright and Maneva [3], and independently by Braunstein and Zecchina [4], that in the context of sparse-graph code compression using decimation over LDGM codes the SP algorithm can be interpreted as a special case of the Belief Propagation (BP) algorithm. More recently Filler and Friedrich proposed a decimation-based BP algorithm, termed bias propagation, that can also perform close to Shannon's rate distortion bound using optimized degree distributions for LDGM codes [5]. A compound sparse-graph code construction was proposed by Martinian and Wainwright in [6], [7], [8], where desirable features of LDPC codes and LDGM codes were combined. They further showed that a randomly chosen code from such ensembles under optimal encoding and decoding achieves the rate-distortion bound with high probability.

Another interesting approach to code construction for lossy compression is based on polar codes introduced by Arikan [9]. Polar codes are based on a deterministic code construction that achieves the channel capacity. It was subsequently shown by Korada and Urbanke that polar codes are also optimal for various lossy compression problems

1068-0314/10 $26.00 © 2010 IEEE
DOI 10.1109/DCC.2010.95

including those for the BES and the BSS [10]. In terms of implementation, however, the encoding and decoding complexities of polar codes are higher than the corresponding complexities for codes on sparse graphs with iterative message-passing.

The first performance bounds for LDGM-based lossy compression of a BSS were derived by Dimakis, Wainwright, and Ramchandran in [11] for ensembles of codes. In contrast Kudekar and Urbanke derived lower bounds on the rate-distortion performance of *individual* LDGM codes for the BSS [12], using two different methods to compute the rate-distortion bounds; namely the counting method and the test channel method. Based on numerical results it was observed in [12] that the two methods lead to identical bounds.

For the BES, it was shown in [1] that sparse graph codes can achieve the optimal rate only for zero distortion. Furthemore, so far the analysis for lossy compression using sparse-graph codes was mainly focus on the BSS case and there are no known bounds for the BES case. The use of a more general source such as the BES would allow to gain fundamental insight into the behaviour of sparse-graph codes used as lossy compressors. As our main contributions in this paper we derive lower bounds for the rate-distortion performance of a BES using LDGM codes by generalizing the counting method and the test channel method proposed in [12], and subsequently prove that both methods lead to identical bounds. Since a BSS is a special case of a BES with zero erasure probability, we also implicitly show the equality of the two bounds for a BSS.

The remainder of the paper is organized as follows. In Section II we formally state the problem and provide the necessary background and definitions. The lower bound for the BES based on the counting method is derived in Section III, and the derivation of the lower bound based on the test channel method is detailed in Section IV. Finally, we prove the equality of the two bounds in Section V and conclude by an open question in Section VI.

## II. Definitions and Background

The BES($\epsilon$) is a ternary-alphabet source with alphabet set $\mathcal{A} = \{0, 1, \star\}$. The symbol $\star$ denotes the erasure symbol, which can be encoded into any of the other two symbols at no cost in terms of distortion. Let $S = \{S_1, \cdots, S_m\}$, $S \in \mathcal{A}^m$ be a random source string, where $\mathbb{P}\{S_i = \star\} = \epsilon$, $\mathbb{P}\{S_i = 0\} = \mathbb{P}\{S_i = 1\} = \frac{1-\epsilon}{2}$, $i \in \{1, \cdots, m\}$. We denote the set of all source sequences of length $m$ by $\mathcal{S}$. The number of erasures in a sequence $s \in \mathcal{S}$ is denoted by $H_E(s)$. We use the notation $\mathcal{S}_b$ to denote the set of all source sequences having $b$ erasures, i.e., $\mathcal{S}_b = \{s \in \mathcal{S},\ \text{s.t. } H_E(s) = b\}$. We are interested in the lossy compression of a BES($\epsilon$) using a LDGM code. Thus, the reconstruction alphabet is $\hat{\mathcal{A}} = \{0, 1\}$.

Let $a \in \mathcal{A}$ and $\hat{a} \in \hat{\mathcal{A}}$. Then the distortion between $a$ and $\hat{a}$ is given by,

$$d(a, \hat{a}) = \begin{cases} 0, & \text{if } a = \star \text{ or } a = \hat{a}, \\ 1, & \text{otherwise.} \end{cases} \tag{1}$$

The Shannon rate-distortion function $R^{\text{sh}}_{\text{E}}(D)$ for a BES($\epsilon$) can easily be shown to be

$$R^{\text{sh}}_{E}(D) = \begin{cases} (1-\epsilon)\left[1 - h\left(\frac{D}{1-\epsilon}\right)\right], & \text{if} \quad D < \frac{1-\epsilon}{2}, \\ 0, & \text{if} \quad D \geq \frac{1-\epsilon}{2}, \end{cases} \tag{2}$$

where $h(\cdot)$ is the binary entropy function. In order to compress a source sequence $s \in \mathcal{S}$, it is mapped to one of the $2^{mR}$ index words $w \in \mathcal{W} = \mathbb{F}_2^{mR}$. Let $f : s \mapsto w$ be this encoding map. Further let $\hat{s}$ denote the reconstructed word associated to $w$: $\hat{s} = wG$, where $G$ is the generator matrix of the LDGM code of interest. We denote the set of all $2^{mR}$ reconstruction sequences by $\hat{S} = \{wG : \forall w \in \mathbb{F}_2^{mR}\}$. Let $g : w \mapsto \hat{s}$ be this decoding map $\hat{s} = wG$. The distortion between $s$ and $\hat{s}$ is given by $d(s, \hat{s}) = \sum_{i=1}^{m} d(s_i, \hat{s}_i)$. As in [12], we call the $mR$ components of the index word the generators, $w = \{w_1, \cdots, w_{mR}\}$. The nodes corresponding to generators in the factor graph representing the LDGM code are called the generator nodes. We denote the fraction of generator nodes with degree $i$ by $L_i$.

The average normalized distortion is $\frac{1}{m}\mathbb{E}\left[d(S, g(f(S)))\right]$, where the average is over all source sequence realizations. Let $\mathcal{C}_b(D)$ be the set of source sequences having $b$ erasures and having an overall normalized distortion less or equal to $D$ with respect to some codeword. Similarly, $\forall \hat{s} \in \hat{\mathcal{S}}$, $\mathcal{B}_b(\hat{s}, D)$ is the set of source sequences having $b$ erasures and having a normalized distortion less than or equal to $D$ from the codeword $\hat{s}$. Then, $\mathcal{C}_b(D) = \cup_{\hat{s} \in \hat{S}} \mathcal{B}_b(\hat{s}, D)$.

We are interested in finding lower bounds on the rate and average normalized distortion trade-off for the BES($\epsilon$) which is valid for all LDGM codes of any length with desired rate, generator degree distribution $L(x)$, and over all encoding functions. In the next section, we derive a lower bound by generalizing the counting method of [12] to the BES($\epsilon$).

## III. Bounds via Counting

The main idea behind the counting method is to find an upper bound on $|\mathcal{C}_b(D)|$, the cardinality of $\mathcal{C}_b(D)$. We show that we are only interested in the cardinality of $\mathcal{C}_b(D)$ for $b = m\epsilon$. Before doing that, we give asymptotic approximation for binomial coefficient in the next lemma.

**Lemma III.1.** *For $\epsilon \in [0, 1]$ and large $m$, the binomial coefficient $\binom{m}{\epsilon m}$ can be approximated as*

$$\binom{m}{\epsilon m} = \frac{2^{mh(\epsilon)}}{\sqrt{2\pi m\epsilon(1-\epsilon)}}(1 + o(1)).$$

*Further, for $\Delta \in \left[-\sqrt{m}(\log m)^{\frac{1}{3}}, \sqrt{m}(\log m)^{\frac{1}{3}}\right]$*

$$\binom{m}{\epsilon m + \Delta} = \binom{m}{\epsilon m}\left(\frac{\epsilon}{1-\epsilon}\right)^{-\Delta} e^{-\frac{\Delta^2}{2m\epsilon(1-\epsilon)}}(1 + o(1).$$

*Proof:* The proof is based on the saddle point approximation of coefficients of power of polynomials. A detailed explanation of which can be found in Appendix D of [13]. The approximations for binomial coefficients can be found in [13, p. 513]. ■

In the next lemma, we derive a lower bound on the average distortion in terms of $|\mathcal{C}_{\epsilon m}(D)|$.

**Lemma III.2.** *Consider lossy compression of BES($\epsilon$) using LDGM code $G$ with encoding map $f : s \mapsto w$, where $s \in \mathcal{S}$ and $w$ is an index word of length $mR$. Let $\hat{s}$ be the reconstruction associated to the word $w$ with the decoding map $g$ given $\hat{s} = wG$. If*

$$\lim_{m\to\infty} \frac{1}{m} \log\left(\epsilon^{\epsilon m}\left(\frac{1-\epsilon}{2}\right)^{m-\epsilon m} |\mathcal{C}_{\epsilon m}(D)|\right) < 0, \tag{3}$$

*then*

$$\frac{1}{m}\mathbb{E}\left[d(S, g(f(S)))\right] \geq D(1 + o(1)). \tag{4}$$

*Proof:* We observe that, $\forall s \in \mathcal{S}_b$,

$$d\left(s, g\left(f\left(s\right)\right)\right) \geq \left\{ \begin{array}{ll} 0, & \text{if } s \in \mathcal{C}_b(D), \\ mD, & \text{if } s \notin \mathcal{C}_b(D). \end{array} \right.$$

and $|\mathcal{S}_b| = \binom{m}{b} 2^{m-b}$. Defining $\delta(m) = \sqrt{m}(\log m)^{1/3}$ and using the above arguments, we obtain

$$\begin{aligned}
\frac{1}{m}\mathbb{E}\left[d(S, g\left(f(S)\right))\right] &= \frac{1}{m}\sum_{b=0}^{m}\mathbb{P}\{S \in \mathcal{S}_b\}\mathbb{E}\left[d(S, g\left(f(S)\right)) \middle| S \in \mathcal{S}_b\right], \\
&\geq \frac{1}{m}\sum_{b=\epsilon m-\delta(m)}^{\epsilon m+\delta(m)}\mathbb{P}\{S \in \mathcal{S}_b\}\mathbb{E}\left[d(S, g\left(f(S)\right)) \middle| S \in \mathcal{S}_b\right], \\
&\geq \frac{1}{m}\sum_{b=\epsilon m-\delta(m)}^{\epsilon m+\delta(m)}\mathbb{P}\{S \in \mathcal{S}_b\}\sum_{s \in \mathcal{S}_b}\mathbb{P}\{S = s | S \in \mathcal{S}_b\} d\left(s, g(f(s))\right), \\
&\geq \frac{1}{m}\sum_{b=\epsilon m-\delta(m)}^{\epsilon m+\delta(m)}\binom{m}{b}\epsilon^b(1-\epsilon)^{m-b}\sum_{s \in \mathcal{S}_b \setminus \mathcal{C}_b(D)}\frac{Dm}{\binom{m}{b}2^{m-b}}, \\
&\geq D\underbrace{\sum_{b=\epsilon m-\delta(m)}^{\epsilon m+\delta(m)}\binom{m}{b}\epsilon^b(1-\epsilon)^{m-b}}_{A} \\
&\quad -D\underbrace{\sum_{b=\epsilon m-\delta(m)}^{\epsilon m+\delta(m)}\epsilon^b\left(\frac{1-\epsilon}{2}\right)^{m-b}|\mathcal{C}_b(D)|}_{B}. \qquad (5)
\end{aligned}$$

As $|\mathcal{C}_b(D)| = e^{o(m)}\mathcal{C}_{\epsilon m}(D)$ for $b \in [\epsilon m - a(m), \epsilon m + a(m)]$, summation $B$ is of the order of $o(1)$ due to (3). We use Stirling's formula and Laplace's method of summation [14, p. 755] to show that $A = 1 + o(1)$.

$$\begin{aligned}
A &= \sum_{b=\epsilon m-\delta(m)}^{\epsilon m+\delta(m)}\binom{m}{b}\epsilon^b(1-\epsilon)^{m-b}, \\
&\stackrel{(i)}{=} \sum_{\Delta=-\delta(m)}^{\Delta=\delta(m)}\binom{m}{\epsilon m+\Delta}\epsilon^{\epsilon m+\Delta}(1-\epsilon)^{m-\epsilon m-\Delta}, \\
&\stackrel{(ii)}{=} \sum_{\Delta=-\delta(m)}^{\Delta=\delta(m)}\frac{e^{-\frac{\Delta^2}{2m\epsilon(1-\epsilon)}}}{\sqrt{2\pi m\epsilon(1-\epsilon)}}(1+o(1)), \\
&\stackrel{(iii)}{=} \frac{1}{\sqrt{2\pi m\epsilon(1-\epsilon)}}\int_{x=-\infty}^{\infty}e^{-\frac{x^2}{2m\epsilon(1-\epsilon)}}(1+o(1)), \\
&= 1 + o(1).
\end{aligned}$$

In step $(i)$, we do a change of variable $b = \epsilon m + \Delta$. Step $(ii)$ follows from Lemma III.1, where we use the approximation of $\binom{m}{m\epsilon}$ and the approximation of $\binom{m}{m\epsilon+\Delta}$ in terms of $\binom{m}{m\epsilon}$. In step $(iii)$, we replace the sum by an integral, as the "variance" of summation terms, which is equal to $m\epsilon(1-\epsilon)$, tends to infinity as $m$ tends to infinity [14, p. 761]. This completes the proof of the lemma. ■

In the next theorem, we state the rate-distortion bound using counting.

**Theorem III.1.** *Consider lossy compression of a BES($\epsilon$) using a LDGM code of block-length $m$ with generator node degree distribution $L(x)$. Let $\hat{S}$ be the set of codewords of the LDGM code. Further let*

$$f(x) = \prod_{i=0}^{d}(1+x^i)^{L_i}, \quad a(x) = \sum_{i=0}^{d} iL_i \frac{x^i}{1+x^i}, \tag{6}$$

$$R(x) = (1-\epsilon)\frac{\left(1-h\left(\frac{x}{1+x}\right)\right)}{1-\log_2\left(\frac{f(x)}{x^{a(x)}}\right)}, \tag{7}$$

$$D(x) = \frac{x(1-\epsilon)}{1+x} - R(x)a(x). \tag{8}$$

*Then, for any block-length $m$, the achievable rate-distortion performance of a LDGM code of rate $R$ and generator degree distribution $L(x)$ is lower bounded by the parametric curve $(D(x), R(x))$, $x \in [0,1]$.*

*Proof:* As was argued in [12], we only need to prove our bound for the limit of block-length tending to infinity. We use Lemma III.2 to prove the theorem. Thus, we need to bound the cardinality of $\mathcal{C}_{\epsilon m}(D)$. Pick $w \in \mathbb{N}$ such that $Dm + w \leq \frac{m-\epsilon m}{2}$. Let $A_m(w)$ denote the number of codewords of Hamming weight at most $w$. From the arguments in [12], we obtain

$$|\mathcal{C}_{\epsilon m}(D)| = |\bigcup_{\hat{s}\in\hat{\mathcal{S}}} \mathcal{B}_{\epsilon m}(\hat{s}, D)| \leq \frac{1}{A_m(w)} \sum_{\hat{s}\in\hat{\mathcal{S}}} \left|\mathcal{B}_{\epsilon m}\left(\hat{s}, D + \frac{w}{m}\right)\right|. \tag{9}$$

Then by using

$$\left|\mathcal{B}_{\epsilon m}\left(\hat{s}, D + \frac{w}{m}\right)\right| = \binom{m}{\epsilon m} \sum_{i=0}^{Dm+w} \binom{m-\epsilon m}{m-\epsilon m - i}, \tag{10}$$

Stirling's approximation, and the observation that the entropy function $h(x)$ is increasing for $x \in [0, 1/2]$ and decreasing for $x \in [1/2, 1]$, we obtain

$$\left|\mathcal{B}_{\epsilon m}\left(\hat{s}, D + \frac{w}{m}\right)\right| \leq 2^{mh(\epsilon)} 2^{(m-\epsilon m)h\left(\frac{Dm+w}{m(1-\epsilon)}\right) + o(m-\epsilon m)}. \tag{11}$$

As was shown in [12], $A_m(w) \geq \sum_{i=0}^{w} \mathrm{coef}(f(x)^{mR}, x^i)$, where $\mathrm{coef}(f(x)^{mR}, x^i)$ denotes the coefficient in front of $x^i$ in the expansion of $f(x)^{mR}$. From Theorem 1 in [15],

$$\mathrm{coef}(f(x)^{mR}, x^w) \leq \inf_{x>0} \frac{f(x)^{mR}}{x^w} = \frac{f(x_\omega)^{mR}}{x_\omega^w} \leq A_m(w), \tag{12}$$

where $x_\omega$ is the unique positive solution to the equation $a(x) = \omega$ and we define $\omega = w/(mR)$. As $Dm + w \leq m\frac{1-\epsilon}{2}$, the maximum of the summation term in (10) occurs at $i = Dm + w$. Thus, we obtain

$$|\mathcal{C}_{\epsilon m}(D)| \leq 2^{m\left[-R\log_2 \frac{f(x_\omega)}{x_\omega^{a(x_\omega)}} + R + h(\epsilon) + (1-\epsilon)h\left(\frac{D+Ra(x_\omega)}{1-\epsilon}\right) + \right] + o(m-\epsilon m)}, \tag{13}$$

where the relation $a(x_\omega) = \omega$. From Lemma III.2 we know that if $\epsilon^{\epsilon m} \left(\frac{1-\epsilon}{2}\right)^{m-\epsilon m} |\mathcal{C}_{\epsilon m}(D)|$ is exponentially small in comparison to $m$, then the average distortion is at least $D$. We now upper bound the growth rate $\epsilon^{\epsilon m} \left(\frac{1-\epsilon}{2}\right)^{m-\epsilon m} |\mathcal{C}_{\epsilon m}(D)|$ using (13), and we obtain

$$\lim_{m\to\infty} \left[\frac{1}{m} \log_2 \left(\epsilon^{m\epsilon} \left(\frac{1-\epsilon}{2}\right)^{m(1-\epsilon)} |\mathcal{C}_{\epsilon m}(D)|\right)\right] \le g(D, R), \tag{14}$$

where

$$g(D, R) = \inf_{\substack{D+a(x)R \le \frac{1-\epsilon}{2} \\ x \ge 0}} h_1(x), \tag{15}$$

with

$$h_1(x) = -R \log_2 \frac{f(x)}{x^{a(x)}} + R + (1-\epsilon) \left(h\left(\frac{D + Ra(x)}{1-\epsilon}\right) - 1\right). \tag{16}$$

If $g(D, R) < 0$ the average normalized distortion is lower bounded by $D$, and thus, we obtain the rate-distortion bound by considering the condition

$$g(D, R) = 0. \tag{17}$$

In order to compute the infimum in (15), we take the derivative of $h_1(x)$ and obtain

$$\frac{dh_1(\beta, x)}{dx} = Ra'(x) \log_2 \left[x \left(\frac{1}{\alpha} - 1\right)\right], \tag{18}$$

where $\alpha = \frac{D+Ra(x)}{1-\epsilon}$. The vanishing derivative conditions imply that

$$\alpha = \frac{x}{1+x}. \tag{19}$$

From (19), we see that if $x \le 1$, then $D+Ra(x) \le (1-\epsilon)/2$. By using (19), the condition $g(D, R) = 0$, and varying $x$ over the interval $x \in [0, 1]$, we obtain the desired parametric lower bound on the rate-distortion performance. ■

*Remark:* Based on the arguments of [12], the parametric bound can be improved for low rates using the straight-line bound. More precisely, let $L' = L'(1)$ and if $R \in \left[0, \frac{1}{L'(1)}\right]$, then

$$D = \frac{1}{2}\left(1 - RL'\left(1 - 2\left(\frac{x\left(\frac{1}{L'}\right)(1-\epsilon)}{1 + x\left(\frac{1}{L'}\right)} - \frac{a\left(x\left(\frac{1}{L'}\right)\right)}{L'}\right)\right)\right), \tag{20}$$

where $x\left(\frac{1}{L'}\right)$ is the unique solution of $R(x) = \frac{1}{L'}$, and $R(x)$ is defined in (7).

In the next section, we derive the rate-distortion bound by using the test-channel-based probabilistic arguments developed in [12].

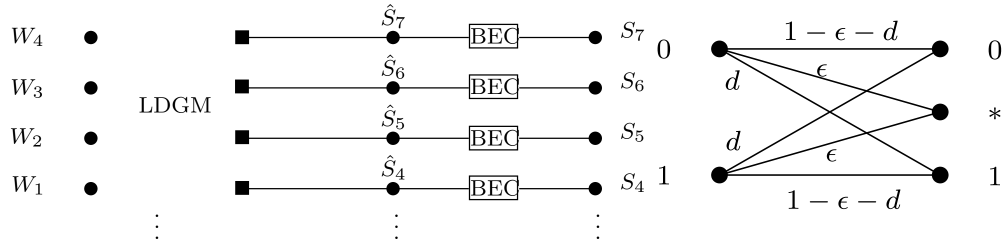
## IV. Bounds via Test Channel

In this section we use a probabilistic argument to bound the cardinality of $\mathcal{C}_b(D)$. We consider the test channel model represented in Figure 1(a) with the binary error/erasure channel represented in Figure 1(b). More precisely, choose an index word $w \in \mathcal{W}$ at random. This generates the corresponding codeword $\hat{S}$, then send each component of $\hat{S}$ through the binary error/erasure channel defined in Figure 1(b).

We prove the rate-distortion bound in the following theorem. We only consider the regular case for the sake of simplicity of exposition. As was argued in [12], it suffices to prove the bound only for the limit of block-length tending to infinity. Also, the obtained bound can be strengthened for low rates using the straight-line bound.

**Theorem IV.1.** *Consider lossy compression of a $BES(\epsilon)$ using a LDGM code with codeword set $\hat{S}$, regular generator degree distribution $L(x) = x^l$, block-length $m$, and rate $R$. Then for an average normalized distortion $D$, the rate $R$ is lower bounded by,*

$$R \geq \sup_{D \leq d \leq \frac{1-\epsilon}{2}} \frac{(1-\epsilon)\left(1-\log_2(1-\epsilon)\right) + (1-D-\epsilon)\log_2(1-\epsilon-d) + D\log_2(d)}{1-\log_2\left(1+\left(\frac{d}{1-\epsilon-d}\right)^l\right)}. \tag{21}$$




(a) (b)

Figure 1: (a) Test channel, (b) Binary error/erasure channel.

*Proof:* We will use Lemma III.2, so we will focus on source words with number of erasures equal to $\epsilon m$. Consider a particular $s \in \mathcal{C}_{\epsilon m}(D)$. This implies that $\exists \hat{s} \in \hat{\mathcal{S}}$ such that $d(s, \hat{s}) \leq Dm$. Let $c = \frac{d}{1-\epsilon-d}$ and $D \leq d \leq (1-\epsilon)/2$. Then,

$$\begin{aligned}
&\mathbb{P}\{S = s\} \\
&= \textstyle\sum_{\hat{s}' \in \hat{\mathcal{S}}} \mathbb{P}\{S = s, \hat{S} = \hat{s}'\}, && (22)\\
&= \textstyle\sum_{w=0}^{m} \sum_{\hat{s}' \in \hat{\mathcal{S}}: d(\hat{s}', \hat{s}) = w} \mathbb{P}\{S = s, \hat{S} = \hat{s}'\}, && (23)\\
&= 2^{-mR}\epsilon^{\epsilon m}(1-\epsilon-d)^{m-\epsilon m} \textstyle\sum_{w=0}^{m} \sum_{\hat{s}' \in \hat{\mathcal{S}}: d(\hat{s}', \hat{s}) = w} c^{d(s, \hat{s}')}, && (24)\\
&\overset{\text{(i)}}{\geq} 2^{-mR}\epsilon^{\epsilon m}(1-\epsilon-d)^{m-\epsilon m} \textstyle\sum_{w=0}^{m} \sum_{\hat{s}' \in \hat{\mathcal{S}}: d(\hat{s}', \hat{s}) = w} c^{d(s, \hat{s}) + d(\hat{s}, \hat{s}')}, && (25)\\
&\overset{\text{(ii)}}{\geq} 2^{-mR}\epsilon^{\epsilon m}(1-\epsilon-d)^{m-\epsilon m} \textstyle\sum_{w=0}^{m} \sum_{\hat{s}' \in \hat{\mathcal{S}}: d(\hat{s}', \hat{s}) = w} c^{Dm + w}, && (26)\\
&\geq 2^{-mR}\epsilon^{\epsilon m}(1-\epsilon-d)^{m-\epsilon m} \textstyle\sum_{w=0}^{m} A_m(w) c^{Dm + w}, && (27)
\end{aligned}$$

where step (i) follows since $d \leq \frac{1-\epsilon}{2}$, and step (ii) follows since $d(\hat{s}, s) \leq Dm$ and $d(\hat{s}, \hat{s}') = w$.

Assuming $\frac{Rc^l}{1+c^l} < \frac{1}{l}$, we obtain the following inequality (see [12] for proof),

$$\sum_{w=0}^{m} A_m(w) c^w \geq \frac{1}{m}\left(1 + c^l\right)^{mR}. \tag{28}$$

Thus for $s \in \mathcal{C}_{\epsilon m}(D)$,

$$\mathbb{P}\{S = s\} \geq \frac{1}{m} 2^{-mR}\left(1 + c^l\right)^{mR} \epsilon^{\epsilon m}(1-\epsilon-d)^{m(1-D)-\epsilon m} d^{Dm}. \tag{29}$$

Also,

$$\binom{m}{\epsilon m} \epsilon^{\epsilon m}(1-\epsilon)^{m-\epsilon m} = \sum_{s \in \mathcal{S}_b} \mathbb{P}\{S = s\} \geq \sum_{s \in \mathcal{C}_b(D)} \mathbb{P}\{S = s\}. \tag{30}$$

We obtain the following upper bound on $|\mathcal{C}_b(D)|$ by combining (29) and (30)

$$|\mathcal{C}_{\epsilon m}(D)| \leq m2^{mR}(1+c^l)^{-mR}(1-\epsilon-d)^{-m(1-D)}(1-\epsilon)^m d^{-Dm}\binom{m}{\epsilon m}\left(\frac{1-\epsilon-d}{1-\epsilon}\right)^{\epsilon m}. \tag{31}$$

By using Lemma III.2, we know that

$$\frac{1}{m}\mathbb{E}\left[d(S, g\left(f(S)\right))\right] \geq D(1+o(1)),$$

if (3) is true i.e.

$$\lim_{m\to\infty}\frac{1}{m}\log\left(\epsilon^{\epsilon m}\left(\frac{1-\epsilon}{2}\right)^{m-\epsilon m}|\mathcal{C}_{\epsilon m}(D)|\right) < 0.$$

We use the bound on $\mathcal{C}_{\epsilon m}(D)$ from (31), plug it in the above condition, and obtain that if

$$\begin{aligned} R - R\log_2(1+c^l) + (1-\epsilon)(-1+\log_2(1-\epsilon)) \\ - (1-D-\epsilon)\log_2(1-\epsilon-d) - D\log_2(d) < 0, \end{aligned} \tag{32}$$

then the distortion is at least $D$. This proves (21).

We still need to prove that $\frac{Rc^l}{1+c^l} < \frac{1}{l}$, which is trivially fulfilled for $R < \frac{1}{l}$ since $c < 1$. Assuming $R > \frac{1}{l}$, we need to prove that $d < \frac{1-\epsilon}{1+(Rl-1)^{\frac{1}{l}}}$. By taking the derivative of the LHS of (32) with respect to $d$, we find the expression for our optimal $d$ to be

$$d = \frac{1-\epsilon}{1+\left[\frac{Rl}{d-D}-1\right]^{\frac{1}{l}}}.$$

As $D \leq d \leq \frac{1-\epsilon}{2}$, it implies $d-D < 1$. This in turn implies that the optimal $d$ is less than $\frac{1-\epsilon}{1+(Rl-1)^{\frac{1}{l}}}$. This gives us the desired result. Thus the inequality $\frac{Rc^l}{1+c^l} < \frac{1}{l}$ always holds and we get (28). ∎

*Remark:* Note that the lower bound on the rate in (21) is lower bounded by the Shannon rate distortion function. In order to see this, we put $d = D$ on the LHS of (21). This give,

$$(1-\epsilon)(1-\log_2(1-\epsilon)) + (1-D-\epsilon)\log_2(1-\epsilon-D) + D\log_2(D).$$

In the above expression, we multiply and divide $D$ by $(1-\epsilon)$. After simplification we obtain

$$\begin{aligned} (1-\epsilon)\left(1+\left(1-\frac{D}{1-\epsilon}\right)\log_2\left(1-\frac{D}{1-\epsilon}\right)+\frac{D}{1-\epsilon}\log_2\left(\frac{D}{1-\epsilon}\right)\right) \\ = (1-\epsilon)\left(1-h\left(\frac{D}{1-\epsilon}\right)\right), \end{aligned}$$

which is the Shannon's rate distortion function for BES($\epsilon$).

## V. The Bounds defined in Theorems III.1 and IV.1 are equal

For the BSS it was numerically observed in [12] that the bounds obtained by the counting method and the test channel method are identical. In the following theorem we prove equality of the two bounds for a BES($\epsilon$). As the BES(0) is a BSS, we also implicitly prove the equality of the two bounds for a BSS. Due to simplicity of exposition, we prove our result for regular LDGM codes.

**Theorem V.1.** *Consider lossy compression of a BES($\epsilon$) using a regular LDGM code with codeword set $\hat{S}$, and generator degree distribution $L(x) = x^l$. Then the lower bounds on the rate-distortion of $\hat{S}$ obtained in Theorem III.1 and Theorem IV.1 are identical.*

*Proof:* Considering (21), we define the function $v(d)$

$$v(d) = \frac{(1-\epsilon)\left(1-\log_2(1-\epsilon)\right) + (1-D-\epsilon)\log_2(1-\epsilon-d) + D\log_2(d)}{1-\log_2\left(1+\left(\frac{d}{1-\epsilon-d}\right)^l\right)} \tag{33}$$

By determining the maximum of $v(d)$ in parametric form, we show that the resulting expression is the same as the bound in Theorem III.1.

We first do a change of variable in $v(d)$ as $x = \frac{d}{1-\epsilon-d}$ and write $v(d)$ as $v(x)$,

$$v(x) = \frac{(1-\epsilon)\left(1-\log_2(1-\epsilon)\right) + (1-D-\epsilon)\log_2\left(\frac{1-\epsilon}{1+x}\right) + D\log_2\left((1-\epsilon)\frac{x}{1+x}\right)}{1-\log_2\left(1+x^l\right)}. \tag{34}$$

In order to compute the maximum, we take the derivative of $v(x)$ which is given by,

$$\frac{dv(x)}{dx} = \frac{1}{1-\log_2(1+x^l)}$$
$$\times\left(\frac{D}{x} - \frac{1-\epsilon}{1+x} + \frac{lx^{l-1}}{(1+x^l)(1-\log_2(1+x^l))}\left((1-\epsilon)\left(1-\log_2(1+x)\right) + D\log_2(x)\right)\right).$$

Equating the derivative to zero and solving for $D$ results in

$$D = (1-\epsilon)\frac{x}{1+x} - a(x)(1-\epsilon)\frac{\frac{x}{1+x}\log_2(x) + 1 - \log_2(1+x)}{1-\log_2\left(\frac{f(x)}{x^{a(x)}}\right)},$$

where

$$f(x) = 1 + x^l, a(x) = \frac{lx^l}{1+x^l}. \tag{35}$$

Note that

$$1 - h\left(\frac{x}{1+x}\right) = 1 - \log_2(1+x) + \frac{x}{1+x}\log_2(x).$$

This results in

$$D = (1-\epsilon)\frac{x}{1+x} - (1-\epsilon)a(x)\frac{1-h\left(\frac{x}{1+x}\right)}{1-\log_2\left(\frac{f(x)}{x^{a(x)}}\right)} = (1-\epsilon)\frac{x}{1+x} - a(x)R(x), \tag{36}$$

where $R(x)$ is the expression for the rate given in Theorem III.1. This gives the same expression for $D$ as in Theorem III.1. Substituting the expression for $D$ from (36) into (34), we obtain the expression for $R$ which is the same as that of Theorem III.1. ■

## VI. Conclusion and open questions

We derived lower bounds on the rate-distortion performance of LDGM codes over the BES using two different methods and then showed that these latter lead to the same result. Although we considered regular LDGM codes for the sake of simplicity, the generalization to the irregular case is straightforward and the two methods we used are valid for any sparse graph code.

An open question was asked in [12]: "What is the relationship between the test channel model and the rate-distortion problem?" It was further conjectured that a $(R, D)$ pair is only achievable for the BSS if the entropy $H(s) = m$ in the test channel model. The equivalent conjecture for BES$(\epsilon)$ will be that an $(R, D)$ pair is only achievable if $H(s) = (1-\epsilon)m$ in the test channel model. Based on this conjecture and Gallager's bound, it was conjectured that for the BSS a stronger rate-distortion bound is valid. The extension of this conjecture to BES$(\epsilon)$ would result in the following bound

$$R \geq \frac{1-\epsilon+h(\epsilon)-\left(\epsilon\log_2\frac{1}{\epsilon}+D\log_2\frac{1}{D}+(1-\epsilon-D)\log_2\frac{1}{1-\epsilon-D}\right)}{1-\sum_{f=0}^{l}\binom{l}{f}\epsilon^f\sum_{w=0}^{l-f}\binom{l-f}{w}D^w(1-\epsilon-D)^{l-f-w}h\left(\frac{1}{1+\left(\frac{D}{1-\epsilon-D}\right)^{l-f-2w}}\right)}$$

Our main future objective is to prove this conjecture.

## References


[1] E. Martinian and J. Yedidia, "Iterative quantization using codes on graphs," in *Proc. 35th Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, 2003.

[2] S. Cilberti, M. Mezard, and R. Zecchina, "Lossy data compression with random gates," *Phyics Review Letters*, vol. 95, 2005.

[3] M. J. Wainwright and E. Maneva, "Lossy source encoding via message-passing and decimation over generalized codewords of LDGM codes," in *Proc. of the IEEE International Symposium on Information Theory*, Adelaide, Australia, Sep. 2005, pp. 1493–1497.

[4] A. Braunstein and R. Zecchina, "Survey propagation as local equilibrium equations," *Journal of Statistical Mechanics*, Jun. 2004.

[5] T. Filler and J. Fridrich, "Binary quantization using belief propagation with decimation over factor graphs of LDGM codes," in *Proc. of the Allerton Conference on Communication, Control, and Computing*, Sep. 2007.

[6] E. Martinian and M. J. Wainwright, "Analysis of LDGM and compound codes for lossy compression and binning," in *Proc. of the IEEE Information Theory Workshop*, San Diego, CA, USA, Feb. 2006.

[7] ——, "Low-density codes achieve the rate-distortion bound," in *Proc. of the Data Compression Conference*, Snowbird, UT, Mar. 2006.

[8] ——, "Low-density construction can achieve the Wyner-Ziv and Gelfand-Pinsker bounds," in *Proc. of the IEEE International Symposium on Information Theory*, Jul. 2006, pp. 484–488.

[9] E. Arikan, "Polar codes: A deterministic construction of capacity achieving codes," *IEEE Transactions on Information Theory*, vol. 50, no. 6, Jul. 2009.

[10] S. Korada and R. Urbanke, "Optimality of polar codes for lossy compression," submitted to IEEE Transactions on Information Theory.

[11] A. Dimakis, M. J. Wainwright, and K. Ramchandran, "Lower bounds on the rate-distortion function of LDGM codes," in *Proc. of the IEEE Information Theory Workshop*, 2007.

[12] S. Kudekar and R. Urbanke, "Lower bounds on the rate-distortion function of individual LDGM codes," in *5th International Symposium on Turbo Codes and Related Topics*, Lausanne, Switzerland, 2008.

[13] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[14] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*. Cambridge University Press, 2009.

[15] D. Burshtein and G. Miller, "Asymptotic enumeration methods for analyzing LDPC codes," *IEEE Transactions on Information Theory*, vol. 50, no. 6, Jun. 2004.