Lower Bounds on the Rate-Distortion Function of Individual LDGM Codes

Shrinivas Kudekar and Rüdiger Urbanke
EPFL, School of Computer and Communication Sciences, Lausanne 1015, Switzerland

***Abstract*— We consider lossy compression of a binary symmetric source by means of a low-density generator-matrix code. We derive two lower bounds on the rate distortion function which are valid for any low-density generator-matrix code with a given node degree distribution $L(x)$ on the set of generators and for any encoding algorithm. These bounds show that, due to the sparseness of the code, the performance is strictly bounded away from the Shannon rate-distortion function. In this sense, our bounds represent a natural generalization of Gallager's bound on the maximum rate at which low-density parity-check codes can be used for reliable transmission. Our bounds are similar in spirit to the technique recently developed by Dimakis, Wainwright, and Ramchandran, but they apply to *individual* codes.**

I. Introduction

We consider lossy compression of a binary symmetric source (BSS) using a low-density generator-matrix (LDGM) code as shown in Figure 1. More precisely, let $S \in \mathbb{F}_2^m$ represent the binary source of length m. We have $S = \{S_1, S_2, \dots, S_m\}$, where the $\{S_i\}_{i=1}^m$ are iid random variables with $\mathbb{P}\{S_i = 1\} = \frac{1}{2}$, $i \in [m]$. Let $\mathcal{S}$ denote the set of all source words.

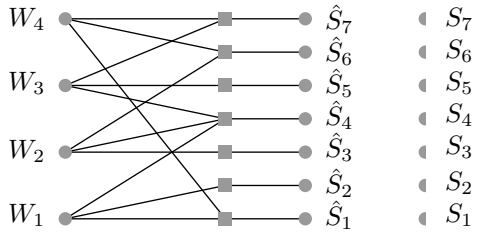

Fig. 1. The Tanner graph corresponding to a simple LDGM code used for lossy compression of a BSS. We have $m = 7$, $R = \frac{4}{7}$, and $L(x) = x^3$.

Given a source word $s \in \mathcal{S}$, we compress it by mapping it to one of the 2^{mR} index words $w \in \mathcal{W} = \mathbb{F}_2^{mR}$, where R is the *rate*, $R \in [0, 1]$. We denote this encoding map by $\mathrm{f} : s \mapsto W$ (the map can be random). The reconstruction is done via an LDGM code determined by a sparse binary $mR \times m$ generator matrix G. Let $\hat{s}$ denote the reconstructed word associated to w. We have $\hat{s} = wG$. We denote this decoding map by $\mathrm{g} : w \mapsto \hat{s}$. Let $\hat{\mathcal{S}}$ denote the code, $\hat{\mathcal{S}} = \{\hat{s}^{(1)}, \dots, \hat{s}^{(2^{mR})}\}$, $\hat{s}^{(i)} \in \mathbb{F}_2^m$. The codewords are not necessarily distinct.

We call the components of the index word $w = \{w_1, \dots, w_{mR}\}$ the *generators* and the associated nodes in the factor graph representing the LDGM code the *generator nodes*. We assume that these generators nodes have a normalized degree distribution $L(x) = \sum_i L_i x^i$. This means that L_i represents the fraction (out of mR) of generator nodes of degree i.

We are interested in the trade-off between rate and distortion which is achievable in this setting. Let $\mathrm{d}(\cdot, \cdot)$ denote the Hamming distortion function, $\mathrm{d} : \mathbb{F}_2^m \times \mathbb{F}_2^m \to \mathbb{N}$. The average distortion is then given by

$$\frac{1}{m}\,\mathbb{E}[d(S, \mathrm{g}(\mathrm{f}(S)))].$$

We are interested in the minimum of this average distortion, where the minimum is taken over all LDGM codes of a given rate, generator degree distribution $L(x)$, and length, as well as over all encoding functions.

II. Review

Given the success of sparse graph codes applied to the channel coding problem, it is not surprising that there is also interest in the use of sparse graph codes for the source coding problem. Martinian and Yedidia [1] were probably the first to work on lossy compression using sparse graph codes. They considered a memoryless ternary source with erasures and demonstrated a duality result between compression of this source and the transmission problem over a binary erasure channel (both using iterative encoding/decoding). Mezard, Zecchina, and Ciliberti [2] considered the lossy compression of the BSS using LDGM codes with a Poisson distribution on the generators. They derived the one-step replica symmetry-breaking (1RSB) solution and the average rate-distortion function. According to this analysis, this ensemble approaches the Shannon rate-distortion curve exponentially fast in the average degree. They observed that the iterative interpretation associated to the 1RSB analysis gives rise to an algorithm, which they called *survey propagation*. In [3] the same authors implement an encoder that utilizes a Tanner graph with random non-linear functions at the check nodes and a *survey propagation* based decimation algorithm for data compression of the BSS. In [4], Wainwright and Maneva also considered the lossy compression of a BSS using an LDGM code with a given degree distribution. They showed how survey propagation can be interpreted as belief propagation algorithm (as did Braunstein and Zecchina [5]) on an enlarged set of assignments and demonstrated that the survey propagation algorithm is a practical and efficient encoding scheme. Recently, Filler and Friedrich [6] demonstrated experimentally that even standard belief propagation based decimation algorithms using optimized degree distributions for LDGM codes

and a proper initialization of the messages can achieve a rate-distortion trade-off very close to the Shannon bound. Martinian and Wainwright [7], [8], [9] constructed *compound LDPC and LDGM code ensembles* and gave rigorous *upper bounds* on their distortion performance. A standard LDGM code ensemble is a special case of their construction, hence they also provide *upper bounds* on the rate-distortion function of LDGM ensembles. By using the first and second moment method they proved that a code chosen randomly from the *compound ensemble* under optimal encoding and decoding achieves the Shannon rate-distortion curve with high probability. Finally, they pointed out that such constructions are useful also in a more general context (e.g., the Wyner-Ziv or the Gelfand-Pinsker problem). Dimakis et al [10] were the first authors to provide rigorous *lower bounds* on the rate-distortion function of LDGM code ensembles.

Theorem 1 (Dimakis, Wainwright, Ramchandran [10]): Let $\hat{\mathcal{S}}$ be a binary code of blocklength m and rate R chosen uniformly at random from an ensemble of left Poisson LDGM Codes with check-node degree $\mathtt{r}$. Suppose that we perform MAP decoding. With high probability the rate-distortion pair (R, D) achieved by $\hat{\mathcal{S}}$ fulfills

$$R \geq \frac{1-h(D)}{1-e^{-\frac{(1-D)\mathtt{r}}{R}}} > 1-h(D).$$

A. Outline

In the spirit of Gallager's information theoretic bound for LDPC codes, we are interested in deriving lower bounds on the rate-distortion function which are valid for *any* LDGM code with a given generator node degree distribution $L(x)$. Our approach is very simple. Pick a parameter D, $D \in [0, \frac{1}{2}]$ (think of this parameter as the distortion). Consider the set of "covered" sequences

$$\mathcal{C}(D) = \bigcup_{\hat{s}\in\hat{\mathcal{S}}} \mathcal{B}(\hat{s}, Dm), \tag{1}$$

where $\mathcal{B}(x,i)$, $x \in \mathbb{F}_2^m$, $i \in [m]$, is the Hamming ball of radius i centered at x. In words, $\mathcal{C}(D)$ represents the set of all those source sequences that are within Hamming distance at most Dm from at least one code word.

Recall that for any $s \in \mathcal{S}$, $\mathrm{f}(s) \in \mathcal{W}$ represents the index word and that $\mathrm{g}(\mathrm{f}(s))$ denotes the reconstructed word. We have

$$\mathrm{d}(s, \mathrm{g}(\mathrm{f}(s))) \geq \begin{cases} 0, & s \in \mathcal{C}(D), \\ Dm, & s \in \mathbb{F}_2^m \setminus \mathcal{C}(D). \end{cases}$$

Therefore,

$$\begin{aligned} &\frac{1}{m}\,\mathbb{E}[\mathrm{d}(S, \mathrm{g}(\mathrm{f}(S)))] \\ &= \frac{1}{m}\sum_{s\in\mathbb{F}_2^m} 2^{-m}\,\mathrm{d}(s, \mathrm{g}(\mathrm{f}(s))) \geq \frac{2^{-m}}{m}\sum_{s\in\mathbb{F}_2^m\setminus\mathcal{C}(D)} \mathrm{d}(s, \mathrm{g}(\mathrm{f}(s))) \\ &\geq 2^{-m} D|\mathbb{F}_2^m \setminus \mathcal{C}(D)| \geq D\big(1-2^{-m}|\mathcal{C}(D)|\big). \end{aligned} \tag{2}$$

If the codewords are well spread out then we know from Shannon's random coding argument that for a choice $D = h^{-1}(1-R)$, $|\mathcal{C}(D)| \approx 2^m$, [11]. But the codewords of an LDGM code are clustered since changing a single generator symbol only changes a constant number of symbols in the codeword. There is therefore substantial overlap of the balls. We will show that there exists a D which is strictly larger than the distortion corresponding to Shannon's rate-distortion bound so that $|\mathcal{C}(D)|$ is exponentially small compared to 2^m regardless of the specific code. From (2) this implies that the distortion is at least D.

To derive the required upper bound on $|\mathcal{C}(D)|$ we use two different techniques. In Section III we use a simple combinatorial argument. In Section IV, on the other hand, we employ a probabilistic argument based on the "test channel" which is typically used to show the achievability of the Shannon rate-distortion function.

Although both bounds prove that the rate-distortion function is strictly bounded away from the Shannon rate-distortion function for the whole range of rates and any LDGM code, we conjecture that a stronger bound is valid. We pose our conjecture as an open problem in Section V.

III. Bound Via Counting

Theorem 2 (Bound Via Counting): Let $\hat{\mathcal{S}}$ be an LDGM code with blocklength m and with generator node degree distribution $L(x)$ and define $L' = L'(1)$. Let

$$f(x) = \prod_{i=0}^{d}(1+x^i)^{L_i}, \quad a(x) = \prod_{i=0}^{d} iL_i \frac{x^i}{1+x^i},$$
$$\hat{R}(x) = \frac{1-h(\frac{x}{1+x})}{1-\log\frac{f(x)}{x^{a(x)}}}, \quad \hat{D}(x) = \frac{x}{1+x} - a(x)\hat{R}(x).$$

For $R \in [\frac{1}{L'}, 1]$ let $x(R)$ be the unique positive solution of $\hat{R}(x) = R$. Define the curve $D(R)$ as

$$\begin{cases} \frac{1}{2}\Big(1 - RL'\big(1 - 2\big(\frac{x(\frac{1}{L'})}{1+x(\frac{1}{L'})} - \frac{a(x(\frac{1}{L'}))}{1}\big)\big)\Big), & R \in [0, \frac{1}{L'}], \\ \hat{D}(x(R)), R \in [\frac{1}{L'}, 1]. \end{cases}$$

Then, for any blocklength m, the achievable distortion of an LDGM code of rate R and generator degree distribution $L(x)$ is lower bounded by $D(R)$.

Discussion: (i) As stated above, if we are considering a single code of rate R then the lower bound on the distortion is $D(R)$. If, on the other hand we are considering a family of codes, all with the same generator degree distribution $L(x)$ but with different rates R, then it is more convenient to plot the lower bound in a parametric form. First plot the curve $(\hat{D}(x), \hat{R}(x))$ for $x \in [0,1]$. Then connect the point $(D = \frac{1}{2}, R = 0)$ to the point on the $(\hat{D}(x), \hat{R}(x))$ curve with $\hat{R}(x) = \frac{1}{L'}$ by a straight line. The resulting upper envelope gives the stated lower bound for the whole range. This construction is shown in Figure 2. (ii) Although this is difficult to glance from the expressions, we will see in the proof that for any bounded generator degree distribution $L(x)$ the performance is strictly bounded away from the Shannon rate-distortion function. From a practical perspective however the gap to the rate-distortion bound decreases quickly in the degree.

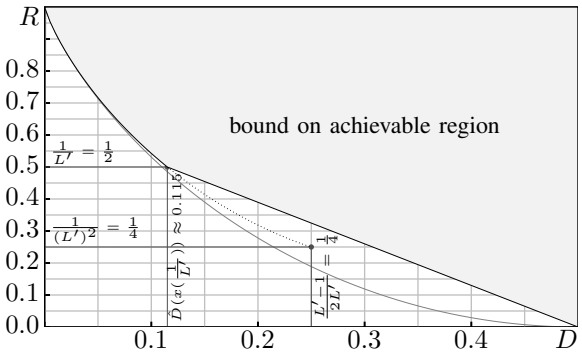

Fig. 2. Construction of the bound for codes with $L(x) = x^2$ so that $L' = 2$ (all generator nodes have degree 2). The solid gray curve corresponds to the Shannon rate-distortion curve. The black curve just above, which is partially solid and partially dotted, corresponds to the curve $(\hat{D}(x), \hat{R}(x))$ for $x \in [0, 1]$. It starts at the point $(0, 1)$ (which corresponds to $x = 0$) and ends at $(\frac{L'-1}{2L'} = \frac{1}{4}, \frac{1}{(L')^2} = \frac{1}{4})$ which corresponds to $x = 1$. The straight line goes from the point $(\hat{D}(x(\frac{1}{L'})), \frac{1}{L'})$ to the point $(\frac{1}{2}, 0)$. Any achievable (R, D) pair must lie in the lightly shaded region. This region is strictly bounded away from the Shannon rate-distortion function over the whole range.

Example 1 (Generator-Regular LDGM Codes): Consider codes with generator degree equal to $\mathtt{l}$ and an arbitrary degree distribution on the check nodes. In this case we have $f(x) = 1 + x^{\mathtt{l}}$ and $a(x) = \frac{\mathtt{l}x^{\mathtt{l}}}{1+x^{\mathtt{l}}}$. Figure 3 compares the lower bound to the rate-distortion curve for $\mathtt{l} = 1$, 2, and 3. For each case the achievable region is strictly bounded away from the Shannon rate-distortion curve.

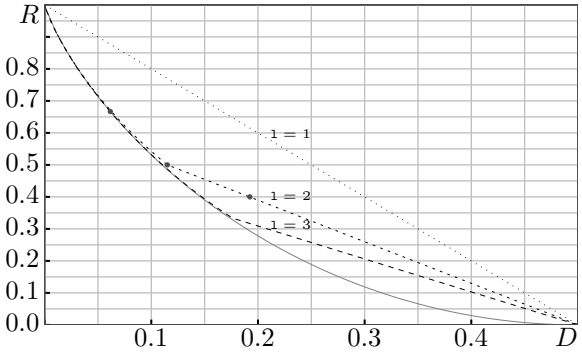

Fig. 3. Bounds for $L(x) = x^{\mathtt{l}}$ for $\mathtt{l} = 1$, 2, and 3. For $\mathtt{l} = 2$ the 3 gray dots correspond to the special cases $R = \frac{2}{3}$, $R = \frac{1}{2}$, and $R = \frac{2}{5}$ respectively. The corresponding lower bounds on the distortion are $D(\frac{2}{3}) \geq 0.0616 > 0.0614905$ (rate-distortion bound), $D(\frac{1}{2}) \geq 0.115 > 0.11$ (rate-distortion bound), and $D(\frac{2}{5}) \geq 0.1924 > 0.1461$ (rate-distortion bound).

Example 2 ($(\mathtt{l}, \mathtt{r})$-Regular LDGM Codes): In this case we have $R = \mathtt{l}/\mathtt{r}$ and $L(x) = x^{\mathtt{l}}$. The same bound as in Example 1 applies. The three special cases $(\mathtt{l} = 2, \mathtt{r} = 3)$, $(\mathtt{l} = 2, \mathtt{r} = 4)$, and $(\mathtt{l} = 2, \mathtt{r} = 5)$, which correspond to $R = \frac{2}{3}$, $R = \frac{1}{2}$, and $R = \frac{2}{5}$ respectively, are marked in Figure 3 as gray dots.

Example 3 ($\mathtt{r}$-Regular LDGM Codes of Rate R): Assume that all check nodes have degree $\mathtt{r}$ and that the connections are chosen uniformly at random with repetitions. For large blocklengths this implies that the degree distribution on the variable nodes converges to a Poisson distribution, i.e., we have in the limit

$$L(x) = \sum_{i=1}^{\infty} L_i x^i = e^{\frac{\mathtt{r}}{R}(x-1)}.$$

Let us evaluate our bound for this generator degree distribution. Note that since the average degree of the *check* nodes is fixed we have a different generator degree distribution $L(x)$ for each rate R. Figure 4 compares the resulting bound with the Shannon rate-distortion function as well as the bound of Theorem 1. The new bound is slightly tighter. But more importantly, it applies to *any* LDGM code.

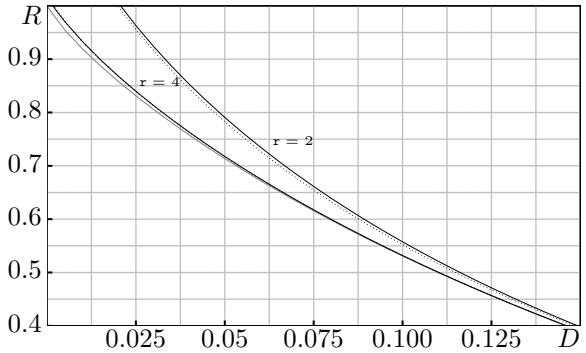

Fig. 4. Lower bound on achievable (R, D) pairs for $\mathtt{r}$-regular LDGM codes with a Poisson generator degree distribution and $\mathtt{r} = 2, 4$. The dashed curve corresponds to the bound of Theorem 1 and the solid black curve represents the bound of Theorem 2. The gray curve is the Shannon rate-distortion tradeoff.

Proof of Theorem 2. From the statement in Theorem 2 you see that the bound consists of a portion of the curve $(\hat{D}(x), \hat{R}(x))$ and a straight-line portion. The straight-line portion is easily explained. Assume that all generator nodes have degree $\mathtt{l}$ (for the general case replace all mentions of $\mathtt{l}$ by the average degree L'). Then the maximum number of check nodes that can depend on the choice of generator nodes is $n\mathtt{l}$. Therefore, if the rate R is lower than $\frac{1}{\mathtt{l}}$ then at least a fraction $(1 - R\mathtt{l})$ of the check nodes cannot be connected to any generator node. For those nodes the average distortion is $\frac{1}{2}$, whereas for the fraction $R\mathtt{l}$ of the check nodes which are (potentially) connected to at least one generator node the best achievable distortion is the same for any $0 \leq R \leq \frac{1}{\mathtt{l}}$. It suffices therefore to restrict our attention to rates in the range $[\frac{1}{L'}, 1]$ and to prove that their (R, D) pairs are lower bounded by the curve $(\hat{D}(x), \hat{R}(x))$.

As a second simplification note that although the bound is valid for all blocklengths m we only need to prove it for the limit of infinite blocklengths. To see this, consider a particular code of blocklength m. Take k identical copies of this code and consider these k copies as one code of blocklength km. Clearly, this large code has the same rate R, the same generator degree distribution $L(x)$, and the same distortion D as each component code. By letting k tend to infinity we can construct an arbitrarily large code of the same characteristics and apply the bound to this limit. Since our bound below is valid for

any sequence of codes whose blocklength tends to infinity the claim follows.

Pick $w \in \mathbb{N}$ so that $Dm + w \leq \frac{m}{2}$. Then

$$\begin{aligned}|\mathcal{C}(D)| &= |\bigcup_{\hat{s}\in\hat{\mathcal{S}}} \mathcal{B}(\hat{s}, Dm)| \\ &\overset{\text{(i)}}{\leq} \frac{1}{A_m(w)} \sum_{\hat{s}\in\hat{\mathcal{S}}} |\mathcal{B}(\hat{s}, Dm + w)| \\ &\overset{\text{(ii)}}{\leq} 2^{-mR\log\frac{f(x_\omega)}{x_\omega^\omega}+o_m(1)} 2^{mR} 2^{mh(D+w/m)} \\ &\overset{\text{(iii)}}{=} 2^{m(-R\log\frac{f(x_\omega)}{x_\omega^{a(x_\omega)}}+R+h(D+a(x_\omega)R)+o_m(1))}.\end{aligned}$$

To see (i) note that a "big" sphere $\mathcal{B}(\hat{s}, Dm + w)$, where $\hat{s} \in \hat{\mathcal{S}}$, contains all "small" spheres of the form $\mathcal{B}(\hat{s}', Dm)$, where $\hat{s}' \in \hat{\mathcal{S}}$ so that $\mathrm{d}(\hat{s}, \hat{s}') \leq w$. Let $A_m(w)$ be the number of codewords of Hamming weight at most w. Then, by symmetry, each small sphere $\mathcal{B}(\hat{s}', Dm)$ is in exactly $A_m(w)$ big spheres $\mathcal{B}(\hat{s}, Dm+w)$. It follows that every point in $\bigcup_{\hat{s}\in\hat{\mathcal{S}}} \mathcal{B}(\hat{s}, Dm)$ is counted at least $A_m(w)$ times in the expression $\sum_{\hat{s}\in\hat{\mathcal{S}}} |\mathcal{B}(\hat{s}, Dm + w)|$.

Consider now step (ii). We need a lower bound on $A_m(w)$. Assume at first that all generator nodes have degree $\mathtt{l}$. Assume that exactly g generator nodes are set to 1 and that all other nodes are set to 0. There are $\binom{mR}{g}$ ways of doing this. Now note that for each such constellation the weight of the resulting codeword is at most $w = g\mathtt{l}$. It follows that in the generator regular case we have

$$A_m(w) \geq \sum_{g=0}^{w/\mathtt{l}} \binom{mR}{g}. \tag{3}$$

We can rewrite (3) in the form

$$A_m(w) \geq \sum_{i=0}^{w} \mathrm{coef}\{(1 + x^{\mathtt{l}})^{mR}, x^i\}, \tag{4}$$

where $\mathrm{coef}\{(1 + x^{\mathtt{l}})^{mR}, x^i\}$ indicates the coefficient of the polynomial $(1 + x^{\mathtt{l}})^{mR}$ in front of the monomial x^i. The expression (4) stays valid also for irregular generator degree distributions $L(x)$ if we replace $(1 + x^{\mathtt{l}})^{mR}$ with $f(x)^{mR}$, where $f(x) = \prod_i (1+x^i)^{L_i}$ as defined in the statement of the theorem. This of course requires that n is chosen in such a way that $nL_i \in \mathbb{N}$ for all i.

Define $N_m(w) = \sum_{i=0}^{w} \mathrm{coef}\{f(x)^{mR}, x^i\}$, so that (4) can be restated as $A_m(w) \geq N_m(w)$. Step (ii) now follows by using the asymptotic expansion of $N_m(w)$ stated as Theorem 1 [12], where we define $\omega = w/(mR)$ and where x_ω is the unique positive solution to $a(x) = \omega$.

Finally, to see (iii) we replace w by $mRa(x_\omega)$ and thus we get the claim. Since this bound is valid for any $w \in \mathbb{N}$ so that $Dm + w \leq \frac{m}{2}$ we get the bound

$$\lim_{m\to\infty} \frac{1}{m} \log |\mathcal{C}(D)| \leq g(D, R),$$

where

$$g(D, R) = \inf_{\substack{x\geq 0 \\ D+a(x)R\leq\frac{1}{2}}} -R\log\frac{f(x)}{x^{a(x)}} + R + h(D + a(x)R).$$

Now note that as long as $g(D, R) < 1$, $|\mathcal{C}(D)|$ is exponentially small compared to 2^m. Therefore, looking back at (2) we see that in this case the average distortion converges to at least D in the limit $m \to \infty$. We get the tightest bound by looking for the condition for equality, i.e. by looking at the equation $g(R, D) = 1$. If we take the derivative with respect to x and set it to 0 then we get the condition

$$\frac{x}{1 + x} = D + Ra(x).$$

Recall that $D + a(x)R \leq \frac{1}{2}$, so that this translates to $x \leq 1$. This means that $x \leq 1$. Replace $D + a(x)R$ in the entropy term by $\frac{x}{1+x}$, set the resulting expression for $g(R, x)$ equal to 1, and solve for R. This gives R as a function of x and so we also get D as a function of x. We have

$$R(x) = \frac{1 - h(\frac{x}{1+x})}{1 - \log\frac{f(x)}{x^{a(x)}}},\ D(x) = \frac{x}{1 + x} - a(x)R(x).$$

A check shows that $x = 0$ corresponds to $(D, R) = (0, 1)$ and that $x = 1$ corresponds to $(D, R) = (\frac{L'-1}{2L'}, \frac{1}{(L')^2})$. Further, R and D are monotone functions of x. Recall that we are only interested in the bound for $R \in [\frac{1}{L'}, 1]$. We get the corresponding curve by letting x take values in $[0, x(\frac{1}{L'})]$. For smaller values of the rate we get the aforementioned straight-line bound.

Looking at the above expression for $g(D, R)$ one can see why this bound is strictly better than the rate-distortion curve for $D \in (0, \frac{1}{2})$. Assume at first that the generator degree distribution is regular. Let the degree be $\mathtt{l}$. In this case a quick check shows that $-R\log\frac{f(x)}{x^{a(x)}}$ is equal to $-Rh(\frac{a(x)}{\mathtt{l}})$. Since $a(0) = 0$ we get the rate distortion bound if we set $x = 0$. The claim follows by observing that $a(x)$ is a continuous strictly increasing function and that $h(x)$ has an infinite derivative at $x = 0$ while $h(D + a(x)R)$ has a finite derivative at $x = 0$. It follows that there exists a sufficiently small x so that $Rh(\frac{a(x)}{\mathtt{l}})$ is strictly larger than $h(D + a(x)R) - h(D)$ and so that $D + a(x)R \leq \frac{1}{2}$. Hence, $g(D, R)$ is strictly decreasing as a function of x at $x = 0$. This bounds the achievable distortion strictly away from the rate-distortion bound. The same argument applies to an irregular generator degree distribution; the simplest way to see this is to replace $\mathtt{l}$ by the maximum degree of $L(x)$.

IV. Bound Via Test Channel

Instead of using a combinatorial approach to bound $|\mathcal{C}(D)|$ one can also use a probabilistic argument using the "test channel" shown in Figure 5.

For the cases we have checked the resulting bound is numerically identical to the bound of Theorem 2 (excluding the straight-line portion). We restrict our exposition to the regular case. The generalization to the irregular case is straightforward.

Theorem 3 (Bound Via Test Channel): Let $\hat{\mathcal{S}}$ be an LDGM code with blocklength m, generator degree distribution $L(x) = x^{\mathtt{l}}$, and rate R. Then for any pair (R, D), where

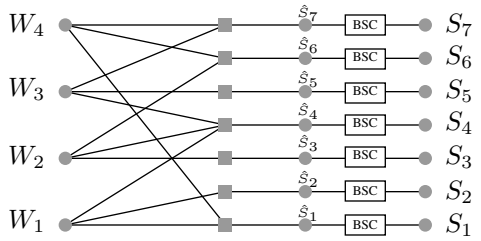

Fig. 5. The generator words W are chosen uniformly at random from $\mathcal{W}$. This generates a codeword $\hat{S}$ uniformly at random. Each component of $\hat{S}$ is then sent over a binary symmetric channel with transition probability D'.

D is the average distortion, we have

$$R \geq \sup_{D \leq D' \leq \frac{1}{2}} \frac{1 - h(D) - \mathrm{KL}(D\|D')}{1 - \log_2\Big(1 + \frac{(D')^{1}}{(1-D')^{1}}\Big)} \geq \frac{1 - h(D)}{1 - \log_2\Big(1 + \frac{D^{1}}{(1-D)^{1}}\Big)} > 1 - h(D),$$

where $\mathrm{KL}(D\|D') = D\log_2(D/D') + (1-D)\log_2((1-D)/(1-D'))$.

Proof. The same remark as in the proof of Theorem 2 applies: although the bound is valid for any blocklength it suffices to prove it for the limit of blocklengths tending to infinity. Also, for simplicity we have not stated the bound in its strengthened form which includes a straight-line portion. But the same technique that was applied in the proof of Theorem 2 applies also to the present case.

As remarked earlier, the idea of the proof is based on bounding $|\mathcal{C}(D)|$ by using the "test channel." More precisely, choose W uniformly at random from the set of all binary sequences of length mR. Subsequently compute $\hat{S}$ via $\hat{S} = WG$, where G is the generator matrix of the LDGM code. Finally, let $S = \hat{S} + Z$, where Z has iid components with $\mathbb{P}\{Z_i = 1\} = D'$.

Consider the set of sequences $s \in \mathcal{C}(D)$. For each such s we know that there exists an $\hat{s} \in \hat{\mathcal{S}}$ so that $\mathrm{d}(s,\hat{s}) \leq Dm$. We have

$$\begin{aligned}
&\mathbb{P}\{S = s \mid s \in \mathcal{C}(D)\} \\
&= \sum_{\hat{s}' \in \hat{\mathcal{S}}} \mathbb{P}\{S = s, \hat{S} = \hat{s}' \mid s \in \mathcal{C}(D)\} \\
&= \sum_{w=0}^{m} \sum_{\hat{s}' \in \hat{\mathcal{S}}: \mathrm{d}(\hat{s}',\hat{s}) = w} \mathbb{P}\{S = s, \hat{S} = \hat{s}' \mid s \in \mathcal{C}(D)\} \\
&= \sum_{w=0}^{m} A_m(w) \mathbb{P}\{S = s, \hat{S} = \hat{s}' \mid s \in \mathcal{C}(D), \mathrm{d}(\hat{s}', \hat{s}) = w\} \\
&= \sum_{w=0}^{m} A_m(w) 2^{-mR} \Big(\frac{D'}{1-D'}\Big)^{\mathrm{d}(s,\hat{s}')} (1-D')^m \\
&\geq \sum_{w=0}^{m} A_m(w) 2^{-mR} \Big(\frac{D'}{1-D'}\Big)^{\mathrm{d}(s,\hat{s})+\mathrm{d}(\hat{s},\hat{s}')} (1-D')^m \\
&\overset{\mathrm{d}(\hat{s}',\hat{s})=w}{=} \sum_{w=0}^{m} A_m(w) 2^{-mR} \Big(\frac{D'}{1-D'}\Big)^{\mathrm{d}(s,\hat{s})+w} (1-D')^m \\
&\overset{\mathrm{d}(s,\hat{s}) \leq Dm}{\geq} \sum_{w=0}^{m} A_m(w) 2^{-mR} \Big(\frac{D'}{1-D'}\Big)^{Dm+w} (1-D')^m \\
&= 2^{-mR - mh(D) - m\mathrm{KL}(D\|D')} \sum_{w=0}^{m} A_m(w) \Big(\frac{D'}{1-D'}\Big)^{w},
\end{aligned}$$

where $A_m(w)$ denotes the number of codewords in $\hat{\mathcal{S}}$ of Hamming weight w. Due to the linearity of the code this is also the number of codewords in $\hat{\mathcal{S}}$ of Hamming distance w from $\hat{s}$. Using summation by parts and setting $c = D'/(1-D') < 1$, we have

$$\begin{aligned}
&\sum_{w=0}^{m} A_m(w) c^w \\
&= c^{m+1} 2^{mR} + \sum_{w=0}^{m} \Big(\sum_{i=0}^{w-1} A_m(i)\Big)(c^w - c^{w+1}) \\
&\overset{(4)}{\geq} c^{m+1} 2^{mR} + \sum_{w=0}^{m} \Big(\sum_{i=0}^{\lfloor (w-1)/\mathtt{l} \rfloor} \binom{mR}{i}\Big)(c^w - c^{w+1}) \\
&= \sum_{w=0}^{\lfloor m/\mathtt{l} \rfloor} \binom{mR}{w} c^{\mathtt{l}w} + c^{m+1}\Big(2^{mR} - \sum_{i=0}^{\lfloor m/\mathtt{l} \rfloor} \binom{mR}{i}\Big) \\
&\geq \sum_{w=0}^{\lfloor m/\mathtt{l} \rfloor} \binom{mR}{w} c^{\mathtt{l}w} \geq \frac{1}{m}(1 + c^{\mathtt{l}})^{mR}.
\end{aligned}$$

The last step is valid as long as $\frac{Rc^{\mathtt{l}}}{1+c^{\mathtt{l}}} < \frac{1}{\mathtt{l}}$. In this case the maximum term (which appears at $\frac{Rc^{\mathtt{l}}}{1+c^{\mathtt{l}}} m$) is included in the sum (which goes to $m/\mathtt{l}$) and is thus greater than equal to the average of all the terms, which is $\frac{1}{m}(1+c^{\mathtt{l}})^{mR}$. This condition is trivially fulfilled for $R\mathtt{l} < 1$. Assume for a moment that it is also fulfilled for $R\mathtt{l} \geq 1$ and the optimum choice of D'. It then follows that

$$\mathbb{P}\{S = s \mid s \in \mathcal{C}(D)\} \geq \frac{1}{m} 2^{-m(R + h(D) + \mathrm{KL}(D\|D') - R\log_2(1+c^{\mathtt{l}}))}.$$

Since

$$\begin{aligned}
1 &= \sum_{s \in \mathbb{F}_2^m} \mathbb{P}\{S = s\} \geq \sum_{s \in \mathcal{C}(D)} \mathbb{P}\{S = s\} \\
&\geq |\mathcal{C}(D)| \frac{1}{m} 2^{-m(R + h(D) + \mathrm{KL}(D\|D') - R\log_2(1+c^{\mathtt{l}}))},
\end{aligned}$$

we have $|\mathcal{C}(D)| \leq m 2^{m(R + h(D) + \mathrm{KL}(D\|D') - R\log_2(1+c^{\mathtt{l}}))}$. Proceeding as in (2), we have

$$\begin{aligned}
&\mathbb{E}[\mathrm{d}(S, \mathrm{g}(\mathrm{f}(S)))] \geq D\big(1 - 2^{-m}|\mathcal{C}(D)|\big) \\
&\geq D\big(1 - m 2^{m(R + h(D) + \mathrm{KL}(D\|D') - R\log_2(1+c^{\mathtt{l}}) - 1)}\big).
\end{aligned}$$

We conclude that if for some $D \leq D' \leq \frac{1}{2}$, $R + h(D) + \mathrm{KL}(D\|D') - R\log_2(1 + \frac{(D')^{1}}{(1-D')^{1}}) - 1 < 0$ then the distortion is at least D. All this is still conditioned on $\frac{R\mathtt{l}c^{\mathtt{l}}}{1+c^{\mathtt{l}}} < 1$ for

the optimum choice of D'. For $R\mathtt{1} < 1$ we already checked this. So assume that $R\mathtt{1} \geq 1$. The above condition can then equivalently be written as $D' < \frac{1}{1+(R\mathtt{1}-1)^{\frac{1}{\mathtt{1}}}}$. On the other hand, taking the derivative of our final expression on the rate-distortion function with respect to D' we get the condition for the maximum to be $D' = \frac{1}{1+(1+\frac{R\mathtt{1}}{D'-D})^{\frac{1}{\mathtt{1}}}} < \frac{1}{1+(R\mathtt{1}-1)^{\frac{1}{\mathtt{1}}}}$. We see therefore that our assumption $\frac{R\mathtt{1}c^{\mathtt{1}}}{1+c^{\mathtt{1}}} < 1$ is also correct in the case $R\mathtt{1} \geq 1$.

Numerical experiments show that the present bound yields for the regular case identical results as plotting the curve corresponding to $g(D,R) = 1$, where $g(D,R)$ was defined in the proof of Theorem 2. This can be interpreted as follows. Choose D' equal to the optimal radius of the Hamming ball in the proof of Theorem 2. Then the points $\hat{s}'$ that contribute most to the probability of $S = s$ must be those that have a distance to $\hat{s}$ of $m(D' - D)$.

V. Discussion and Open Questions

In the preceding sections we gave two bounds. Both of them are based on the idea of counting the number of points that are "covered" by spheres centered around the codewords of an LDGM code. In the first case we derived a bound by double counting this number. In the second case we derived a bound by looking at a probabilistic model using the test channel.

An interesting open question is to determine the exact relationship of the test channel model to the rate-distortion problem. More precisely, it is tempting to conjecture that a pair (R,D) is only achievable if $H(S) = m$ in this test channel model. This would require to show that only elements of the typical set of $\mathcal{S}$ under the test channel model are covered, i.e., have code words within distance D. For the test channel model it is very easy to determine a criterion in the spirit of Gallager's original bound. We have

$$\begin{aligned} H(S) &= H(W) + H(S \mid W) - H(W \mid S) \\ &= mR + mh(D) - \sum_{g=1}^{mR} H(W_g \mid S, W_1, \ldots, W_{g-1}) \\ &\overset{\text{(i)}}{\leq} mR + mh(D) - \sum_{g=1}^{mR} H(W_g \mid S, W_{\sim g}) \\ &\overset{\text{(ii)}}{=} mR + mh(D) - \sum_{g=1}^{mR} H(W_g \mid S_g, W_{\sim g}), \end{aligned}$$

where S_g denotes the subset of the components of the S vectors which are connected to the generator g. Step (i) follows since conditioning decreases entropy. Step (ii) follows since knowing $(S_g, W_{\sim g})$, W_g is not dependent on $S_{\sim g}$. The term $H(W_g \mid S_g, W_{\sim g})$ represents the EXIT function of a repetition code when transmitting over BSC(D) channel. If one could show that $H(S) = m$ is a necessary condition for achieving average distortion of D then a quick calculation shows that the resulting bound would read

$$R \geq \frac{1 - h(D)}{1 - \sum_{i=0}^{\mathtt{1}} \binom{\mathtt{1}}{i} (1-D)^i D^{\mathtt{1}-i} \log_2\left(1 + \left(\frac{D}{1-D}\right)^{2i-1}\right)}.$$

This "bound" is similar in spirit to the original bound given by Gallager, except that in Gallager's original bound for LDPC codes we have a term corresponding to the entropy of single-parity check codes, whereas here we have terms that correspond to the entropy of repetition codes; this would be quite fitting given the duality of the problems.

Acknowledgment

We gratefully acknowledge the support by the Swiss National Science Foundation under grant number 200020-113412.

References

[1] E. Martinian and J. Yedidia, "Iterative quantization using codes on graphs," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Oct. 2003.

[2] S. Ciliberti and M. Mezard, "The theoretical capacity of the parity source coder," *J. Stat. Mech.*, 2005.

[3] S. Ciliberti, M. Mezard, and R. Zecchina, "Lossy data compression with random gates," *Phys. Rev. Lett.*, vol. 95, 2005.

[4] M. J. Wainwright and E. Maneva, "Lossy source coding via message-passing and decimation over generalized codewords of LDGM codes," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Adelaide, Australia, Sept. 2005, pp. 1493–1497.

[5] A. Braunstein and R. Zecchina, "Survey propagation as local equilibrium equations," *J. Statistical Mechanics: Theory and Experiment*, June 2004.

[6] T. Filler and J. Fridrich, "Binary quantization using belief propagation with decimation over factor graphs of ldgm codes," in *Proc. of the Allerton Conf. on Commun., Control, and Computing*, Sept. 2007.

[7] E. Martinian and M. J. Wainwright, "Low-density codes achieve the rate-distortion bound," in *Proc. of the Data Compression Conference*, Snowbird, UT, Mar. 2006.

[8] ——, "Low-density constructions can achieve the Wyner-Ziv and Gelfand-Pinsker bounds," in *Proc. of the IEEE Int. Symposium on Inform. Theory*, Seattle, WA, USA, July 2006, pp. 484–488.

[9] ——, "Analysis of LDGM and compound codes for lossy compression and binning," in *Proc. of the IEEE Inform. Theory Workshop*, San Diego, CA, USA, Feb. 2006.

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

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[12] D. Burshtein and G. Miller, "Asymptotic enumeration methods for analyzing LDPC codes," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1115–1131, June 2004.