Published in: Proceedings of the Data Compression Compression,
Snowbird, Utah. March 2006

Low density codes achieve the rate-distortion bound

Emin Martinian
Mitsubishi Electric Research Labs
Cambridge, MA 02139
Email: {martinian}@merl.com

Martin Wainwright
University of California at Berkeley
Berkeley, CA 94720
Email: {wainwrig}@eecs.berkeley.edu

Abstract: We propose a new construction for low-density source codes with multiple parameters that can be tuned to optimize the performance of the code. In addition, we introduce a set of analysis techniques for deriving upper bounds for the expected distortion of our construction, as well as more general low-density constructions. We show that (with an optimal encoding algorithm) our codes achieve the rate-distortion bound for a binary symmetric source and Hamming distortion. Our methods also provide rigorous upper bounds on the minimum distortion achievable by previously proposed low-density constructions.

1 Introduction

While low-density parity check (LDPC) codes can provably approach the channel coding capacity [10] for point-to-point transmission, currently there are relatively few theoretical results on low-density codes for lossy source coding, channel coding with encoder side information, and source coding with decoder side information. Note that the latter three scenarios all involve some aspect of quantization. Even though quantization and error correction are closely related, the standard LDPC constructions used for channel coding generally fail [8]. One viable option is trellis-based quantization (TCQ) [7], which has been used both for lossy source coding, as well as for distributed source coding [13, 2, 14]. However, saturating fundamental bounds with TCQ requires taking the constraint length to infinity [11], which incurs exponential complexity even for message-passing decoders/encoders. Consequently, it is of considerable interest to develop low-density constructions that are also capable of saturating the information-theoretic bounds.

Previous work [8] has shown that low-density generator matrix (LDGM) codes, which are dual to LDPC codes, are provably optimal for binary erasure quantization (a special type of source coding). This motivates the use of LDGM codes and variants for more general compression problems. Indeed, recent work [3, 9, 12] has shown empirically that LDGM codes, in conjunction with variants of sum-product message-passing for encoding, can approach the rate-distortion bound for a binary symmetric source (BSS). In addition, non-rigorous replica or cavity method calculations [3, 9] also suggest that the theoretical performance of LDGM codes is close to optimal.

Emin Martinian was supported by Mitsubishi Electric Research Labs while Martin Wainwright was supported by NSF Grant DMS-0528488 and an Alfred P. Sloan Foundation Fellowship.

This paper makes two primary contributions to this area. First, we propose a new low-density construction for lossy source coding with multiple parameters that can be tuned to optimize the performance of the code. Our construction includes as a special case the ordinary LDGM codes examined previously [3, 8, 9, 12]. Second, we develop methods useful for analyzing the expected distortion of our constructions as well as more general lossy source codes. Using these methods, we show that (with optimal quantization) our codes saturate the rate-distortion bound for a uniform binary source and Hamming distortion. As we will show in a longer version of this paper, our methods also lead to rigorous upper bounds on the distortion achievable by a standard LDGM construction. Thus, provided that a low complexity iterative encoding algorithm can be found, our results suggest that low density codes can provide significant improvements for a wide class of quantization problems.

The remainder of this paper is organized as follows. After introducing some notation, we describe our new low density generator matrix construction in Section 2 and bound its performance in Section 3. In particular, we describe the tools required to analyze low density source codes through a series of lemmas, which we believe illustrate the main insights of the paper. Finally, we close with some concluding remarks in Section 4 and postpone all proofs to the appendix.

Notation: Vectors/sequences are denoted in bold (*e.g.*, $\mathbf{s}$), random variables in sans serif font (*e.g.*, s), and random vectors/sequences in bold sans serif (*e.g.*, $\mathbf{s}$). Similarly, matrixes are denoted using bold capital letters (*e.g.*, $\mathbf{G}$) and random matrixes with bold sans serif capitals (*e.g.*, $\mathbf{G}$). We use $I(\cdot;\cdot)$, $H(\cdot)$, and $D(\cdot||\cdot)$ to denote mutual information, entropy, and relative entropy (Kullback-Leibler distance), respectively. Finally, we use $\operatorname{card}\{\cdot\}$ to denote the cardinality of a set, $||\cdot||_p$ to denote the p-norm of a vector, and $H_b(t)$ to denote the entropy of a Bernoulli(t) random variable.

2 The Compound Construction

The construction considered in the paper is illustrated in Fig. 1: the top section consists of an LDGM code $\mathbf{C}_{\mathrm{t}}$ of rate $R_t = \frac{m}{n}$ with n source bits and m information bits, whereas the bottom section consists of an LDPC code of rate $R_b = 1 - \frac{k}{m}$ with m bits constrained by k checks. The compound code formed by joining the top and bottom code can generate $2^{R_b m} = 2^{m-k}$ possible source reconstructions of length n, so that the overall code $\mathbf{C}$ has rate $R = R_t R_b$. Note that a check-regular LDGM code corresponds to the special case of setting $R_b = 1$.

To quantize a length n binary source vector $\mathbf{s}$ using the compound construction, an encoder finds an assignment for the m bits in the middle layer that satisfy the constraints of the bottom LDPC code. Formally, we can denote the m-by-n generator matrix for the top LDGM code as $\mathbf{G}$ and the k-by-m parity check matrix for the bottom LDPC code as $\mathbf{H}$. Then $\mathbf{q}$ is a codeword of the overall code if $\mathbf{q} = \mathbf{w}\,\mathbf{G}$ and $\mathbf{H}\,\mathbf{w}' = 0$ for some assignment of the middle layer, which we denote as $\mathbf{w}$. Thus, an optimal encoder for $\mathbf{s}$ would find the codeword minimizing the Hamming

distance, $d_H(\mathbf{wG}, \mathbf{s})$, such that $\mathbf{Hw}' = 0$. Since the vector $\mathbf{w}$ has length m, storing or transmitting $\mathbf{w}$ directly would achieve only compression rate R_t. Instead, we can use the fact that there are only 2^{m-k} valid choices for $\mathbf{w}$, to store $\mathbf{w}$ using only k bits, resulting in compression rate R. For example, we could store the k-bit information vector that when encoded with the bottom LDPC code yields $\mathbf{w}$.

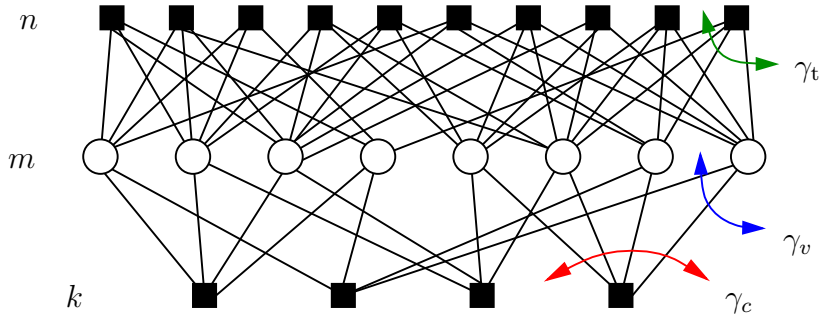

Figure 1. Illustration of the compound code construction, involving an LDGM (top section) with $\gamma_t = 4$ and an LDPC (bottom section) with $(\gamma_v, \gamma_c) = (2, 4)$.

Random LDPC Ensemble: For the bottom LDPC code, we use the standard (γ_v, γ_c)-regular LDPC ensemble studied by Gallager [5]. Specifically, each of the m variable nodes in the middle layer connects to γ_v check nodes in the bottom layer. Similarly, each of the k check nodes in the bottom layer connects to γ_c variable nodes in the middle layer. For convenience, we restrict ourselves to even check degrees γ_c. Note that these degrees are linked to the rate via the relation $\frac{\gamma_v}{\gamma_c} = 1 - R_b$. A random LDPC code $\mathbf{C}_\mathrm{b} \equiv \mathbf{C}_\mathrm{b}(\gamma_v, \gamma_c)$ is generated by choosing uniformly from this ensemble.

Random LDGM ensemble: For the top LDGM code, each of the n checks at the top are randomly connected to γ_t variable nodes in the middle layer chosen uniformly at random. This leads to a Poisson degree distribution on the information bits and makes the resulting distribution of a random codeword easy to characterize:

Lemma 1. *Let* G *be a random generator matrix obtained by placing* γ_t *ones in each column uniformly at random. Then for any vector* $\mathbf{w} \in \{0,1\}^m$ *with a fraction of* v *ones, the distribution of the corresponding codeword* $\mathbf{w}\,\mathsf{G}$ *is Bernoulli*$(\delta(v; \gamma_t))$ *where*

$$\delta(v; \gamma_t) = \frac{1}{2} \cdot \left[1 - (1 - 2v)^{\gamma_t}\right]. \tag{1}$$

3 Main Results

Although our methods apply to the compound construction more generally, we state our main result in application to the special case with $R_t = 1$ and $R_b = \mathrm{R}$. For these choices, we can guarantee that our compound construction approaches the optimal rate-distortion trade-off as the blocklength n tends to infinity using finite

choices of degrees in our LDGM/LDPC construction.[1]

Theorem 1. *Consider an arbitrary rate distortion pair* $(D, \mathrm{R}(D))$*. For any* $\Delta > 0$*. there exists a finite LDGM degree* $\gamma_{\mathrm{t}}(\Delta, D)$ *and an LDPC code with finite degrees* $\gamma_v(\Delta, D)$ *and* $\gamma_c(\Delta, D)$ *such that a randomly chosen code with rate* $\mathrm{R} \stackrel{\Delta}{=} \mathrm{R}(D) + \Delta$ *in the associated LDGM ensemble achieves distortion* D *with probability* $1 - \exp(-cn)$ *for some constant* c*.*

As a particular example of our results illustrated later in Fig. 2, the degree choices $\gamma_{\mathrm{t}} = 4$, $\gamma_{\mathrm{c}} = 8$ for rate $\mathrm{R}(D) = 1/2$, are sufficient to make the gap Δ zero (within the precision of our numerical calculations). The proof of Theorem 1 consists of several steps, which we motivate and describe in the following text. Proofs of these auxiliary results are provided in the appendix.

3.1 Expected Number of Good Codewords

For a length n code $\mathbf{C}$ and a source vector $\mathbf{s}$, we define $z(\mathbf{C}, \mathbf{s}, D)$ to be the number of codewords that are within Hamming distance Dn of $\mathbf{s}$. Specifically, let $x_i(\mathbf{C}, \mathbf{s}, D)$ be 1 if the ith codeword in the compound code $\mathbf{C}$ is within Hamming distance Dn of the source $\mathbf{s}$, and 0 otherwise. Then

$$z(\mathbf{C}, \mathbf{s}, D) \stackrel{\Delta}{=} \sum_i x_i(\mathbf{C}, \mathbf{s}, D). \tag{2}$$

Ideally, $z(\mathbf{C}, \mathbf{s}, D)$ should be large and there should be many good codewords provided that the rate exceeds the rate-distortion function: $\mathrm{R} > 1 - H_b(D)$. Specifically, if we consider a random source vector $\mathbf{s}$ and a randomly generated code $\mathbf{C}$, then the probability that the code is successful is simply[2] $\Pr[\mathsf{z}(D) > 0]$.

Since analyzing this probability directly is generally difficult, most random coding arguments [4] consider the expectation $E[\mathsf{z}(D)]$. For essentially any code (and in particular the compound construction), it is possible to show that the expected number of good codewords is large:

Lemma 2.

$$E[\mathsf{z}(D)] \geq \frac{1}{n+1} 2^{n\left(\mathrm{R} - \left[1 - H(D)\right]\right)}. \tag{3}$$

3.2 Typical Number of Good Codewords

Unfortunately, the fact that the expected number of codewords is large is insufficient to show that the code achieves the rate-distortion bound. Rather, in order to show that the code is good, we must show that the typical number of good codewords is not too far from the expected number of good codewords (or at least non-zero).

[1] Our methods also yield upper bounds on the achievable distortion of the check-regular LDGM construction ($R_t = \mathrm{R}$ and $R_b = 1$). Subsequent work will describe the use of alternative rate pairs (R_t, R_t) for source and channel coding with side information.

[2] When the source and/or code are random, then we drop the indexing of random quantities and write $x_i(\mathbf{C}, \mathbf{s}, D)$ and $z(\mathbf{C}, \mathbf{s}, D)$ as random variables $\mathsf{x}_i(D)$ and $\mathsf{z}(D)$.

For high density codes, this can be done by using Chebyshev's inequality, which depends on the variance of $z(D)$. For most low density constructions (including our own), the variance is too large for Chebyshev's inequality to yield a useful bound. Consequently, we instead use Shepp's second moment method as summarized in the following proposition:[3]

Proposition 1. *For any positive integer valued random variable* z, $\Pr[z > 0] \geq \frac{\mathrm{E}[z]^2}{\mathrm{E}[z^2]}$.

To show that there is typically at least one good codeword, we must upper bound $\mathrm{E}[z(D)^2]$, which can be cast in a more useful form using the following lemma:

Lemma 3.

$$\mathrm{E}[z(D)^2] = E[z(D)] + E[z(D)] \cdot \left\{ \sum_{j \neq 0} \Pr[x_j(D) = 1 \mid x_0(D) = 1] \right\} \tag{4}$$

Lemma 3 illustrates one of the main differences between low density and high density constructions. Specifically, in a high density construction, each codeword can be chosen independently yielding $\Pr[x_j(D) = 1 \mid x_0(D) = 1] = \Pr[x_j(D) = 1]$ and implying $\mathrm{E}[z(D)^2] \leq E[z(D)] + E[z(D)]^2$. In contrast, for low density codes, there will usually be some dependence between the codewords. For example, in the usual LDGM construction, when the information bits $\mathbf{w}$ have low weight, then the resulting codeword $\mathbf{w}\,\mathbf{G}$ will also have low weight. Consequently, if the all-zero codeword is within Hamming distance Dn of the source, then these low weight codewords probably are as well and so $\Pr[x_j(D) = 1 \mid x_0(D) = 1]$ can be much larger than $\Pr[x_j(D) = 1]$. In particular, we can bound $\Pr[x_j(D) = 1 \mid x_0(D) = 1]$ by considering the weight of the information sequence $\mathbf{w}_j$ used to generate the jth codeword:

Lemma 4. *Let* $\mathbf{w}_j\,\mathbf{G}$ *be the* j*th codeword obtained by multiplying a weight* v_j *vector* $\mathbf{w}_j$ *by a random matrix from the LDGM ensemble and let* $x_j(D)$ *denote the event that codeword* j *is within Hamming distance* Dn *of a random Bernoulli*$(1/2)$ *source. Then for any even degree* γ_{t}*, letting* $v_0 = 0$ *yields*

$$\Pr[x_j(D) = 1 \mid x_0(D) = 1] \leq \begin{cases} 1 & \text{if } 0 \leq v_j \leq \nu^*(D; \gamma_{\mathrm{t}}) \\ 2^{-n\,\mathrm{KL}(D \,\|\, \delta(v_j; \gamma_{\mathrm{t}}))} & \text{otherwise,} \end{cases} \tag{5}$$

where

$$\nu^*(D; \gamma_{\mathrm{t}}) = \frac{1}{2}\left[1 - \left(1 - 2D\right)^{\frac{1}{\gamma_{\mathrm{t}}}}\right]. \tag{6}$$

Lemma 4 shows that $\Pr[x_j(D) = 1 \mid x_0(D) = 1]$ is small whenever the weight of the information sequence for a codeword is large. So to characterize the sum over this probability we must consider how many vectors of a given weight in the middle layer satisfy the constraints of the bottom LDPC code $\mathbf{C}_{\mathrm{b}}$. Specifically, we denote

[3] Proposition 1 can be established by defining an indicator random variable, $r(D)$, for the event $\{z(D) > 0\}$ and applying the Cauchy-Schwartz inequality to obtain $\mathrm{E}[z(D)]^2 = \mathrm{E}[z(D)r(D)]^2 \leq \mathrm{E}[z(D)^2] \cdot \mathrm{E}[r(D)]$, which is equivalent to the desired result.

the average (log domain) weight enumerator of $\mathbf{C}_\mathrm{b}$ (*i.e.*, the rate of codewords of $\mathbf{C}_\mathrm{b}$ with a given weight) as

$$\mathcal{A}_{\mathbf{C}_\mathrm{b}}(\omega) \triangleq \frac{1}{n} \operatorname{card} \{\mathbf{q} \in \mathbf{C}_\mathrm{b} \mid ||\mathbf{q}||_1 = \omega \cdot n\}. \tag{7}$$

Intuitively, by combining (7) with Lemma 4, we can bound the term in braces of (4):

$$\sum_{j \neq 0} \Pr[\mathsf{x}_j(D) = 1 \mid \mathsf{x}_0(D) = 1] \leq \sum_{t=1}^{\nu^*(D;\gamma_\mathrm{t})} 2^{\mathcal{A}_{\mathbf{C}_\mathrm{b}}(t/n)} + \sum_{t=\nu^*(D;\gamma_\mathrm{t})}^{n} 2^{n\left[\mathcal{A}_{\mathbf{C}_\mathrm{b}}(t/n) - \mathrm{KL}(D \,\|\, \delta(t;\gamma_\mathrm{t}))\right]}. \tag{8}$$

Formally, we can use this idea to obtain the following result:

Theorem 2. *Consider a sequence of rate* R *compound codes of increasing blocklength* n*. Suppose that the following inequality holds for all sufficiently large blocklengths:*

$$\mathrm{R} - [1 - H_b(D)] > \frac{1}{n} \log \left\{ \sum_{t=1}^{\nu^*(D;\gamma_\mathrm{t})} 2^{\mathcal{A}_{\mathbf{C}_\mathrm{b}}(\frac{t}{n})} + \sum_{t=\nu^*(D;\gamma_\mathrm{t})}^{n} 2^{n\left[\mathcal{A}_{\mathbf{C}_\mathrm{b}}(\frac{t}{n}) - \mathrm{KL}(D \,\|\, \delta(\frac{t}{n};\gamma_\mathrm{t}))\right]} \right\} \tag{9}$$

Then the probability that a code in the sequence fails to quantize a source with distortion at most Dn *goes to zero as* $n \to \infty$*.*

3.3 Reducing Dependency Between Codewords

The bracketed term on the RHS of (9) corresponds to the *excess rate* required beyond the minimum $1 - H_b(D)$ and is plotted in Fig. 2 for the compound code in Fig. 1. The first term represents the number of low weight codewords of the bottom code. Since the bound from Lemma 4 does not become active until weight $\nu^*(D;\gamma_\mathrm{t})$, making the first term negligible requires choosing the LDPC ensemble so that the minimum distance is greater than the weight $\nu^*(D;\gamma_\mathrm{t})$ resulting from the choice of the degree γ_t in the LDGM ensemble. The exponent of the second term in (9) is the sum of the weight enumerator and the bound from Lemma 4. For this term to be negligible, the bottom LDPC code must have a weight enumerator that grows less quickly than the error exponent in (5).

Using the exact formula for the asymptotic weight enumerator of regular LDPC codes developed by Litsyn and Shevelev [6], it is possible to prove the following result:[4]

Proposition 2. *There exist choices for* γ_t*,* γ_v*, and* γ_c *such that the term in braces in (9) becomes negligible.*

[4]The proof essentially requires showing that the sum of the weight enumerator and the bound from Lemma 4 is negative for all weights in $[\nu^*(D;\gamma_\mathrm{t}), 1/2]$. This can be done by checking the appropriate derivatives of the sum and is omitted for brevity.

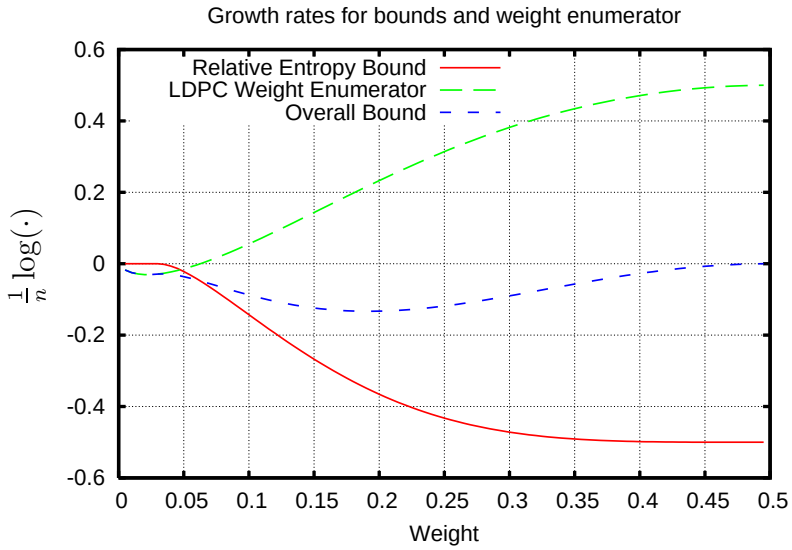

Figure 2. Log of bounds and weight enumerator for R $= 1/2$, $\gamma_{\mathrm{t}} = 4$, $\gamma_{\mathrm{c}} = 8$, at distortion $D \approx 0.11$ normalized by the blocklength n. The relative entropy bound from (5) is zero for weights below $\nu^*(D;\gamma_{\mathrm{t}})$ and then quickly goes to $2^{-n/2}$. The (log domain) weight enumerator for a regular rate $1/2$ LDPC code is negative for weights below the minimum distance and then rises to $2^{n/2}$. As long as the relative entropy bound is stronger than the weight enumerator, the excess rate in (9) of Theorem 2 will be negligible.

4 Concluding Remarks

In this paper, we proposed a new construction for low density source codes and introduced tools to analyze low density generator matrix codes. As stated in Lemma 4 and illustrated in Fig. 2, our main insight was that the source coding performance of a low density code can be bounded by considering the weight of the codewords. Thus, by using a compound code to control the weight spectrum we obtained codes that approach the rate-distortion function. A future paper will describe and analyze these types of compound constructions in application to source and channel coding with side information.

A Proofs

Proof of Lemma 1: By construction of the LDGM ensemble for G, each bit of the codeword $\mathbf{w}\,\mathsf{G}$ is independent of the others and is the modulo-2 sum of γ_{t} randomly and independently selected bits of $\mathbf{w}$. So the resulting codeword has a Bernoulli distribution and all that remains is to determine the probability that a given bit is one, which we denote as $\delta(v;\gamma_{\mathrm{t}})$.

For any output bit, let the random variable e_i denote whether the ith one in a column of G occurs in a position where $\mathbf{w}$ has a one (*i.e.*, e_i is the value of the variable node connected to the ith link of a given check node at the top of Fig. 1). Then $\delta(v;\gamma_{\mathrm{t}})$ is exactly the probability that $\sum_{i=1}^{\gamma_{\mathrm{t}}} \mathsf{e}_i$ is even. Letting $\Delta_v(z)$ denote

the generating function (*i.e.*, the z-transform) of $\sum_{i=1}^{\gamma_t} \mathsf{e}_i$ yields

$$\delta(v;\gamma_t) - (1-\delta(v;\gamma_t)) = \Delta_v(z=-1) = \prod_{i=1}^{\gamma_t} \left(\Pr[\mathsf{e}_i = 0] + z^{-1}\Pr[\mathsf{e}_i = 1]\right)\Big|_{z=-1} \tag{10}$$

$$= (1 - v + v\cdot z^{-1})^{\gamma_t}\Big|_{z=-1} = (1-2v)^{\gamma_t}. \tag{11}$$

Equating the leftmost term of (10) and the rightmost term of (11) and solving for $\delta(v;\gamma_t)$ yields the desired result. □

Proof of Lemma 2:

$$E[\mathsf{z}(D)] = E\left[\sum_i \mathsf{x}_i(D)\right] = \sum_i E[\mathsf{x}_i(D)] = \sum_i \Pr[d_H(\mathbf{q}_i, \mathbf{s}) \le Dn] \tag{12}$$

$$\ge \sum_i \frac{2^{-n\,\mathrm{KL}(D\,\|\,1/2)}}{(n+1)^2} = \sum_i \frac{2^{-n[1-H_b(D)]}}{(n+1)^2} = \frac{2^{n\{R-[1-H_b(D)]\}}}{(n+1)^2} \tag{13}$$

The first line follows by repeatedly expanding the definition of the random variables $\mathsf{z}(D)$ and $\mathsf{x}_i(D)$. For the next line, we lower bound the probability that a given codeword $\mathbf{q}_i$ is within distortion Dn using standard large deviations results (Theorem 12.1.4, [4]). Note that nothing in this argument depends on the actual code construction itself (except for the number of codewords). □

Proof of Lemma 3:

$$\mathrm{E}[\mathsf{z}(D)^2] = \mathrm{E}\left[\sum_i \sum_j \mathsf{x}_i(D)\mathsf{x}_j(D)\right] = E[\mathsf{z}(D)] + \sum_i \sum_{j\ne i} \mathrm{E}\left[\mathsf{x}_i(D)\mathsf{x}_j(D)\right] \tag{14}$$

$$= E[\mathsf{z}(D)] + \sum_{\mathbf{s}} \sum_i \sum_{j\ne i} \Pr[d_H(\mathbf{q}_j, \mathbf{s}) \le Dn, d_H(\mathbf{q}_i, \mathbf{s}) \le Dn] \Pr[\mathbf{s} = \mathbf{s}] \tag{15}$$

$$= E[\mathsf{z}(D)] + \sum_{\mathbf{s}} \sum_i \sum_{j\ne i} \Pr[d_H(\mathbf{q}_j \oplus \mathbf{q}_i, \mathbf{s} \oplus \mathbf{q}_i) \le Dn,$$

$$d_H(0, \mathbf{s} \oplus \mathbf{q}_i) \le Dn] \Pr[\mathbf{s} = \mathbf{s}] \tag{16}$$

$$= E[\mathsf{z}(D)] + \sum_{\mathbf{s}} \sum_i \sum_{j'\ne 0} \Pr[d_H(\mathbf{q}_{j'}, 0) \le Dn, d_H(0, \mathbf{s}') \le Dn] \Pr[\mathbf{s} = \mathbf{s}] \tag{17}$$

$$= E[\mathsf{z}(D)] + \sum_i \sum_{j'\ne 0} \Pr[\mathsf{x}_{j'}(D) = 1, \mathsf{x}_0(D) = 0] \tag{18}$$

$$= E[\mathsf{z}(D)] + \left\{\sum_i \Pr[\mathsf{x}_0(D) = 1]\right\} \cdot \left\{\sum_{j'\ne 0} \Pr[\mathsf{x}_{j'}(D) = 1 | \mathsf{x}_0(D) = 1]\right\} \tag{19}$$

To obtain (14) we consider the diagonal terms separately from the off-diagonal terms, and note that $E[\mathsf{x}_i(D)^2] = E[\mathsf{x}_i(D)]$ since the $\mathsf{x}_i(D)$ are indicator variables. Next,

we apply the definition of $\mathsf{x}_i(D)$ to get (15) and then add $\mathbf{q}_i$ to each side of the $d_H(\cdot,\cdot)$ terms to obtain (16). Since the code is linear, adding the codewords $\mathbf{q}_i$ and $\mathbf{q}_j$ yields another codeword which we denote $\mathbf{q}_{j'}$. This observation combined with writing $\mathbf{s}' = \mathbf{s} \oplus \mathbf{q}_i$ yields (17). To go from (17) to (18), we note that for a uniformly random source, $\Pr[\mathbf{s} = \mathbf{s}] = \Pr[\mathbf{s} = \mathbf{s}']$. Finally, to obtain the desired result from (19), we note that $\mathsf{x}_i(D)$ is independent of i and hence $E[\mathsf{x}_i(D)] = E[\mathsf{x}_0(D)]$. □

Proof of Lemma 4: We focus on the case when $\delta(v_j;\gamma_{\mathrm{t}}) \geq D$. Solving this relation for v_j yields the formula for $\nu^*(D;\gamma_{\mathrm{t}})$ in (6) and so $\delta(v_j;\gamma_{\mathrm{t}}) < D$ corresponds to the trivial bound in the top case of (5). Thus, for $\delta(v_j;\gamma_{\mathrm{t}}) \geq D$ we have

$$\Pr\left[\mathsf{x}_j(D) = 1 \mid \mathsf{x}_0(D) = 1\right] \leq \Pr\left[d_H(\mathbf{q}_j, \mathbf{s}) \leq Dn \mid d_H(\mathbf{s}, 0) \leq Dn\right] \tag{20}$$

$$\overset{(a)}{\leq} \max_{t \leq Dn} \Pr\left[d_H(\mathbf{q}_j, 1^t 0^{n-t}) \leq Dn\right] \overset{(b)}{\leq} \Pr\left[d_H(\mathbf{q}_j, 0^n) \leq Dn\right] \overset{(c)}{\leq} 2^{-n\,\mathrm{KL}(D \,\|\, \delta(v_j;\gamma_{\mathrm{t}}))}. \tag{21}$$

We obtain (20) from the definition of the random variable $\mathsf{x}_j(D)$. For (a), since $\mathbf{q}_j$ is a Bernoulli sequence, without loss of generality we can imagine that all the ones in $\mathbf{s}$ occur at the start of the sequence. To obtain an upper bound, we put in as many such ones as required to maximize the desired probability. In (b), we note that $\delta(v_j;\gamma_{\mathrm{t}}) \leq 1/2$ implies that it is more likely that a given position of $\mathbf{q}_j$ is zero than one so $t = 0$ gives the largest value for the maximization. Finally, to obtain (c), we apply Sanov's Theorem (Theorem 12.1.4, [4]). Note that the reason we required $\delta(v_j;\gamma_{\mathrm{t}}) \geq D$ originally is that this condition is required by Sanov's Theorem in (c). □

Before proving Theorem 2, we require the following lemma:

Lemma 5. *For a compound code that satisfies (9),* $\Pr[\mathsf{z}(D) > 0] > (1/2) \cdot (n+1)^{-2}$.

Proof. First, assume that $\sum_{j \neq 0} \Pr[\mathsf{x}_j(D) \mid \mathsf{x}_0(D)] \geq 1$ because if this is not the case then (4) immediately implies that $\Pr[\mathsf{z}(D) > 0] \geq 1/2$ and the proof is complete. Therefore continuing from the assumption that $\sum_{j \neq 0} \Pr[\mathsf{x}_j(D) \mid \mathsf{x}_0(D)] \geq 1$ yields

$$\Pr[\mathsf{z}(D) > 0] \overset{(a)}{\geq} \frac{\mathrm{E}[\mathsf{z}(D)]^2}{\mathrm{E}[\mathsf{z}(D)^2]} \overset{(b)}{=} \frac{\mathrm{E}[\mathsf{z}(D)]^2}{\mathrm{E}[\mathsf{z}(D)]\{1 + \sum_{j \neq 0} \Pr[\mathsf{x}_j(D) \mid \mathsf{x}_0(D)]\}} \tag{22}$$

$$\overset{(c)}{\geq} \frac{\mathrm{E}[\mathsf{z}(D)]^2}{2 \cdot \mathrm{E}[\mathsf{z}(D)] \cdot \{\sum_{j \neq 0} \Pr[\mathsf{x}_j(D) | \mathsf{x}_0(D)]\}} = \frac{\mathrm{E}[\mathsf{z}(D)]/2}{\sum_{j \neq 0} \Pr[\mathsf{x}_j(D) | \mathsf{x}_0(D)]} \tag{23}$$

$$\overset{(d)}{\geq} \frac{\mathrm{E}[\mathsf{z}(D)]/2}{2^{n\{\mathrm{R} - [1 - H_b(D)]\}}} \overset{(e)}{\geq} \frac{2^{n\{\mathrm{R} - [1 - H_b(D)]\}}}{2(n+1)^2 \cdot 2^{n\{\mathrm{R} - [1 - H_b(D)]\}}} = \frac{1}{2(n+1)^2} \tag{24}$$

where (a) follows from Proposition 1, (b) comes from Lemma 3, (c) follows from the assumption in the first sentence, (d) comes from (9), and (e) comes from Lemma 2. □

Proof of Theorem 2: Lemma 5 tells us that the probability that at least codeword is found within distortion D is at at least $(1/2)/(n+1)^2$, *i.e.*, there is at least a small chance that a good codeword exists. The key insight of the remainder of the proof is that while $(1/2)/(n+1)^2$ may be small, it is not *exponentially* small. Hence if we can show that the distortion for a compound code is concentrated near its typical value except with some *exponentially* small probability, then Lemma 5 immediately implies that the event $\{\mathsf{z}(D) > 0\}$ must correspond to the typical distortion. To prove exponential concentration, we show that the actual error probability, $\Pr[\mathsf{z}(D) = 0]$ is smaller than e^{-cn} for some constant c using martingale arguments [1, 10].

Specifically, we define a Doob martingale $\mathsf{m}_i(\mathbf{C}_\mathrm{b})$ that is the expected value of the distortion between the best codeword and the source (conditioned on the bottom code $\mathbf{C}_\mathrm{b}$) when the first i columns of the generator matrix $\mathbf{G}$ (*i.e.*, the connections from the first i checks to their respective variables) of the top code in Fig. 1 have been revealed. Going from step i to $i+1$ and revealing check $i+1$ can only change the value of the martingale $\mathsf{m}_i(\mathbf{C}_\mathrm{b})$ by at most 1. Hence, by the Azuma-Hoeffding inequality, the probability that a sample path of the martingale differs from its expected value by more than ϵ is less than $2e^{-n\epsilon^2}$.

Since Lemma 5 shows that the probability that $\Pr[\mathsf{z}(D) > 0]$ is at least an inverse polynomial (and hence *not* exponentially small), the event $\{\mathsf{z}(D) > 0\}$ must determine the expected value of the martingale. Therefore other events that result in a distortion larger than D (*e.g.*, $\{\mathsf{z}(D) = 0\}$) must be exponentially small. □

Proof of Theorem 1: Combining Theorem 2 with Proposition 2 establishes this result. □

References

[1] N. Alon and J. Spencer. *The Probabilistic Method*. Wiley Interscience, New York, 2000.

[2] J. Chou, S. S. Pradhan, and K. Ramchandran. Turbo and trellis-based constructions for source coding with side information. In *Data Compression Conference*, 2003.

[3] S. Ciliberti, M. Mézard, and R. Zecchina. Message-passing algorithms for non-linear nodes and data compression. Technical report, November 2005. arXiv:cond-mat/0508723.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.

[5] R. G. Gallager. *Low-density parity-check codes*. MIT Press, Cambridge, MA, 1963.

[6] S. Litsyn and V. Shevelev. On ensembles of low-density parity-check codes: asymptotic distance distributions. *IEEE Trans. Info. Theory*, 48(4):887–908, April 2002.

[7] M. W. Marcellin and T. R. Fischer. Trellis coded quantization of memoryless and Gauss-Markov sources. *IEEE Trans. Comm.*, 38(1):82–93, 1990.

[8] E. Martinian and J. S. Yedidia. Iterative quantization using codes on graphs. In *Proc. Allerton Conf. Comm., Control, and Computing*, Monticello, IL, Oct. 2003.

[9] T. Murayama. Thouless-Anderson-Palmer approach for lossy compression. *Physical Review E*, 69:035105(1)–035105(4), 2004.

[10] T. J. Richardson and R. L. Urbanke. The capacity of low-density parity-check codes under message-passing decoding. *IEEE Trans. Info. Theory*, 47(2):599–618, Feb. 2001.

[11] A. J. Viterbi and J. K. Omura. Trellis encoding of memoryless discrete-time sources with a fidelity criterion. *IEEE Trans. Info. Theory*, IT-20(3):325–332, 1974.

[12] M. J. Wainwright and E. Maneva. Lossy source coding via message-passing and decimation over generalized codewords of LDGM codes. In *Proc. ISIT*, Sep. 2005.

[13] X. Wang and M. T. Orchard. Design of trellis codes for source coding with side-information. In *Data Compression Conference*, 2001.

[14] Y. Yang, V. Stankovic, Z. Xiong, and W. Zhao. On multiterminal source code design. In *Data Compression Conference*, pages 43–52, 2005.