

Low-density constructions can achieve the Wyner-Ziv and Gelfand-Pinsker bounds

Emin Martinian

Mitsubishi Electric Research Labs
Cambridge, MA 02139, USA
Email: martinian@merl.com

Martin J. Wainwright

Dept. of Statistics and Dept. of EECS,
UC Berkeley, Berkeley, CA 94720
Email: wainwrig@{eecs,stat}.berkeley.edu

Abstract—We describe and analyze sparse graphical code constructions for the problems of source coding with decoder side information (the Wyner-Ziv problem), and channel coding with encoder side information (the Gelfand-Pinsker problem). Our approach relies on a combination of low-density parity check (LDPC) codes and low-density generator matrix (LDGM) codes, and produces sparse constructions that are simultaneously good as both source and channel codes. In particular, we prove that under maximum likelihood encoding/decoding, there exist low-density codes (i.e., with finite degrees) from our constructions that can saturate both the Wyner-Ziv and Gelfand-Pinsker bounds.

I. INTRODUCTION

Sparse graphical codes, particularly low-density parity check (LDPC) codes, are widely used and well understood in application to channel coding problems [10]. For other communication problems, especially those involving aspects of both channel and source coding, there remain various open questions associated with using low-density code constructions. Two important examples are source coding with side information (the Wyner-Ziv problem), and channel coding with side information (the Gelfand-Pinsker problem). This paper focuses on the design and analysis of low-density codes—more specifically, constructions based on a combination of LDPC and low-density generator matrix (LDGM) codes—for source and channel coding with side information. It builds on our previous work [7], in which we proved that low-density constructions and ML decoding can saturate the rate-distortion bound for a symmetric Bernoulli source.

Related work: It is well-known that random constructions of nested codes can saturate the Wyner-Ziv and Gelfand-Pinsker bounds [13], [15]. However, an unconstrained random construction leads to a high-density code, which is of limited practical use. One practically viable approach to lossy compression is trellis coded quantization (TCQ) [6]. A number of researchers have exploited TCQ as a quantizer for the Wyner-Ziv and related multiterminal source coding problems [2], [14] as well as for channel coding with side information []. A disadvantage of TCQ is that

saturating rate-distortion bounds requires that the trellis constraint length be taken infinity [11]; consequently, the computational complexity of decoding, even using message-passing algorithms, grows exponentially. It is therefore of considerable interest to develop low-density graphical constructions for such problems. Past work by a number of researchers [8], [12], [3], [9] has suggested that LDGM codes, which arise as the duals of LDPC codes, are well-suited to various types of quantization.

Our contributions: In this paper, we describe a sparse graphical construction for generating nested codes that are simultaneously good as both source and channel codes. We build on our previous work [7], in which we analyzed constructions, based on a combination of LDPC and LDGM codes, for the problem of standard lossy compression. Here we prove that there exist variants of these joint LDPC/LDGM constructions with finite degrees such that, when decoded/encoded using maximum likelihood, can saturate the Wyner-Ziv and Gelfand-Pinsker bounds. Although ML decoding is not practically viable, the low-density nature of our construction means that they have low degree, and with high probability (w.h.p.) high girth and expansion, all of which are important for the application of efficient message-passing.

The remainder of this paper is organized as follows. Section II provides background on source coding with side information (SCSI, or the Wyner-Ziv problem), and channel coding with side information (CCSI, or the Gelfand-Pinsker problem). Section III introduces our joint LDGM/LDPC construction, and provides a high-level overview of its use for the SCSI and CCSI problems. In Section IV, we prove that our construction produces codes that are simultaneously “good” for both source and channel coding. We conclude with a discussion in Section V.

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 $\text{card}\{\cdot\}$ to denote the cardinality of a set,

EM was supported by Mitsubishi Electric Research Labs and MJW was supported by an Alfred P. Sloan Foundation Fellowship, an Okawa Foundation Research Grant, and NSF Grant DMS-0528488.

$\|\cdot\|_p$ to denote the p -norm of a vector, $\text{Ber}(t)$ to denote a Bernoulli- t distribution, and $H_b(t)$ to denote the entropy of a $\text{Ber}(t)$ random variable.

II. BACKGROUND

A. Source and channel coding

We begin with definitions of “good” source and channel codes that are useful for future reference.

Definition 1. (a) A code family is a good D -distortion binary symmetric source code if for any $\epsilon > 0$, there exists a code with rate $R < 1 - H_b(D) + \epsilon$ that achieves distortion D .

(b) A code family is a good $\text{BSC}(p)$ -noise channel code if for any $\epsilon > 0$ there exists a code with rate $R > 1 - H_b(p) - \epsilon$ with error probability less than ϵ .

B. Wyner-Ziv problem

Suppose that we wish to compress a symmetric Bernoulli source $\mathbf{s} \sim \text{Ber}(\frac{1}{2})$ so as to be able to reconstruct it with Hamming distortion D . By classical rate distortion theory [4], the minimum achievable rate is given by $R(D) = 1 - H_b(D)$. In the Wyner-Ziv extension [13], there is an additional source of side information about \mathbf{s} —say in the form $\mathbf{y} = \mathbf{s} \oplus \mathbf{w}$ where $\mathbf{w} \sim \text{Ber}(\delta)$ is observation noise—that is available only at the decoder. In this setting, the minimum achievable rate takes the form $R_{\text{WZ}}(D, p) = \text{l.c.e.} \{H_b(D * p) - H_b(D), (p, 0)\}$, where l.c.e. denotes the lower convex envelope. Note that in the special case $p = \frac{1}{2}$, the side information is useless, so that the Wyner-Ziv rate reduces to classical rate-distortion.

C. Gelfand-Pinsker problem

Now consider the binary information embedding problem: the channel has the form $\mathbf{y} = \mathbf{u} \oplus \mathbf{s} \oplus \mathbf{z}$, where \mathbf{u} is the channel input, \mathbf{s} is a host signal (not under control of the encoder), and $\mathbf{z} \sim \text{Ber}(p)$ is channel noise. The encoder is free to choose the input vector $\mathbf{u} \in \{0, 1\}^n$, subject to the channel constraint $\|\mathbf{u}\|_1 \leq wn$, so as to maximize the rate of information transfer. We write $\mathbf{u} \equiv \mathbf{u}_m$ where \mathbf{m} is the underlying message to be transmitted. The decoder wishes to recover the embedded message from the corrupted observation \mathbf{y} . It can be shown [1] that the capacity in this set-up is given by $R_{\text{IE}}(w, p) \text{ u.c.e.} \{H_b(w) - H_b(p), (0, 0)\}$, where u.c.e. denotes the upper convex envelope.

III. GENERALIZED COMPOUND CONSTRUCTION

In this section, we describe a compound construction that produces codes that are simultaneously “good”, in the senses previously defined, as source and channel codes. We then describe how the nested codes generated by this compound construction apply to the SCSI and CCSI problems.

A. Code construction

Consider the compound code construction illustrated in Fig. 1, defined by a factor graph with three layers. The top layer consists of n bits, each attached to an associated parity check. These parity checks connect to m variable nodes in

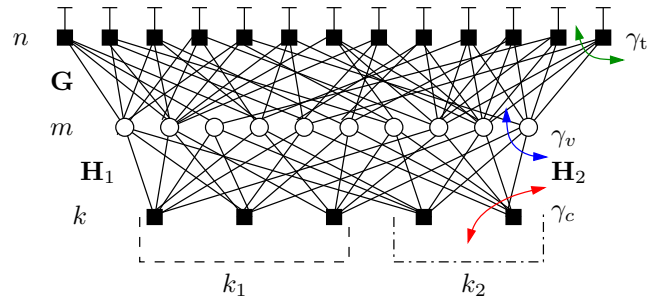


Fig. 1. Illustration of compound LDGM and LDPC code construction. The top section consists of an (n, m) LDGM code with generator matrix \mathbf{G} and constant check degrees $\gamma_t = 4$; its rate is $R(\mathbf{G}) = \frac{m}{n}$. The bottom section consists of (m, k_1) and (m, k_2) LDPC codes with degrees $(\gamma_v, \gamma_c) = (3, 6)$, described by parity check matrices \mathbf{H}_1 and \mathbf{H}_2 and rates $R(\mathbf{H}_1) = 1 - \frac{k_1}{m}$ and $R(\mathbf{H}_2) = 1 - \frac{k_2}{m}$ respectively. The overall rate of the compound construction is $R_{\text{com}} = R(\mathbf{G})R(\mathbf{H})$, where $R(\mathbf{H}) = R(\mathbf{H}_1) + R(\mathbf{H}_2)$.

the middle layer, and in turn these middle variable nodes are connected to $k = k_1 + k_2$ parity checks in the bottom layer.

Random LDGM ensemble: The top two layers define an (n, m) LDGM code. We construct it by connecting each of the n checks at the top randomly to γ_t variable nodes in the middle layer chosen uniformly at random. We use $\mathbf{G} \in \{0, 1\}^{m \times n}$ to denote the resulting generator matrix; by construction, each column of \mathbf{G} has exactly γ_t ones, whereas each row (corresponding to a variable node) has an (approximately) Poisson number of ones. An advantage of this regular-Poisson degree ensemble is that the resulting distribution of a random codeword is extremely easy to characterize:

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

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

Random LDPC code: The bottom two layers define a pair of LDPC codes, with parameters (m, k_1) and (m, k_2) ; we choose these codes from a standard standard (γ_v, γ_c) -regular LDPC ensemble originally studied by Gallager. 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 . Dividing the k check bits into two subsets, of size k_1 and k_2 with respective parity check matrices \mathbf{H}_1 and \mathbf{H}_2 , allows for the construction of nested codes, which will be needed for both the Wyner-Ziv and Gelfand-Pinsker problems.

B. Good source and channel codes

The key theoretical properties of this joint LDGM/LDPC construction are summarized in the following results:

Theorem 1 (Good source code). With appropriate finite degrees, there exist (n, m, k) constructions that are D -good source codes for all rates above $R(D) = 1 - H_b(D)$.

Theorem 2 (Good channel code). *With appropriate finite degrees, there exist (n, m, k) constructions that are good p -channel codes for all rates below capacity $C = 1 - H_b(p)$.*

Theorem 1 on source coding was proved in our previous work [7], whereas a proof of Theorem 2 is given in Section IV. We now describe how these two theorems allow us to establish that our low-density construction achieves the Wyner-Ziv and Gelfand-Pinsker bounds. At a high level, our approach is closely related to standard approaches to SCSI/CSCI coding; the key novelty is that appropriately nested codes can be constructed using low-density architectures.

C. Coding for Wyner-Ziv

We focus only on achieving rates of the form $H_b(D * p) - H_b(D)$, as any remaining rates on the Wyner-Ziv curve can be achieved by time-sharing with the point $(p, 0)$. To do this, we use the compound code in Fig. 1. Specifically, a source \mathbf{s} is encoded to $\mathbf{H}_2\mathbf{w}$ where \mathbf{w} is chosen to minimize the distortion $\|\mathbf{s} - \mathbf{w}'\mathbf{G}\|_1$ subject to the constraint that $\mathbf{H}_1\mathbf{w} = 0$. Theorems 1 and 2 show that maximum likelihood decoding of $\mathbf{H}_2\mathbf{w}$ using side information \mathbf{y} approaches the Wyner-Ziv bound in the sense that this construction yields a good D -distortion binary source code, and a nested subcode that is a good $D * p$ -noise channel code. Details follow.

Source coding component: The D -distortion source code component of the construction involves the n variable nodes representing the source bits, the m intermediate variable nodes, and the subset of k_1 lower layer check nodes. This subgraph, represented by the generator matrix \mathbf{G} and parity check matrix \mathbf{H}_1 (see Fig. 1), define a code (on the n variable nodes) with effective rate

$$R_1 := \frac{m(1 - \frac{k_1}{m})}{n} = \frac{m - k_1}{n}. \quad (2)$$

Choosing the middle and lower layer sizes m and k_1 such that $R_1 = 1 - H_b(D)$ guarantees (from Theorem 1) the existence of finite degrees such that this code is a good D -distortion source code.

Channel coding component: Now suppose that the source \mathbf{s} has been quantized, and is represented (up to distortion D) by the compressed sequence $\hat{\mathbf{x}} \in \{0, 1\}^m$. We transmit the associated sequence $\mathbf{H}_2\hat{\mathbf{x}} \in \{0, 1\}^{k_2}$ of parity bits associated with the code \mathbf{H}_2 ; doing so requires rate $R_{\text{trans}} = \frac{k_2}{n}$. The task of the decoder is as follows: given these k_2 parity bits as well as the k_1 zero-valued parity bits, the decoder seeks to recover the quantized sequence $\hat{\mathbf{x}}$ on the basis of the observed side-information \mathbf{y} . Note that from the decoder's perspective, the effective code rate is given by

$$R_2 = \frac{m - k_1 - k_2}{n} \quad (3)$$

Suppose that we choose k_2 such that $R_2 = 1 - H_b(D * p)$; then Theorem 2 guarantees that the decoder will (w.h.p.) be able to recover a codeword corrupted by $(D * p)$ -Bernoulli noise. Note that the side information can be written as $\mathbf{y} = \hat{\mathbf{s}} \oplus \mathbf{e} \oplus \mathbf{v}$, where $\mathbf{e} := \mathbf{s} \oplus \hat{\mathbf{s}}$ is the quantization noise,

and $\mathbf{v} \sim \text{Ber}(p)$ is the channel noise. If the quantization noise \mathbf{e} were i.i.d. $\text{Ber}(D)$, then the overall effective noise $\mathbf{e} \oplus \mathbf{v}$ would be i.i.d. $\text{Ber}(D * p)$. In reality, the quantization noise is not exactly i.i.d. $\text{Ber}(D)$, but it can be shown [15] that it can be treated as such for theoretical purposes.

In summary then, the overall transmission rate of this scheme for the Wyner-Ziv problem is given by

$$\left(\frac{m - k_1}{n}\right) - \left(\frac{m - k_1 - k_2}{n}\right) = H_b(D * p) - H_b(D). \quad (4)$$

Thus, by applying Theorems 1 and 2, we conclude that our low-density scheme saturates the Wyner-Ziv bound.

D. Coding for Gelfand-Pinsker

The construction for the Gelfand-Pinsker problem is similar, but with the order of the code nesting reversed. In particular, the Gelfand-Pinsker problem requires a good p -noise channel code, and a nested subcode that is a good w -distortion source code. As before, we focus only on achieving rates of the form $H_b(w) - H_b(p)$. To encode a message \mathbf{m} with side information \mathbf{y} , the channel input is $\mathbf{w}'\mathbf{G}$ where \mathbf{w} is chosen to minimize $\|\mathbf{y} - \mathbf{w}'\mathbf{G}\|_1$ subject to $\mathbf{H}_1\mathbf{w} = \mathbf{m}$. Details follow.

Source coding component: We begin by describing the nested subcode for the source coding component. The idea is to embed a message into the transmitted signal during the quantization process. The first set of k_1 lower parity bits remain fixed to zero throughout the scheme. On the other hand, we use the remaining k_2 lower parity bits to specify a particular message $\mathbf{m} \in \{0, 1\}^{k_2}$ that the decoder would like to recover. With the lower parity bits specified in this way, we use the resulting code to quantize a given source sequence \mathbf{s} to a compressed version $\hat{\mathbf{s}}$. If we choose n, m and k such that

$$R_1 = \frac{m - k_1 - k_2}{n} = 1 - H_b(w), \quad (5)$$

then Theorem 1 guarantees that the resulting code is a good w -distortion source code. Otherwise stated, we are guaranteed that w.h.p, the error $\mathbf{e} := \mathbf{s} \oplus \hat{\mathbf{s}}$ in our quantization has Hamming weight upper bounded by wn . Thus, transmitting the error \mathbf{e} ensures that the channel constraint is met.

Channel coding component: At the decoder, the k_1 lower parity bits remain set to zero; the remaining k_2 parity bits, which represent the message \mathbf{m} , are unknown to the coder. We now choose k_1 such that the effective code used by the decoder has rate

$$R_2 = \frac{m - k_1}{n} = 1 - H_b(p). \quad (6)$$

In addition, the decoder is given a noisy channel observation of the form $\mathbf{y} = \mathbf{e} \oplus \mathbf{s} \oplus \mathbf{v} = \hat{\mathbf{s}} \oplus \mathbf{v}$ and its task is to recover $\hat{\mathbf{s}}$. With the channel coding rate chosen as in equation (6) and channel noise $\mathbf{v} \sim \text{Ber}(p)$, Theorem 2 guarantees that the decoder will w.h.p. be able to recover $\hat{\mathbf{s}}$. By design of the quantization procedure, we have the equivalence $\mathbf{m} = \hat{\mathbf{s}}\mathbf{H}_2$ so that a simple syndrome-forming procedure allows the decoder to recover the hidden message. Thus, by applying Theorems 1 and 2, we conclude that our low-density scheme saturates the Gelfand-Pinsker bound under ML encoding/decoding.

IV. PROOF OF THEOREM 2

As described in the previous sections, Theorems 1 and 2 allow us to establish that the Wyner-Ziv and Gelfand-Pinsker bounds can be saturated under ML encoding/decoding. The source coding part—namely Theorem 1—was proved in our earlier work [7]. Here we provide a proof of Theorem 2, which ensures that these joint LDGM/LDPC constructions are good channel codes. We consider a joint construction, as illustrated in Fig. 1, consisting of a rate $R(\mathbf{G})$ LDGM top code, and a rate $R(\mathbf{H})$ lower LDPC code. Recall that the overall rate of this compound construction is given by $R_{\text{com}} = R(\mathbf{G})R(\mathbf{H})$. Note that an LDGM code on its own (i.e., without the lower LDPC code) is a special case of this construction with $R(\mathbf{H}) = 1$. However, a standard LDGM of this variety is *not* a good channel code, due to the large number of low-weight codewords. Essentially, the following proof shows that using a non-trivial LDPC lower code (with $R(\mathbf{H}) < 1$) can eliminate these troublesome low-weight codewords.

If the codeword \mathbf{c} is transmitted, then the receiver observes $\mathbf{y} = \mathbf{c} \oplus \mathbf{v}$ where \mathbf{v} is a $\text{Ber}(p)$ random vector. Our goal is to bound the probability that maximum likelihood (ML) decoding fails where the probability is taken over the randomness in both the channel noise and the code construction. To simplify the analysis, we focus on the following sub-optimal (non-ML) decoding procedure:

Definition 2 (Decoding Rule:). *With threshold $d(n) := (p + n^{-1/3})n$, decode to codeword $\mathbf{c}_i \iff \|\mathbf{c}_i \oplus \mathbf{y}\|_1 \leq d(n)$, and no other codeword is within $d(n)$ of \mathbf{y} .*

(The extra factor of $n^{-1/3}$ in the threshold $d(n)$ is of theoretical convenience.) Due to the linearity of the code construction, we may assume without loss of generality that the all zeros codeword 0^n was transmitted (i.e., $\mathbf{c} = 0^n$). In this case, the channel output is simply $\mathbf{y} = \mathbf{v}$ and so our decoding procedure will fail if and only if either (i) $\|\mathbf{v}\|_1 > d(n)$, or (ii) there exists some codeword “middle layer codeword” $\mathbf{z} \in \{0, 1\}^m$ satisfying the parity check equation¹ $\mathbf{H}\mathbf{z} = 0$ and corresponding to a codeword $\mathbf{c}_i = \mathbf{z}\mathbf{G}$ such that $\|\mathbf{z}\mathbf{G} \oplus \mathbf{v}\|_1 \leq d(n)$. Using the following two lemmas, we establish that this procedure has arbitrarily small probability of error, whence ML decoding (which is at least as good) also has arbitrarily small error probability.

Lemma 2. *The probability of decoding error vanishes asymptotically provided that*

$$R(\mathbf{G})\mathcal{A}(v) - D(p|\delta(v; \gamma_t) * p) < 0 \text{ for all } v \in (0, \frac{1}{2}] \quad (7)$$

where $\mathcal{A}(v) := \lim_{m \rightarrow +\infty} \mathcal{A}_m(v)$ is the asymptotic log-domain weight enumerator of the LDPC code, with $\mathcal{A}_m(v)$ being the average log-domain weight enumerator defined as

$$\mathcal{A}_m(v) := \frac{1}{m} \log \mathbb{E} \text{card} \{ \mathbf{z} \mid \|\mathbf{z}\|_1 = vm \}. \quad (8)$$

¹To be more precise, for the channel decoding step of the Wyner-Ziv problem, the middle layer codeword must satisfy $\mathbf{H}_1 \mathbf{z} = 0$ and $\mathbf{H}_2 \mathbf{z} = \mathbf{m}$ where \mathbf{m} is the output of the Wyner-Ziv encoder. For the channel decoding step of the Gelfand-Pinsker problem, the middle layer codeword must only satisfy $\mathbf{H}_1 \mathbf{z} = 0$, since \mathbf{m} is unknown until decoding is complete.

Proof. Let $N = 2^{nR_{\text{com}}}$ denote the total number of codewords in the joint LDGM/LDPC code. Then we can upper bound the probability of error using the union bound as follows:

$$p_{\text{err}} \leq \mathbb{P}[\|\mathbf{v}\|_1 > d(n)] + \sum_{i=2}^N \mathbb{P}[\|\mathbf{z}_i \mathbf{G} \oplus \mathbf{v}\|_1 \leq d(n)]. \quad (9)$$

By Bernstein’s inequality, the probability of the first error event vanishes for large n . Now focusing on the second sum, let us condition on the event that $\|\mathbf{z}\|_1 = \ell$. Then Lemma 1 guarantees that $\mathbf{z}\mathbf{G}$ has i.i.d. $\text{Ber}(\delta(\frac{\ell}{m}; \gamma_t))$ elements, so that the vector $\mathbf{z}\mathbf{G} \oplus \mathbf{v}$ has i.i.d. $\text{Ber}(\delta(\frac{\ell}{m}; \gamma_t) * p)$ elements. Applying Sanov’s theorem yields the upper bound

$$\mathbb{P}[\|\mathbf{z}\mathbf{G} \oplus \mathbf{v}\|_1 \geq d(n) \mid \|\mathbf{z}\|_1 = \ell] \leq 2^{-nD(p|\delta(\frac{\ell}{m}; \gamma_t) * p)}.$$

We can then upper bound the second error term (9) via

$$\begin{aligned} & 2^{nR_{\text{com}}} \sum_{\ell=0}^m \mathbb{P}[\|\mathbf{z}\|_1 = \ell] 2^{-nD(p|\delta(\frac{\ell}{m}; \gamma_t) * p)} \\ &= \sum_{\ell=0}^m 2^{\{nR_{\text{com}} + m[\mathcal{A}_m(\frac{\ell}{m}) - R(\mathbf{H})] - nD(p|\delta(\frac{\ell}{m}; \gamma_t) * p)\}} \\ &= \sum_{\ell=0}^m 2^{n\{R(\mathbf{G})\mathcal{A}_m(\frac{\ell}{m}) - D(p|\delta(\frac{\ell}{m}; \gamma_t) * p)\}} \\ &= \sum_{\ell=0}^m 2^{n\{R(\mathbf{G})[\mathcal{A}_m(\frac{\ell}{m}) - \mathcal{A}(\frac{\ell}{m}) + \mathcal{A}(\frac{\ell}{m})] - D(p|\delta(\frac{\ell}{m}; \gamma_t) * p)\}} \\ &\leq \sum_{\ell=0}^m 2^{n\{R(\mathbf{G})|\mathcal{A}_m(\frac{\ell}{m}) - \mathcal{A}(\frac{\ell}{m})|^+ + R(\mathbf{G})\mathcal{A}(\frac{\ell}{m}) - D(p|\delta(\frac{\ell}{m}; \gamma_t) * p)\}} \end{aligned}$$

where we have replaced $R_{\text{com}} = R(\mathbf{G})$ with $R(\mathbf{H})$ in the third line and used the notation $|x|^+$ to denote $\max(0, x)$. Finally, we notice that by the definition of the asymptotic weight enumerator, $\mathcal{A}(v)$, the $|\mathcal{A}_m(v) - \mathcal{A}(v)|^+$ term converges to zero uniformly² for $v \in [0, 1]$ leaving only the error exponent (7), which is negative by assumption. \square

Lemma 3. *For any $p \in (0, 1)$ and total rate $R_{\text{com}} := R(\mathbf{G})R(\mathbf{H}) < 1 - H_b(p)$, it is possible to choose the code parameters γ_t , γ_c and γ_v such that (7) is satisfied.*

Proof. For brevity, let $F(v) = R(\mathbf{G})\mathcal{A}(v) - D(p|\delta(v; \gamma_t) * p)$. It is well-known that a regular LDPC code with rate $R(\mathbf{H}) = \frac{\gamma_v}{\gamma_c} < 1$ has linear minimum distance; in particular, there exists a threshold $v^* = v^*(\gamma_v, \gamma_c)$ such that $\mathcal{A}(v) \leq 0$ for all $v \in [0, v^*]$. Hence, for $v \in (0, v^*]$, we have $F(v) < 0$.

Turning now to the interval $[v^*, \frac{1}{2}]$, consider the function

$$G(v) := R_{\text{com}}H_b(v) - D(p|\delta(v; \gamma_t)).$$

Since $\mathcal{A}(v) \leq R(\mathbf{H})H_b(v)$, we have $F(v) \leq G(v)$, so that it suffices to upper bound G . Observe that $G(\frac{1}{2}) = R_{\text{com}} - (1 - H_b(p)) < 0$. Therefore, it suffices to show that, by appropriate choice of γ_t , we can ensure that $G(v) \leq G(\frac{1}{2})$. Noting that G is infinitely differentiable and taking derivatives (details

²The definition of $\mathcal{A}(v)$ implies pointwise convergence of $|\mathcal{A}_m(v) - \mathcal{A}(v)|^+$ for $v \in [0, 1]$. But since the domain is compact, pointwise convergence implies uniform convergence.

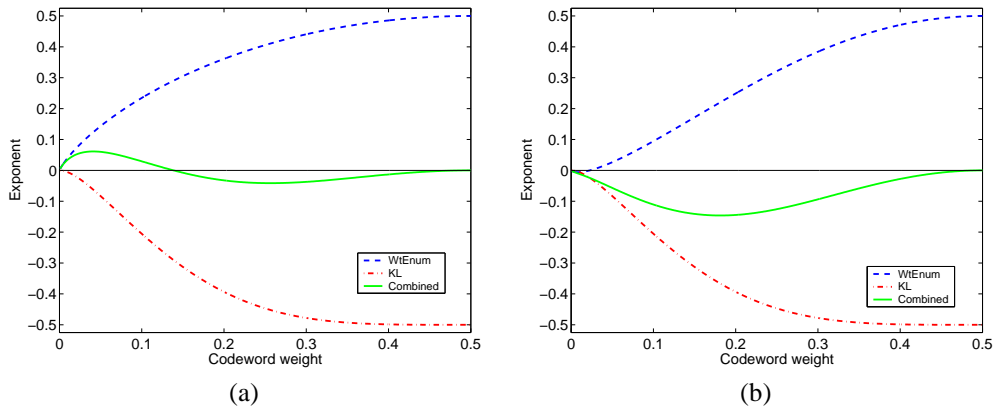


Fig. 2. Plots of different terms in error exponent (7). The combined curve must remain negative for all ω in order for the error probability to vanish asymptotically. (a) A LDGM $\gamma_t = 4$ construction without any LDPC lower code: here the weight enumerator \mathcal{A} is given by $H_b(\omega)$, and it dominates the Kullback-Leibler term for low ω . (b) The same $\gamma_t = 4$ LDGM combined with a $(\gamma_v, \gamma_c) = (3, 6)$ LDPC lower code: here the LDPC weight enumerator is dominated for all ω by the KL error exponent.

omitted), it can be shown that $G'(\frac{1}{2}) = 0$ and $G''(\frac{1}{2}) < 0$. Hence, a second order Taylor series expansion yields that $G(v) \leq G(\frac{1}{2})$ for all $v \in (\mu, \frac{1}{2}]$ for some $\mu < \frac{1}{2}$. It remains to bound G on the interval $[\nu^*, \mu]$. On this interval, we have $G(v) \leq R_{\text{com}} H_b(\mu) - D(p|\delta(\nu^*; \gamma_t))$. By examining (1), we see that choosing γ_t sufficiently large will ensure that on the interval $[\nu^*, \mu]$, the RHS is less than $R_{\text{com}} - (1 - H_b(p))$ as required. \square

Theorem 2 follows by combining the previous lemmas.

At first glance, Lemma 3 may seem unsatisfying, since it might require a very large top degree γ_t . Note, however, that this degree does not depend on the block length, hence our claim that good low density codes can be constructed with *finite* degree. Of course, for the claim of finite degree codes to be practically meaningful, the degree required for γ_t should be reasonably small. To investigate this issue, we plot the error exponent (7) for rate $R_{\text{com}} = 0.5$, LDGM top degree $\gamma_t = 4$, and different choices of lower code with $R(\mathbf{H})$ in Figure 2. Without any lower LDPC code, then $R(\mathbf{H}) = 1$ and the effective asymptotic weight enumerator is simply $H_b(\omega)$. Panel (a) shows the behavior in this case: note that the error exponent exceeds zero in a region around $v = 0$ where the weight enumerator dominates the negative KL term. In contrast, panel (b) shows the case of a $(\gamma_v, \gamma_c) = (3, 6)$ LDPC code, where we have used the results of Litsyn and Shevelev [5] in plotting the asymptotic weight enumerator. This code family has a linear minimum distance, so that the log-domain weight enumerator is negative in a region around $v = 0$. Thus, the error exponent (7) remains negative for all $v \in [0, 0.5]$. Thus, provided that a $(3, 6)$ lower LDPC code is used, a very reasonable top degree of $\gamma_t = 4$ is sufficient.

V. DISCUSSION

We have established that sparse graphical constructions that exploit both LDGM and LDPC codes can saturate fundamental bounds for problems of source coding with side information, and channel coding with side information. Although the present results are based on ML encoding/decoding, the sparsity and graphical structure of our constructions render them

suitable candidates for practical message-passing schemes, which remains to be investigated in future work.

REFERENCES

- [1] R. J. Barron, Brian Chen, and G. W. Wornell. The duality between information embedding and source coding with side information and some applications. *IEEE Trans. Info. Theory*, 49(5):1159–1180, May 2003.
- [2] J. Chou, S. S. Pradhan, and K. Ramchandran. Turbo and trellis-based constructions for source coding with side information. In *Data Compression Conference (DCC)*, 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] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.
- [5] 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.
- [6] M. W. Marcellin and T. R. Fischer. Trellis coded quantization of memoryless and Gauss-Markov sources. *IEEE Trans. Comm.*, 38(1):82–93, 1990.
- [7] E. Martinian and M. J. Wainwright. Low density codes achieve the rate-distortion bound. In *Data Compression Conference*, volume 1, page To appear, March 2006.
- [8] E. Martinian and J. S. Yedidia. Iterative quantization using codes on graphs. In *Allerton Conference on Communication, Control, and Computing*, Monticello, IL, October 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, February 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. Int. Symp. Info. Theory*, September 2005.
- [13] A. D. Wyner and J. Ziv. The rate-distortion function for source encoding with side information at the encoder. *IEEE Trans. Info. Theory*, IT-22:1–10, January 1976.
- [14] Y. Yang, V. Stankovic, Z. Xiong, and W. Zhao. On multiterminal source code design. In *Data Compression Conference*, 2005.
- [15] R. Zamir, S. Shamai (Shitz), and U. Erez. Nested linear/lattice codes for structured multiterminal binning. *IEEE Trans. Info. Theory*, 6(48):1250–1276, 2002.