

Nonlinear Sparse-Graph Codes for Lossy Compression

Ankit Gupta, *Student Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

Abstract—We propose a scheme for lossy compression of discrete memoryless sources: The compressor is the decoder of a nonlinear channel code, constructed from a sparse graph. We prove asymptotic optimality of the scheme for any separable (letter-by-letter) bounded distortion criterion. We also present a suboptimal compression algorithm, which exhibits near-optimal performance for moderate block lengths.

Index Terms—Discrete memoryless sources, lossy data compression, rate–distortion theory, source–channel coding duality, sparse-graph codes.

I. INTRODUCTION

EVEN for simple sources and distortion criteria, such as Bernoulli processes with bit-error-rate distortion, the construction of compression–decompression algorithms that perform near the rate–distortion function with reasonable complexity lags well behind the construction of capacity-achieving error-correcting codes. One reason for this is the fact that while linear codes achieve capacity for discrete channels with additive noise [1] (and the minimum lossless compression rate for arbitrary sources [2]), linear compressors cannot approach the rate–distortion function [3], (see also [4]). However, suppose that for a binary-symmetric source with bit-error-rate distortion, the codewords of a linear code for a binary symmetric channel are used as reconstruction codewords. Then, if the compressor is the maximum-likelihood channel decoder it is possible to find a sequence of linear codes that attain the rate–distortion function [5]. More generally, using nonbinary linear codes it is possible to approach the rate–distortion function of discrete memoryless sources arbitrarily closely as long as the distortion function is separable [6].

The advances in sparse-graph codes that perform close to capacity with low encoding–decoding complexity have spurred a number of recent works in the lossy data compression literature where a decoder for a low-density parity check code (LDPC) or low-density generator matrix code (LDGM) is

Manuscript received June 24, 2007; revised January 08, 2009. Current version published April 22, 2009. This work was supported in part by the National Science Foundation under Grant CCR-0312839. The material in this paper was presented in part at the IEEE Information Theory Workshop, Lake Tahoe, CA, September 2007.

The authors are with the Department of Electrical Engineering, Princeton University, Princeton NJ 08544 USA (e-mail:ankitg@princeton.edu; verdu@princeton.edu).

Communicated by M. Effros, Associate Editor for Source Coding.

Color versions of Figures 1–7 in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2009.2016040

used as the compressor. A sequence of LDPC codes is constructed in [7] that attains the rate–distortion function of the binary symmetric source with bit-error-rate distortion when the maximum-likelihood channel decoder is used as the lossy compressor. Unfortunately, the belief propagation decoder fails when used as a lossy encoder for this code. Furthermore, a polynomial-complexity encoder with near-optimal performance has not been found for this code. LDGM codes were proposed for this problem in [8]. In [9], generalized LDGM codes were constructed by substituting modulo 2 addition by other Boolean operations. Both [8] and [9] also propose low-complexity compressors based on the survey propagation algorithm [10] that show excellent empirical performance. However, the asymptotic optimality of LDGM codes for this problem is still open. Another approach using an LDPC–LDGM hybrid code with bounded check degrees is proposed in [11] and proven to be asymptotically optimal (with the computationally intensive maximum-likelihood decoder used as the compressor). Sparse-graph lossy compression systems for more general rate–distortion problems have been studied in [12] and [13]. In [12], asymptotically optimal LDPC codes for compressing the nonredundant (i.e., memoryless and equiprobable) q -ary source with a Hamming distortion criterion were proposed. In [13], an asymptotically optimal lossy compressor (based on LDPC codes) for compressing the Bernoulli source with Hamming distortion is proposed, but no computationally feasible compression algorithm is known for the codes in [12] and [13]. A sparse-graph-based lossy compressor for compressing discrete memoryless sources with an arbitrary separable distortion criterion has not been found in the literature yet. In fact, no code (linear or nonlinear; sparse-graph-based or not) is known that exhibits both asymptotic optimality and computationally feasible compression algorithms with near-optimal performance in the finite block length regime.

In the literature, various types of matrix sparsity are referred to as “low-density.” In the strong sense, this means that the nonzero entries per column (or row) in the (parity check, or generator) matrix remain bounded as the block length n grows [9], [8], [11], [14]. A weaker notion is that they are allowed to scale sublinearly with n [7], [12], [13]. It was shown in [15] that any LDGM code with bounded ones per column cannot achieve the optimal rate–distortion tradeoff, for the binary symmetric source with Hamming distortion.

In this paper, we propose a new construction of nonlinear codes based on LDGM matrices, which are asymptotically optimal for compressing discrete memoryless sources with separable distortion criterion. This construction is low-density in the weaker sense that the number of nonzero entries per row

in the generator matrix scale as $\log^2 n$ with the block length n . We also provide suboptimal compressors for these codes, which have excellent empirical performance, even at moderate block lengths. Our code design can be viewed as an intermediate on a continuum of block codes, with the linear codebook and the random nonlinear codebook as the two extremes.

The remainder of this paper is organized as follows. Section II presents the code design and proof of the asymptotic optimality of the construction for compressing the binary symmetric source with a Hamming distortion criterion. Section III extends the codes presented in Section II for compressing discrete memoryless sources with a separable (i.e., letter-by-letter) and bounded distortion criterion. Section IV proposes suboptimal compression algorithms whose performance is illustrated in Section V.

II. CODE CONSTRUCTION AND ANALYSIS FOR THE BINARY SOURCE

A. Code Construction

A binary (n, k) codebook has block length n and 2^k codewords. If there is no underlying structure to this set (for example, a random codebook) then exponential complexity is required for channel decoding (or lossy compression). A binary linear codebook, on the other hand, is a much restricted set of codewords: all the binary n -vectors \mathbf{c} that can be written as

$$\mathbf{c} = \mathbf{G}^T \mathbf{u} \quad (1)$$

for all possible choices of a binary k -vector \mathbf{u} , for a given binary $k \times n$ matrix \mathbf{G} . Note that if \mathbf{u} is not allowed to range over all 2^k choices then the ensuing codebook is, in general, nonlinear. In fact, any codebook (linear or nonlinear) can be described by (1) if \mathbf{u} is allowed to range over only the vectors with unit Hamming weight, and \mathbf{G} has as many rows as codewords, i.e., 2^k .

In this paper, we propose a class of nonlinear codebooks that has some convenient structure by letting \mathbf{u} range over the k -vectors with a given Hamming weight $\lceil k\omega \rceil$, where $0 < \omega < 1$. Further, we let

$$k = \lceil n \log_2 n \rceil \quad (2)$$

and

$$\omega = \frac{R}{\log_2 n \log_2 \log_2 n}. \quad (3)$$

We denote by $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M\}$ the binary k -strings of Hamming weight $\lceil k\omega \rceil$ in lexicographic order. The codebook is given by $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_M\}$ where $\mathbf{c}_i = \mathbf{G}^T \mathbf{u}_i$. The number of codewords in the codebook is equal to

$$M = \binom{k}{\lceil k\omega \rceil}. \quad (4)$$

Lemma 1: The asymptotic rate of the code converges to

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 M = R \quad (5)$$

with the choice of parameters in (2) and (3).

Proof: See Appendix I. \square

A convenient low-density choice of the $k \times n$ binary matrix \mathbf{G} is by independent and identically distributed generation of its coefficients where

$$\mathbb{P}[\mathbf{G}_{ij} = 1] = \frac{\log_2^2 n}{n}. \quad (6)$$

The lossy compressor is the minimum Hamming distance decoder and the decompressor is simply the encoder.

B. Code Analysis

We now turn to the analysis of the code introduced in Section II-A. We show that with high probability the lossy compressor described in Section II-A asymptotically attains the rate-distortion function of the memoryless binary symmetric source with Hamming distortion. More formally we show the following result.

Theorem 1: Let a codebook be constructed as specified in Section II-A with block length n and $R > 1 - h(d)$. As $n \rightarrow \infty$, the Hamming distortion obtained by representing an arbitrary n -length source realization \mathbf{s} with the nearest codeword in the codebook is less than d almost surely.

Proof: Pick a random codebook as outlined in Section II-A. Label the codewords as $\mathbf{c}_i, i \in \{1, 2, \dots, M\}$. Denote

$$L_i = \mathbf{1}\{w_H(\mathbf{c}_i \oplus \mathbf{s}) \leq nd\} \quad (7)$$

and

$$Z = \sum_{i=1}^M L_i. \quad (8)$$

The event $Z > 0$ is equivalent to the event that at least one codeword in the codebook is within Hamming distortion d from the given source. Thus, if we show that $\mathbb{P}[Z > 0]$ goes to 1 as $n \rightarrow \infty$ the theorem will be proved. However we will see later that using martingale arguments it is sufficient to show that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[Z > 0] = 0 \quad (9)$$

to claim that

$$\lim_{n \rightarrow \infty} \mathbb{P}[Z > 0] = 1. \quad (10)$$

Therefore, we first show (9) and then we prove (9) \implies (10) to complete the proof of Theorem 1. The proof is structured as a sequence of intermediate lemmas.

The Cauchy-Schwarz inequality gives a lower bound on $\mathbb{P}[Z > 0]$ in terms of the first and second moments of the nonnegative random variable Z

$$\mathbb{P}[Z > 0] \geq \frac{\mathbb{E}^2[Z]}{\mathbb{E}[Z^2]}. \quad (11)$$

To compute $\mathbb{E}[Z^2]$ we make use of the following result.

Lemma 2:

$$\frac{\mathbb{E}[Z^2]}{\mathbb{E}[Z]} = \sum_{j=1}^M \mathbb{P}[L_j = 1 | L_1 = 1]. \quad (12)$$

Proof of Lemma 2: Although this result is similar to Lemma 3 in [11], we cannot use the proof therein because it requires the code to be linear

$$\mathbb{E}[Z^2] = \mathbb{E} \left[\sum_{i=1}^M \sum_{j=1}^M L_i L_j \right] \quad (13)$$

$$= \sum_{i=1}^M \mathbb{E}[L_i] + \sum_{\substack{j=1 \\ j \neq i}}^M \sum_{i=1}^M \mathbb{E}[L_j L_i]. \quad (14)$$

Therefore

$$\begin{aligned} \mathbb{E}[Z^2] &= \sum_{i=1}^M \mathbb{P}[L_i = 1] \\ &+ \sum_{\substack{j=1 \\ j \neq i}}^M \sum_{i=1}^M \mathbb{P}[L_j = 1 | L_i = 1] \mathbb{P}[L_i = 1] \end{aligned} \quad (15)$$

$$= \mathbb{E}[Z] \left(1 + \sum_{\substack{j=1 \\ j \neq i}}^M \mathbb{P}[L_j = 1 | L_i = 1] \right) \quad (16)$$

$$= \mathbb{E}[Z] \sum_{j=1}^M \mathbb{P}[L_j = 1 | L_1 = 1], \quad (17)$$

where (17) follows from (16), because by symmetry $\sum_{\substack{j=1 \\ j \neq i}}^M \mathbb{P}[L_j = 1 | L_i = 1]$ does not depend on i . \square

We now show that asymptotically each codeword behaves like a sequence of fair coin flips, as formalized by the following result.

Lemma 3: For every $i \in \{1, 2, \dots, M\}$ and $n = 1, 2, \dots$ the bits $(\mathbf{c}_i[1], \mathbf{c}_i[2], \dots, \mathbf{c}_i[n])$ of the codeword \mathbf{c}_i are independent and identically distributed. If $i \neq j$ and A_i, A_j are disjoint subsets of $\{1, 2, \dots, n\}$, then $\{\mathbf{c}_i[l], l \in A_i\}$ and $\{\mathbf{c}_j[l], l \in A_j\}$ are independent. Furthermore

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbf{c}_i[m] = 1] = \frac{1}{2}, \quad (18)$$

for every $m \in \{1, 2, \dots, n\}$.

Proof of Lemma 3: Recall that

$$\mathbf{c}_i = \mathbf{G}^T \mathbf{u}_i. \quad (19)$$

By definition, $\mathbf{c}_i[m] = 1$ if and only if the $\lceil k\omega \rceil$ positions corresponding to the ones in \mathbf{u}_i select an odd number of ones in the m th row of \mathbf{G}^T . These events are independent with identical probabilities for different m , because the coefficients of \mathbf{G}^T are independent and identically distributed. Thus, if $A_i \cap A_j = \emptyset$ then $\{\mathbf{c}_i[l], l \in A_i\}$ and $\{\mathbf{c}_j[l], l \in A_j\}$ are independent and the bits $\{\mathbf{c}_i[1], \mathbf{c}_i[2], \dots, \mathbf{c}_i[n]\}$ are independent and identically distributed. Furthermore

$$\begin{aligned} \mathbb{P}[\mathbf{c}_i(m) = 1] &= \sum_{l=\text{odd}} \binom{\lceil k\omega \rceil}{l} \left(\frac{\log^2 n}{n} \right)^l \left(1 - \frac{\log^2 n}{n} \right)^{\lceil k\omega \rceil - l} \\ &= \frac{1}{2} \left(1 - \frac{\log^2 n}{n} + \frac{\log^2 n}{n} \right)^{\lceil k\omega \rceil} \end{aligned} \quad (20)$$

$$= \frac{1}{2} \left(1 - \frac{\log^2 n}{n} - \frac{\log^2 n}{n} \right)^{\lceil k\omega \rceil} \quad (21)$$

$$= \frac{1}{2} \left[1 - \left(1 - \frac{2 \log^2 n}{n} \right)^{\lceil k\omega \rceil} \right] \quad (22)$$

$$= \frac{1}{2} \left[1 - e^{-\frac{2R \log^2 n}{\log \log n}} \right] + o(1) \quad (23)$$

$$\rightarrow \frac{1}{2} \quad (24)$$

where (21) follows from (20) through application of the binomial expansion

$$\begin{aligned} (a+b)^n - (a-b)^n &= \sum_{l=1}^n \binom{n}{l} a^{n-l} b^l - \sum_{l=1}^n \binom{n}{l} a^{n-l} (-b)^l \end{aligned} \quad (25)$$

$$= 2 \sum_{l=\text{odd}} \binom{n}{l} a^{n-l} b^l. \quad (26)$$

\square

To compute probabilities of the form $\mathbb{P}[L_j = 1 | L_i = 1]$, we need the joint statistics of $\mathbf{c}_i, \mathbf{c}_j$. To that end, we have the following result.

Lemma 4: Let $j_n \in \{1, \dots, M\}$ be a sequence such that $w_H(\mathbf{u}_{j_n} \oplus \mathbf{u}_1) > \frac{\lceil k\omega \rceil}{\log_2 n}$, and let $\mathcal{B}_n \subset \{0, 1\}^n$. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[\mathbf{c}_{j_n} \in \mathcal{B}_n | \mathbf{c}_1 \in \mathcal{B}_n] = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[\mathbf{c}_1 \in \mathcal{B}_n] \quad (27)$$

whenever the limits exist.

Proof of Lemma 4:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_{j_n} \in \mathcal{B}_n | \mathbf{c}_1 \in \mathcal{B}_n] &= \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_{j_n} \in \mathcal{B}_n, \mathbf{c}_1 \in \mathcal{B}_n] \\ &- \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 \in \mathcal{B}_n] \end{aligned} \quad (28)$$

$$= (2 \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 |\mathcal{B}_n| - 2) - \left(\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 |\mathcal{B}_n| - 1 \right) \quad (29)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 \in \mathcal{B}_n] \quad (30)$$

where (29) is obtained from Lemmas 21 and 22 and (30) is obtained from Lemma 21. \square

Returning to the proof of Theorem 1, let

$$\mathcal{I} = \left\{ j : w_H(\mathbf{u}_j \oplus \mathbf{u}_1) \leq \frac{\lceil k\omega \rceil}{\log_2 n} \right\} \quad (31)$$

and

$$\mathcal{J} = \left\{ j : w_H(\mathbf{u}_j \oplus \mathbf{u}_1) > \frac{\lceil k\omega \rceil}{\log_2 n} \right\}. \quad (32)$$

In order to compute the quantity in the right-hand side of (12) we write

$$\begin{aligned} \sum_{j=1}^M \mathbb{P}[L_j = 1 | L_1 = 1] &= \sum_{j \in \mathcal{I}} \mathbb{P}[L_j = 1 | L_1 = 1] \\ &+ \sum_{j \in \mathcal{J}} \mathbb{P}[L_j = 1 | L_1 = 1]. \end{aligned} \quad (33)$$

The first term in (33) is less than or equal to $|\mathcal{I}|$, which grows subexponentially with n as the following result shows.

Lemma 5: Let \mathcal{I} be defined as in (31), then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |\mathcal{I}| = 0. \quad (34)$$

Proof of Lemma 5: Let

$$w_{\max} = \max_{j \in \mathcal{I}} w_H(\mathbf{u}_1 \oplus \mathbf{u}_j) \quad (35)$$

and let \mathcal{S} be the set of k -vectors defined as

$$\mathcal{S} = \{\mathbf{d} : w_H(\mathbf{d}) \leq w_{\max}\}. \quad (36)$$

Clearly

$$|\mathcal{I}| \leq |\mathcal{S}| \quad (37)$$

$$= \sum_{l=0}^{w_{\max}} \binom{k}{l} \quad (38)$$

$$\leq (w_{\max} + 1) \binom{k}{w_{\max}} \quad (39)$$

where we used the fact that

$$\binom{k}{l-1} \leq \binom{k}{l} \quad (40)$$

for $l \leq \lfloor \frac{k}{2} \rfloor$.

Next, using (39) we show that $|\mathcal{S}|$ grows subexponentially in n , thus proving (34).

$$\frac{1}{n} \log_2 |\mathcal{S}| \leq \frac{\log_2(w_{\max} + 1)}{n} + \frac{\log_2 \binom{k}{w_{\max}}}{n} \quad (41)$$

$$= \frac{kh(\frac{w_{\max}}{k})}{n} + o(1) \quad (42)$$

$$\leq \frac{n \log_2 n \log_2 \log_2 \log_2 n}{n \log_2^2 n \log_2 \log_2 n} + o(1) \quad (43)$$

$$= \frac{\log_2 \log_2 \log_2 n}{\log_2 n \log_2 \log_2 n} + o(1). \quad (44)$$

Substituting w_{\max} for $\lfloor k\omega \rfloor$ and noticing that by assumption $\frac{w_{\max}}{k} \leq \frac{\omega}{\log_2 n}$ we get (43). From (44) and the lower bound $|\mathcal{I}| \geq 1$ we get (34). \square

We give the asymptotic rate of decay of $\mathbb{P}[L_i = 1]$ in the following lemma.

Lemma 6: Let $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_M\}$ be a random codebook as chosen in Section II-A. For $i \in \{1, 2, \dots, M\}$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \frac{1}{\mathbb{P}[L_i = 1]} = 1 - h(d). \quad (45)$$

Proof of Lemma 6: Although a similar result is given in [14, Lemma 3], we give a self-contained proof due to the different code construction. From Lemma 3, $\mathbf{c}_i[m]$, $m \in \{1, 2, \dots, n\}$ are independent and identically distributed with

$$\mathbb{P}[\mathbf{c}_i[m] = 1] = p_n \quad (46)$$

for $n = 1, 2, \dots$, such that $p_n < 1/2$ and

$$\lim_{n \rightarrow \infty} p_n = \frac{1}{2}. \quad (47)$$

Let $\mathbf{0}$ and $\mathbf{1}$, be the all-zero and all-one n -vectors respectively. Then, using Lemma 14 (Appendix II), and Sanov's theorem (e.g., [16, Theorem 11.4.1]), we obtain

$$\frac{1}{n} \log_2 \frac{1}{\mathbb{P}[L_i = 1]} \leq \frac{1}{n} \log_2 \frac{1}{\mathbb{P}[w_H(\mathbf{1} \oplus \mathbf{c}_i) \leq nd]} \quad (48)$$

$$\leq \frac{2 \log_2 n}{n} + D(d \| 1 - p_n) \quad (49)$$

and

$$\frac{1}{n} \log_2 \frac{1}{\mathbb{P}[L_i = 1]} \geq \frac{1}{n} \log_2 \frac{1}{\mathbb{P}[w_H(\mathbf{0} \oplus \mathbf{c}_i) \leq nd]} \quad (50)$$

$$\geq D(d \| p_n). \quad (51)$$

Together with (47), we obtain the desired result from (49)–(51). \square

Using Lemma 2 in [11] we have the following result.

Lemma 7: For Z defined in (8)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{E}[Z] \geq R - 1 + h(d). \quad (52)$$

Combined with inequality (11) and Lemma 7, the following result gives a lower bound to $\frac{1}{n} \log_2 \mathbb{P}[Z > 0]$.

Lemma 8: If $R = 1 - h(d) + \epsilon$ for an arbitrary $\epsilon > 0$, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{j=1}^M \mathbb{P}[L_j = 1 | L_1 = 1] \leq \epsilon. \quad (53)$$

Proof of Lemma 8: For arbitrary $a_n, b_n > 0$

$$\lim_{n \rightarrow \infty} \frac{\log(a_n + b_n)}{n} = \max \left\{ \lim_{n \rightarrow \infty} \frac{\log a_n}{n}, \lim_{n \rightarrow \infty} \frac{\log b_n}{n} \right\}. \quad (54)$$

Let \mathcal{I} and \mathcal{J} be defined in (31) and (32), respectively, then

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{j=1}^M \mathbb{P}[L_j = 1 | L_1 = 1] \\ &= \max_{S \in \{\mathcal{I}, \mathcal{J}\}} \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{j \in S} \mathbb{P}[L_j = 1 | L_1 = 1] \end{aligned} \quad (55)$$

$$= \left[\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{j \in \mathcal{J}} \mathbb{P}[L_j = 1 | L_1 = 1] \right]^+ \quad (56)$$

where (56) follows from Lemma 5 and (54).

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{j \in \mathcal{J}} \mathbb{P}[L_j = 1 | L_1 = 1] \\ & \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 (M \max_{j \in \mathcal{J}} \mathbb{P}[L_j = 1 | L_1 = 1]) \end{aligned} \quad (57)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 M + \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \max_{j \in \mathcal{J}} \mathbb{P}[L_j = 1 | L_1 = 1] \quad (58)$$

$$= R + \lim_{n \rightarrow \infty} \frac{1}{n} \max_{j \in \mathcal{J}} \log_2 \mathbb{P}[L_j = 1 | L_1 = 1] \quad (59)$$

$$= R + \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{P}[L_1 = 1] \quad (60)$$

$$= R - (1 - h(d)) \quad (61)$$

$$= \epsilon \quad (62)$$

where (60) is obtained by using Lemma 4. Combining (56) and (62) we obtain (53). \square

Now we show that $\mathbb{P}[Z > 0]$ does not decay exponentially in n .

Lemma 9: Let $R > 1 - h(d)$. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[Z > 0] = 0. \quad (63)$$

Proof of Lemma 9: Using inequality (11) and Lemma 2 we have

$$\frac{1}{n} \log \mathbb{P}[Z > 0] \geq \frac{2}{n} \log \mathbb{E}[Z] - \frac{1}{n} \log \mathbb{E}[Z^2] \quad (64)$$

$$= \frac{1}{n} \log \mathbb{E}[Z] - \frac{1}{n} \log \sum_{j=1}^M \mathbb{P}[L_j = 1 | L_1 = 1]. \quad (65)$$

The desired result follows from the asymptotic behavior of both terms in the right-side of (65) found in Lemmas 7 and 8. \square

To finalize the proof of Theorem 1 we use an argument that is virtually identical to the proof of Theorem 2 in [11]; we spell out the details because our code construction is different from the one presented in [11].

To prove Theorem 1 we will also use the following auxiliary bound.

Lemma 10: [17] For a martingale B_1, B_2, \dots if

$$|B_i - B_{i-1}| < \tau \quad (66)$$

for all $i \in \{2, 3, \dots, n\}$, then

$$\mathbb{P}[|B_n - B_0| > n\epsilon] < 2e^{-\frac{n\epsilon^2}{2\tau^2}}. \quad (67)$$

Define the martingale

$$B_i = \mathbb{E} \left[\min_{j \in \{1, \dots, M\}} (w_H(\mathbf{c}_j \oplus \mathbf{s})) | \mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_i \right] \quad (68)$$

where $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n$ are the rows of the \mathbf{G}^T matrix. For this martingale, (66) is satisfied with $\tau = 1$ according to Lemma 15. Note that B_0 is the average (over all the codebooks) of the Hamming distance between the source realization \mathbf{s} and the closest codeword. Furthermore, there is no averaging with respect to the codebook in the definition of B_n , as it is the distance between the source realization and the closest codeword in the codebook defined by \mathbf{G} , and

$$Z > 0 \Leftrightarrow B_n \leq nd. \quad (69)$$

If

$$\limsup_{n \rightarrow \infty} \frac{1}{n} B_0 = d + \epsilon' \quad (70)$$

then there exists a convergent subsequence such that

$$\lim_{i \rightarrow \infty} \frac{1}{n_i} B_0 = d + \epsilon'. \quad (71)$$

From Lemma 10 along this subsequence, we have

$$\lim_{i \rightarrow \infty} \frac{1}{n_i} \log \mathbb{P} \left[\frac{1}{n_i} B_{n_i} \leq d \right] = \lim_{i \rightarrow \infty} \frac{1}{n_i} \log \mathbb{P}[Z > 0] \quad (72)$$

$$\leq -\frac{\epsilon'^2}{2} \quad (73)$$

which contradicts Lemma 9. Thus

$$\limsup_{n \rightarrow \infty} \frac{1}{n} B_0 \leq d \quad (74)$$

which, according to Lemma 10, implies that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{1}{n} B_n > d \right] = 0. \quad (75)$$

Therefore, (10) follows in view of (69). \square

III. CODE CONSTRUCTION AND PROOF OF OPTIMALITY FOR THE DISCRETE MEMORYLESS SOURCE

A. Code Construction for the Nonredundant Source

We begin by generalizing the construction in Section II-A to the discrete nonredundant (i.e., memoryless and equiprobable) source taking values over an alphabet \mathcal{A} . We label the symbols in the source alphabet as $\mathcal{A} = \{0, 1, \dots, q - 1\}$. The codebook $\{\mathbf{c}_i\}, i \in \{1, 2, \dots, M\}$ is defined through a $k \times n$ matrix \mathbf{G} (where addition is over the group $\{0, 1, \dots, q - 1\}$), as

$$\mathbf{c}_i = \mathbf{G}^T \mathbf{u}_i \quad (76)$$

where $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M\}$ are binary k -vectors of Hamming weight $\lceil k\omega \rceil$ in lexicographic order and k and ω are chosen from (2) and (3). The asymptotic rate of the code satisfies (see Lemma 1)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 M = R. \quad (77)$$

The matrix \mathbf{G} is obtained by random independent and identically distributed generation of its coefficients such that

$$\mathbb{P}[\mathbf{G}_{ij} = 0] = 1 - \frac{\log^2 n}{n} \quad (78)$$

and

$$\mathbb{P}[\mathbf{G}_{ij} = l] = \frac{\log^2 n}{n(q-1)} \quad (79)$$

for $l \neq 0$. For this code construction we can show a general version of Lemma 3 replacing $1/2$ by $1/q$ in (18), and of Lemma 4 where $\mathcal{B}_n \subset \mathcal{A}^n$. In addition, this code construction achieves

the ideal rate–distortion tradeoff asymptotically (see [18] for details).

Theorem 2: Construct a codebook as given above with a block length n and $R > R(d)$, where $R(d)$ is the rate–distortion function for the equiprobable q -ary source with Hamming distortion. Let \mathbf{s} be the source realization. If \mathbf{c}_s is the nearest codeword in Hamming distance to \mathbf{s} , then

$$\lim_{n \rightarrow \infty} \mathbb{P}[w_H(\mathbf{s} - \mathbf{c}_s) > nd] = 0 \quad (80)$$

for all $\mathbf{s} \in \mathcal{A}^n$.

B. Code Construction for Discrete Memoryless Sources

Next, we bootstrap the code design for the nonredundant q -ary source with Hamming distortion, to obtain asymptotically optimal codes for compressing general discrete memoryless sources with bounded and separable¹ distortion criteria. The idea for this code construction is similar to the one given in [19] for channel coding and [6] for lossy source coding.

Consider a discrete memoryless source taking values over an alphabet \mathcal{A} with distribution P_S . Let $\hat{\mathcal{A}}$ be the reproduction alphabet and $d(\cdot, \cdot) : \mathcal{A} \times \hat{\mathcal{A}} \mapsto \mathbb{R}^+$ be the per-letter distortion measure. Consider the variational problem corresponding to the rate–distortion function for a given distortion d

$$R(d) = \min_{P_{\hat{S}|S}} I(S; \hat{S}) \quad (81)$$

$$\mathbb{E}[d(S, \hat{S})] \leq d$$

For brevity, we fix d , and denote by $P_{\hat{S}}$ the marginal distribution resulting from the minimization in (81). We will assume that $P_{\hat{S}}$ is a rational distribution, i.e., we can write it as

$$P_{\hat{S}}(\hat{a}) = \frac{l(\hat{a})}{q} \quad (82)$$

where $\hat{a} \in \hat{\mathcal{A}}$ and $l(\hat{a}), q$ are integers. This is not a very restrictive condition as we can design codes to operate arbitrarily closely to any given point on the rate–distortion tradeoff curve. Thus, for every $d, \epsilon > 0 \exists d_r$ such that

$$|d - d_r| < \epsilon \quad (83)$$

$$|R(d) - R(d_r)| < \epsilon \quad (84)$$

and $P_{\hat{S}}$ corresponding to d_r is of the form given in (82).

Given a rate R and $P_{\hat{S}}$ of the form (82), we construct a q -ary codebook \mathcal{C} using LDGM matrices with block length n , asymptotic rate R , and a deterministic mapping $Q : \{0, 1, \dots, q-1\} \mapsto \hat{\mathcal{A}}$, such that Q maps the equiprobable probability distribution over $\{0, 1, \dots, q-1\}$ to $P_{\hat{S}}$ over $\hat{\mathcal{A}}$. The codebook $\mathbf{c}_i, i \in \{1, 2, \dots, M\}$ is obtained by applying Q to each symbol in the codebook \mathcal{C} . The compressor selects the codeword closest to the source realization (according to the distortion criterion defined by $d(\cdot, \cdot)$). Again, since the number of codewords has not changed after applying the deterministic transformation Q , from Lemma 1 the asymptotic rate remains R .

¹Separable distortion means that $d^n(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^n d(\mathbf{x}[i], \mathbf{y}[i])$.

C. Code Analysis

The code construction in Section III-B achieves the ideal rate distortion performance asymptotically, as stated in the following result.

Theorem 3: Consider a memoryless source distributed according to P_S and per-letter distortion $d(\cdot, \cdot) \leq d_{\max}$, with rate–distortion function $R(\cdot)$. Let $\hat{S}_1, \dots, \hat{S}_n$ be the output of the compressor–decompressor in Section III-B designed for distortion d and asymptotic rate $R > R(d)$, when the source realization is S_1, \dots, S_n . Then

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\sum_{i=1}^n d(S_i, \hat{S}_i) > nd \right] = 0. \quad (85)$$

Proof: We begin by showing the following result for the distribution of \mathbf{c}_i .

Lemma 11: For every $i \in \{1, 2, \dots, M\}$ and n , $\{\mathbf{c}_i[1], \mathbf{c}_i[2], \dots, \mathbf{c}_i[n]\}$ are independent and identically distributed. If $i \neq j$ and A_i, A_j are disjoint subsets of $\{1, 2, \dots, n\}$, then $\{\mathbf{c}_i[l], l \in A_i\}$ and $\{\mathbf{c}_j[l], l \in A_j\}$ are independent. Furthermore

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbf{c}_i[m] = \hat{a}] = P_{\hat{S}}(\hat{a}) \quad (86)$$

for all $\hat{a} \in \hat{\mathcal{A}}$.

Proof of Lemma 11: Let $\mathbf{c}'_i, i \in \{1, 2, \dots, M\}$ be the codewords in Section III-B before the mapping Q is applied. From Lemma 3 (generalized to the q -ary case), if $i \neq j$ and A_i, A_j are disjoint subsets of $\{1, 2, \dots, n\}$, then $\{\mathbf{c}'_i[l], l \in A_i\}$ and $\{\mathbf{c}'_j[l], l \in A_j\}$ are independent. Therefore, $\{\mathbf{c}_i[l], l \in A_i\}$ and $\{\mathbf{c}_j[l], l \in A_j\}$ are independent since Q is a deterministic mapping. Further, from Lemma 3 (generalized to the q -ary case), for $l \in \{0, \dots, q-1\}$

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbf{c}'_i[m] = l] = \frac{1}{q}. \quad (87)$$

Therefore

$$\lim_{n \rightarrow \infty} \mathbb{P}[\mathbf{c}_i[m] = \hat{a}] = \lim_{n \rightarrow \infty} \sum_{t \in Q^{-1}(\hat{a})} \mathbb{P}[\mathbf{c}'_i[m] = t] \quad (88)$$

$$= \frac{|Q^{-1}(\hat{a})|}{q} \quad (89)$$

$$= P_{\hat{S}}(\hat{a}) \quad (90)$$

where $Q^{-1}(\hat{a})$ is the subset of $\{0, 1, \dots, q-1\}$ which is mapped to \hat{a} . We get (90) because Q maps the equiprobable probability distribution to $P_{\hat{S}}$. \square

Returning to the proof of Theorem 3, define

$$d^n(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^n d(\mathbf{x}[i], \mathbf{y}[i]) \quad (91)$$

and let $\mathcal{Z}^n \subset \mathcal{A}^n$ be such that

$$\lim_{n \rightarrow \infty} \mathbb{P}[d^n(S^n, \hat{S}^n) > d | S^n \in \mathcal{Z}^n] = 0 \quad (92)$$

then

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}[d^n(S^n, \hat{S}^n) > d] \\ &= \lim_{n \rightarrow \infty} \mathbb{P}[d^n(S^n, \hat{S}^n) > d | S^n \in \mathcal{Z}^n] \mathbb{P}[S^n \in \mathcal{Z}^n] \\ & \quad + \lim_{n \rightarrow \infty} \mathbb{P}[d^n(S^n, \hat{S}^n) > d | S^n \notin \mathcal{Z}^n] \mathbb{P}[S^n \notin \mathcal{Z}^n] \quad (93) \\ & \leq \lim_{n \rightarrow \infty} \mathbb{P}[S^n \notin \mathcal{Z}^n]. \quad (94) \end{aligned}$$

Therefore, to prove Theorem 3 all we need to show is that there exists a set $\mathcal{Z}^n \subset \mathcal{A}^n$ over which (92) holds and

$$\lim_{n \rightarrow \infty} \mathbb{P}[S^n \in \mathcal{Z}^n] = 1. \quad (95)$$

For $\mathbf{x} \in \mathcal{A}^n$, let

$$L_i(\mathbf{x}) = \mathbf{1}\{d^n(\mathbf{x}, \mathbf{c}_i) \leq d\} \quad (96)$$

and define

$$\mathcal{Z}^n = \left\{ \mathbf{x} \in \mathcal{A}^n : \left| \frac{1}{n} \log \frac{1}{\mathbb{P}[L_1(\mathbf{x}) = 1]} - R(d) \right| \leq \epsilon_n \right\} \quad (97)$$

for a vanishing sequence ϵ_n , which will be specified later. We will now show that \mathcal{Z}^n satisfies (92) and (95) in the following two lemmas, completing the proof of Theorem 3.

Lemma 12: For \mathcal{Z}^n defined in (97) and $R > R(d)$, (92) holds.

Proof of Lemma 12: Fix $\mathbf{x} \in \mathcal{Z}^n$, let $Z(\mathbf{x}) = \sum_{i=1}^M L_i(\mathbf{x})$. With $R = R(d) + \epsilon$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{E}[Z(\mathbf{x})] = \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{i=1}^M \mathbb{E}[L_i(\mathbf{x})] \quad (98)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{i=1}^M \mathbb{P}[L_i(\mathbf{x}) = 1] \quad (99)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 (M \mathbb{P}[L_1(\mathbf{x}) = 1]) \quad (100)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 M - \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \frac{1}{\mathbb{P}[L_1(\mathbf{x}) = 1]} \quad (101)$$

$$= \epsilon, \quad (102)$$

where (102) follows from Lemma 1 and (97). Define \mathcal{I}, \mathcal{J} as in (31) and (32). Using Lemma 5 and (54)–(56) we get

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{j=1}^M \mathbb{P}[L_j(\mathbf{x}) = 1 | L_1(\mathbf{x}) = 1] \\ &= \max_{S \in \{\mathcal{I}, \mathcal{J}\}} \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{j \in S} \mathbb{P}[L_j(\mathbf{x}) = 1 | L_1(\mathbf{x}) = 1] \quad (103) \end{aligned}$$

$$= \left[\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{j \in \mathcal{J}} \mathbb{P}[L_j(\mathbf{x}) = 1 | L_1(\mathbf{x}) = 1] \right]^+ \quad (104)$$

Let j_n be a sequence such that $j_n \in \mathcal{J}$, and let $\mathbf{c}'_{j_n} = \mathbf{G}^T \mathbf{u}_{j_n}$ and $\mathbf{c}'_1 = \mathbf{G}^T \mathbf{u}_1$. Using Lemma 4 generalized to the q -ary case: for a sequence of sets $\mathcal{B}_n \subset \{0, 1, \dots, q-1\}^n$, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[\mathbf{c}'_{j_n} \in \mathcal{B}_n | \mathbf{c}'_1 \in \mathcal{B}_n] = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[\mathbf{c}'_1 \in \mathcal{B}_n], \quad (105)$$

whenever the limits exist. Therefore, for a sequence of sets $\mathcal{D}_n \subset \hat{\mathcal{A}}^n$ and $j_n \in \mathcal{J}$, we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[\mathbf{c}_{j_n} \in \mathcal{D}_n | \mathbf{c}_1 \in \mathcal{D}_n] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[\mathbf{c}'_{j_n} \in Q^{-1}(\mathcal{D}_n) | \mathbf{c}'_1 \in Q^{-1}(\mathcal{D}_n)] \quad (106) \end{aligned}$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[\mathbf{c}'_1 \in Q^{-1}(\mathcal{D}_n)] \quad (107)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[\mathbf{c}_1 \in \mathcal{D}_n] \quad (108)$$

whenever the limits exist (where $Q^{-1}(S)$ denotes the pre-image of the set S). Using (108) and (57)–(60) we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{j \in \mathcal{J}} \mathbb{P}[L_j(\mathbf{x}) = 1 | L_1(\mathbf{x}) = 1] \\ & \leq R + \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{P}[L_1(\mathbf{x}) = 1] \quad (109) \end{aligned}$$

$$= R - R(d) \quad (110)$$

$$= \epsilon. \quad (111)$$

Therefore

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{j=1}^M \mathbb{P}[L_j(\mathbf{x}) = 1 | L_1(\mathbf{x}) = 1] \leq \epsilon. \quad (112)$$

Using (112), (102), and (11) for $\mathbf{x} \in \mathcal{Z}^n$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{P}[Z(\mathbf{x}) > 0] = 0. \quad (113)$$

Now we proceed as in the proof of Theorem 1 (see also [11]). Define a martingale B_i $i \in \{1, 2, \dots, n\}$ as

$$B_i = \mathbb{E} \left[\min_{j \in \{1, \dots, M\}} (d^n(\mathbf{x}, \mathbf{c}_j)) | \mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_i \right] \quad (114)$$

where $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n$ are the columns of the \mathbf{G} matrix. For this martingale, (66) is satisfied with $\tau = d_{\max}$ according to Lemma 17. The remainder of the proof proceeds as in the proof of Theorem 1. \square

Now, to complete the proof of Theorem 3 all we need to show is the following result.

Lemma 13: There exists a sequence $\epsilon_n \rightarrow 0$ such that for \mathcal{Z}^n defined in (97), (95) holds.

Proof of Lemma 13: From Lemma 11, for fixed n each codeword \mathbf{c}_i is a sequence of independent and identically distributed random variables with distribution $P_{\hat{S}(n)}$, such that

$$\lim_{n \rightarrow \infty} P_{\hat{S}(n)}(\hat{a}) = P_{\hat{S}}(\hat{a}). \quad (115)$$

Fix a source realization \mathbf{x} and define the random variables $D_l = d(\mathbf{x}[l], \mathbf{c}_1[l]) = d(\mathbf{x}[l], \hat{S}(n))$. Let

$$\Lambda_n(\mathbf{x}, \lambda) = \log \mathbb{E}[e^{\lambda \sum_{l=1}^n D_l}] \quad (116)$$

$$= \sum_{l=1}^n \log \mathbb{E}[e^{\lambda d(\mathbf{x}[l], \hat{S}(n))}] \quad (117)$$

and let

$$\Lambda(\lambda) \triangleq \mathbb{E}[\log \mathbb{E}[e^{\lambda d(S, \hat{S})} | S]]. \quad (118)$$

Define

$$\mathcal{E}^n = \left\{ \mathbf{x} \in \mathcal{A}^n : \left| \frac{1}{n} \sum_{l=1}^n \log \mathbb{E}[e^{\lambda d(\mathbf{x}[l], \hat{S})}] - \Lambda(\lambda) \right| \leq \frac{1}{\sqrt[n]{n}} \right\}. \quad (119)$$

Using the Chebyshev inequality

$$\lim_{n \rightarrow \infty} \mathbb{P}[S^n \in \mathcal{E}^n] = 1. \quad (120)$$

From Lemma 16 for $\mathbf{x} \in \mathcal{E}^n$ and any fixed $\epsilon' > 0$

$$\Lambda(\lambda) - \epsilon' < \lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_n(\mathbf{x}, \lambda) < \Lambda(\lambda) + \epsilon' \quad (121)$$

which implies, since ϵ' is arbitrary

$$\lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_n(\mathbf{x}, \lambda) = \Lambda(\lambda). \quad (122)$$

Using the Gartner–Ellis theorem [20, Theorem 2.3.6], for $\mathbf{x} \in \mathcal{E}^n$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\mathbb{P}[\sum_{l=1}^n D_l \leq dn]} = \lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\mathbb{P}[L_1(\mathbf{x}) = 1]} \quad (123)$$

$$= \Lambda^*(d) \quad (124)$$

where $\Lambda^*(d)$ is the Fenchel–Legendre transform of $\Lambda(\lambda)$ evaluated at d , i.e.,

$$\Lambda^*(d) = \sup_{\lambda \leq 0} [\lambda d - \Lambda(\lambda)]. \quad (125)$$

From [21, Theorem 2], $\Lambda^*(d) = R(d)$. Therefore, for a suitably chosen ϵ_n , $\mathcal{E}^n \subset \mathcal{Z}^n$ and (120) implies that (95) is satisfied. \square

Finally, Theorem 3 follows from Lemmas 12, 13 as well as (92) and (95). \square

IV. SUBOPTIMAL ALGORITHMS FOR COMPRESSION

In this section, we describe a suboptimal algorithm (and its variants) to encode a source using the codebooks described in Sections II and III. This algorithm attempts to locate a codeword in the codebook such that the distortion between the source and codeword is minimum. We note that an optimal algorithm to select such a codeword is NP-complete [22], therefore any polynomial complexity algorithm (like the one presented here) is necessarily suboptimal. However, it should be noted that NP-completeness implies hardness in the sense of worst case input; it does not rule out the existence of a polynomial time algorithm that is able to locate the minimum-distance codeword for most source sequences. Empirical results in Section V demonstrate that our algorithm attains near-optimum performance. The compressor/decompressor work as follows.

Recall that the codebook is specified by the $k \times n$ matrix \mathbf{G} and $0 < \omega < 1$. It consists of all the codewords \mathbf{c} of length n that can be written as

$$\mathbf{c} = Q(\mathbf{G}^T \mathbf{u}), \quad (126)$$

where Q is the identity mapping for the binary and nonredundant case and is a deterministic many to one mapping (see Section III-B) otherwise, where with slight abuse of notation

each component is mapped with the function Q defined in Section III-B. The k -vector \mathbf{u} satisfies

$$w_H(\mathbf{u}) = \lceil k\omega \rceil. \quad (127)$$

The algorithm attempts to find a good approximation to the source string \mathbf{s} of length n among the codewords in an iterative manner. At each step in the iteration, we select a string of length k , \mathbf{u}_{i+1} by flipping one and only one bit in \mathbf{u}_i . The algorithm starts with $\mathbf{u}_0 = [0, 0, \dots, 0]$ and at the completion of the algorithm we have $w_H(\mathbf{u}_t) = \lceil k\omega \rceil$. Let $\Delta_i = \mathbf{G}^T \mathbf{u}_i$. The choice of the bit to flip in \mathbf{u}_i is such that $d^n(\mathbf{s}, Q(\Delta_{i+1}))$ is minimized. To that end, we perform an exhaustive search for all columns of \mathbf{G}^T enumerated as $\mathbf{g}_j, j \in \{1, 2, \dots, k\}$, computing $\Delta_i + \mathbf{g}_j$, and selecting the index j that leads to the lowest distortion metric. This procedure is repeated till $d^n(\mathbf{s}, Q(\Delta_i)) > d^n(\mathbf{s}, Q(\Delta_{i+1}))$. If $d^n(\mathbf{s}, Q(\Delta_i)) \leq d^n(\mathbf{s}, Q(\Delta_{i+1}))$, then if $w_H(\mathbf{u}_i) < \lceil k\omega \rceil$, the algorithm is now constrained to flip only bits which are zero in \mathbf{u}_i that lead to minimum $d^n(\mathbf{s}, Q(\Delta_{i+1}))$, and *vice versa* for the case when $w_H(\mathbf{u}_i) > \lceil k\omega \rceil$. We halt when $w_H(\mathbf{u}_t) = \lceil k\omega \rceil$. It is immaterial how ties are broken by the compressor. At the final configuration of \mathbf{u}_t , the encoder then stores the value of \mathbf{u}_t in the form of an index, using an enumerative encoding scheme [23]. The decoder then uses this index to recover \mathbf{u}_t , and outputs $Q(\mathbf{G}^T \mathbf{u}_t)$. A pseudocode description of the encoder is given in Algorithm 1 at the top of the following page.

Some other variations in this algorithm are also possible. For example, we can fix a recursion depth d . We run multiple copies of this algorithm whenever we have ties for the element with the maximum gain by flipping each maximum gain position in different copies. After d bits have been flipped, the algorithm proceeds as described above in each of the multiple copies. Finally, we choose the winner out of all these multiple copies. Another possible variation is to flip pairs of bits simultaneously, selecting the pair which leads to the best approximation to the source.

For the core algorithm, the complexity analysis may be performed as follows: To compute Δ_{i+1} at each iteration requires $O(k \log^2 n)$ computations on average because we add a total of k columns (\mathbf{g}_i) to Δ_{i+1} and each column contains $\log_2^2 n$ entries on average; therefore, the average computational cost is $O(k \log^2 n)$. The number of iterations is bounded by $n(d_{\max}) + \lceil k\omega \rceil = O(n)$. Thus, the average complexity of the algorithm is $O(nk \log^2(n)) = O(n^2 \log^3(n))$ compared to various message-passing-based approaches such as survey propagation [9] and its variants [8], which incur a complexity of $O(n^2)$.

V. EXPERIMENTS

In this section, we show empirical results obtained with the codes given in Sections II and III and the encoding/decoding algorithms in Section IV for a variety of rate–distortion problems. For each rate we fix a randomly generated codebook and average the distortion obtained for compressing a random source (for 1000 iterations).

LDGM codes and message-passing algorithms perform very close to the rate–distortion function for compressing the binary symmetric source with Hamming distortion for block lengths of the order of thousands as demonstrated by the empirical results in [8] and [9]. However, these algorithms perform far from

Algorithm 1: Encoding Algorithm

```

begin
     $i = 0$   $\mathbf{u}_i = [0, 0, \dots, 0]$   $\Delta_i = \mathbf{G}^T \mathbf{u}_i$ ;
    repeat
         $mindist = \infty$ ;
        for  $j=1:k$  do
             $dist = d^n(\mathbf{s}, Q(\Delta_i + \mathbf{g}_j))$ ;
            if  $dist < mindist$  then
                 $flippos = j$ ;
                 $mindist = dist$ ;
            end
        end
         $i = i + 1$ ;
         $\mathbf{u}_i = \mathbf{u}_{i-1}$ ;
         $\mathbf{u}_i[flippos] = \mathbf{u}_i[flippos] \oplus 1$ ;
         $\Delta_i = \mathbf{G}^T \mathbf{u}_i$ ;
    until  $d^n(\mathbf{s}, Q(\Delta_i)) < d^n(\mathbf{s}, Q(\Delta_{i-1}))$ ;
    if  $w_H(\mathbf{u}_i) > \lceil k\omega \rceil$  then
        repeat
             $list = find(\mathbf{u}_i = 1)$ ;
             $mindist = \infty$ ;
            for  $k = 1 : length(list)$  do
                 $j = list(k)$ ;
                 $dist = d^n(\mathbf{s}, Q(\Delta_i + \mathbf{g}_j))$ ;
                if  $dist < mindist$  then
                     $flippos = j$ ;
                     $mindist = dist$ ;
                end
            end
             $\mathbf{u}_i[flippos] = \mathbf{u}_i[flippos] \oplus 1$ ;
             $\Delta_i = \mathbf{G}^T \mathbf{u}_i$ ;
        until  $w_H(\mathbf{u}_i) = \lceil k\omega \rceil$ ;
    end
    else if  $w_H(\mathbf{u}_i) < \lceil k\omega \rceil$  then
        repeat
             $list = find(\mathbf{u}_i = 0)$ ;
             $mindist = \infty$ ;
            for  $k = 1 : length(list)$  do
                 $j = list(k)$ ;
                 $dist = d^n(\mathbf{s}, Q(\Delta_i + \mathbf{g}_j))$ ;
                if  $dist < mindist$  then
                     $flippos = j$ ;
                     $mindist = dist$ ;
                end
            end
             $\mathbf{u}_i[flippos] = \mathbf{u}_i[flippos] \oplus 1$ ;
             $\Delta_i = \mathbf{G}^T \mathbf{u}_i$ ;
        until  $w_H(\mathbf{u}_i) = \lceil k\omega \rceil$ ;
    end
    return enumerative_encoding( $\mathbf{u}_i$ );
end
    
```

optimal for short block lengths due to the effect of cycles in the graph. On the other hand, our scheme performs well even for short block lengths (such as $n = 400$) as shown in Fig. 1. For block length $n > 1000$, both schemes are very close to the rate–distortion function without any discernible difference in performance as seen in Figs. 2 and 3.

In Figs. 4 and 5, we plot the rate–distortion tradeoff for an equiprobable 4-ary source for block lengths $n = 100$ and $n = 400$, respectively, with a Hamming distortion criterion. These figures show that the performance of our codes and encoding algorithm is very close to the optimal for short block lengths (for the general q -ary source with Hamming distortion). In Fig. 6 we show results obtained with codes from Section III for compressing the 0.4-Bernoulli source with a Hamming distortion criterion and block length $n = 1000$. We now provide an illustration of the code construction in Section III-B for this problem. In this case, the reproduction distribution corresponding to the rate–distortion point $(R(d), d)$ is given as $\{P_S[1], P_S[0]\} = \{(p-d)/(1-2d), (1-p-d)/(1-2d)\}$. Let $d = 0.1$, the corresponding reproduction distribution is $\{P_S[1], P_S[0]\} = \{3/8, 5/8\}$, thus, a reasonable choice of q from (82) is $q = 8$. The mapping $Q(\cdot)$ should map the equiprobable probability distribution over $\{0, 1, \dots, 7\}$ to the distribution $\{5/8, 3/8\}$ over $\{0, 1\}$, a possible choice is $Q(a) = 1, a \in \{0, 1, 2\}$ and 0 otherwise. Thus, to obtain the codebook for compressing the 0.4-Bernoulli source with Hamming distortion 0.1 we construct a q -ary codebook over the alphabet $\{0, 1, \dots, 7\}$ with rate $h(0.4) - h(0.1)$, and apply the mapping $Q(\cdot)$ to each of its codewords.

In Fig. 7, we show results obtained with the codes in Section III, when used for compressing the binary symmetric source (with block length $n = 1000$) where the distortion criterion d satisfies

$$d(0, 1) = 2d(1, 0) = 2. \quad (128)$$

These experiments demonstrate the near-optimal performance of the proposed codes for simple memoryless sources and separable distortion criterion, even for short block lengths. Furthermore, the low complexity of the proposed suboptimal compression algorithm makes the new codes particularly appealing.

APPENDIX I
PROOF OF LEMMA 1

The rate of the code as a function of the block length n is given as

$$R_n = \frac{1}{n} \log_2 \binom{k}{\lceil k\omega \rceil}. \quad (129)$$

From [16, eq. (11.40)]

$$\frac{1}{n+1} 2^{nh(\frac{k}{n})} \leq \binom{n}{k} \leq 2^{nh(\frac{k}{n})}. \quad (130)$$

Therefore

$$R_n = \frac{k}{n} h \left(\frac{\lceil k\omega \rceil}{k} \right) + o(1) \quad (131)$$

$$= \frac{1}{n} (k \log_2 k - k\omega \log_2(k\omega) - k(1-\omega) \log_2(k(1-\omega))) + o(1) \quad (132)$$

$$= \frac{k h(\omega)}{n} + o(1) \quad (133)$$

$$= \frac{\log_2(n) R(\log_2 \log_2(n) + \log_2 \log_2 \log_2(n) - \log_2 R)}{\log_2(n) \log_2 \log_2(n)} + o(1) \quad (134)$$

$$= R + o(1) \quad (135)$$

as we wanted to show.

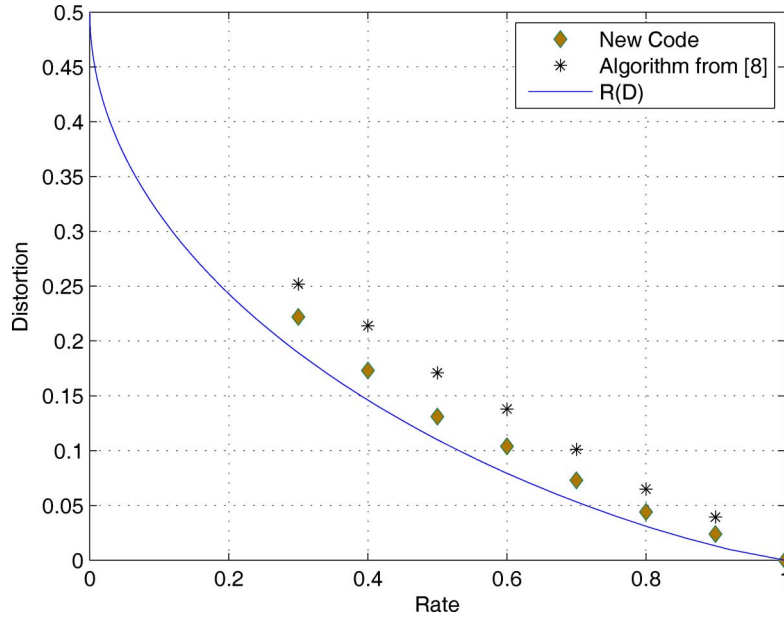


Fig. 1. Empirical performance of the code in Section II-A with the suboptimal encoding algorithm in Section IV compared with LDGM codes and the message-passing heuristic from [8], for the binary symmetric source with bit-error-rate distortion and block length $n = 400$.

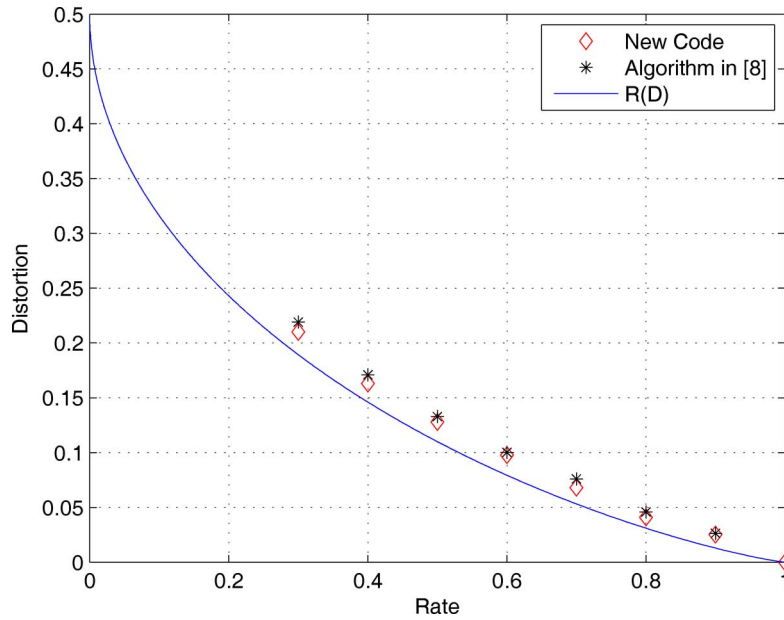


Fig. 2. Empirical performance of the code in Section II-A with the suboptimal encoding algorithm in Section IV compared with LDGM codes and the message-passing heuristic from [8], for the binary symmetric source with bit-error-rate distortion and block length $n = 1000$.

APPENDIX II
AUXILIARY RESULTS

Lemma 14: Let $\{\mathbf{c}[1], \mathbf{c}[2], \dots, \mathbf{c}[n]\}$ be independent and identically distributed binary random variables such that

$$\mathbb{P}[\mathbf{c}[j] = 1] = p \tag{136}$$

where $p < 1/2$. Let \mathbf{s} denote an arbitrary sequence in $\{0, 1\}^n$ and let $\mathbf{0} = [0, 0, \dots, 0]^T$, $\mathbf{1} = [1, 1, \dots, 1]^T$, then for $d \in [0, 1]$

$$\mathbb{P}[w_H(\mathbf{1} \oplus \mathbf{c}) \leq nd] \leq \mathbb{P}[w_H(\mathbf{s} \oplus \mathbf{c}) \leq nd] \tag{137}$$

$$\leq \mathbb{P}[w_H(\mathbf{0} \oplus \mathbf{c}) \leq nd]. \tag{138}$$

Proof: Let b_l, \bar{b}_k $l \neq k$ be independent binary random variables with

$$\mathbb{P}[b_l = 1] = p = \mathbb{P}[\bar{b}_k = 0]. \tag{139}$$

If we show that for arbitrary $j \in \{1, \dots, n-1\}$

$$\mathbb{P}\left[\sum_{i=1}^j b_i + \sum_{i=j+1}^n \bar{b}_i \leq nd\right] \leq \mathbb{P}\left[\sum_{i=1}^{j+1} b_i + \sum_{i=j+2}^n \bar{b}_i \leq nd\right] \tag{140}$$

then (138) follows by induction. Let

$$\mathbb{P}\left[\sum_{i=1}^j b_i + \sum_{i=j+2}^n \bar{b}_i \leq nd\right] = \nu \tag{141}$$

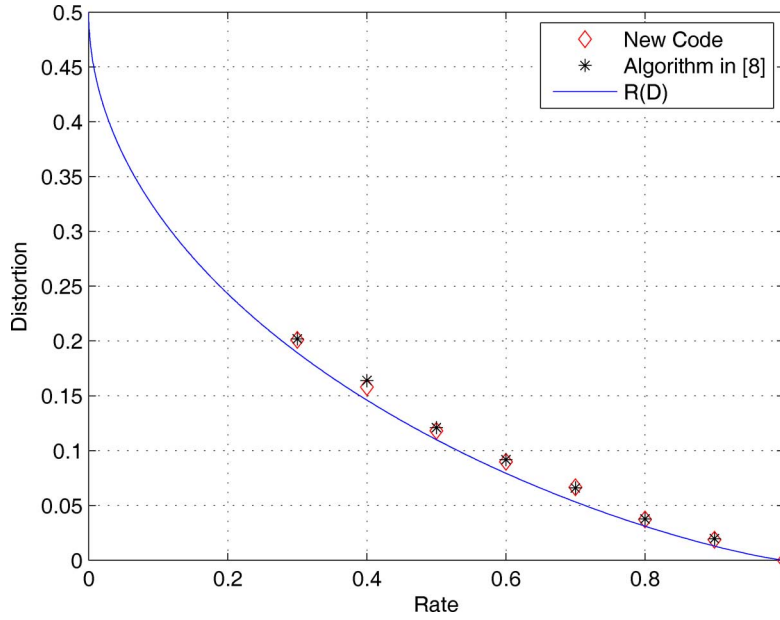


Fig. 3. Empirical performance of the code in Section II-A with the suboptimal encoding algorithm in Section IV compared with LDGM codes and the message-passing heuristic from [8], for the binary symmetric source with bit-error-rate distortion and block length $n = 2000$.

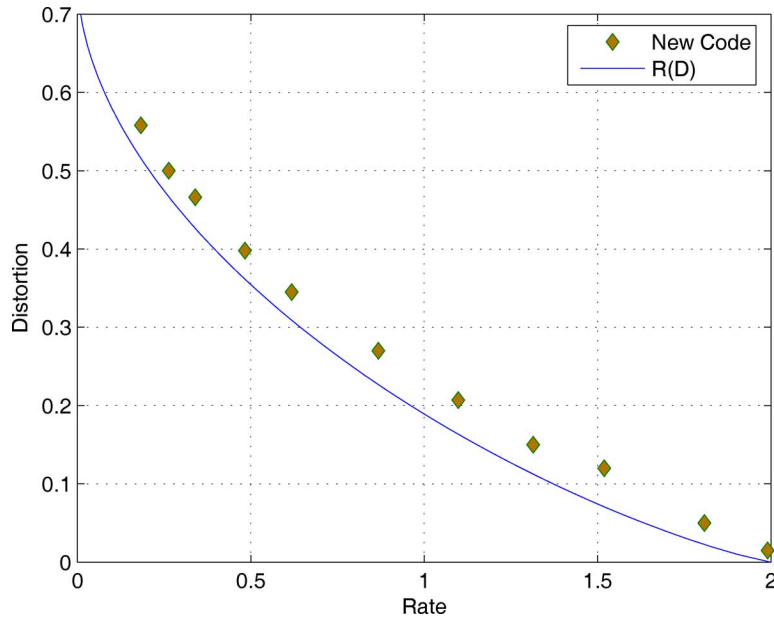


Fig. 4. Empirical performance of the code in Section III for compressing the 4-ary source with block length $n = 100$ and Hamming distortion criterion.

and

$$+ \mathbb{P}[b_{j+1} = 0] \mathbb{P} \left[\sum_{i=1}^j b_i + \sum_{i=j+2}^n \bar{b}_i \leq nd \right] \quad (143)$$

$$\mathbb{P} \left[\sum_{i=1}^j b_i + \sum_{i=j+2}^n \bar{b}_i \leq nd - 1 \right] = \tau, \quad (142) \quad = p\tau + (1-p)\nu. \quad (144)$$

Similarly

$$\mathbb{P} \left[\sum_{i=1}^j b_i + \sum_{i=j+1}^n \bar{b}_i \leq nd \right] = (1-p)\tau + p\nu. \quad (145)$$

Clearly $\tau \leq \nu$, further

Since $(1-p)\tau + p\nu \leq p\tau + (1-p)\nu$, (140) holds. Using (140) and induction on j we have (137) and (138). \square

$$\begin{aligned} & \mathbb{P} \left[\sum_{i=1}^{j+1} b_i + \sum_{i=j+2}^n \bar{b}_i \leq nd \right] \\ &= \mathbb{P}[b_{j+1} = 1] \mathbb{P} \left[\sum_{i=1}^j b_i + \sum_{i=j+2}^n \bar{b}_i \leq nd - 1 \right] \end{aligned}$$

Lemma 15: Let $B_i \ i \in \{1, 2, \dots, n\}$ be defined as in (68) then

$$|B_{i+1} - B_i| \leq 1. \quad (146)$$

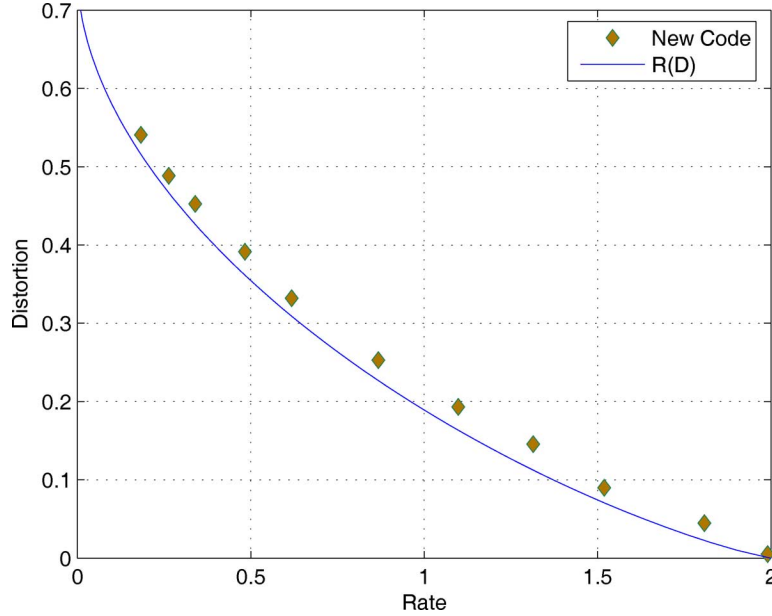


Fig. 5. Empirical performance of the new code for compressing the 4-ary source with block length $n = 400$ and Hamming distortion criterion.

Proof: For a given source realization \mathbf{s} let

$$f(\mathbf{G}) = \min_{j \in \{1, \dots, M\}} w_H(\mathbf{c}_j \oplus \mathbf{s}); \quad (147)$$

further note that the distribution function of $\mathbf{g}_j, \dots, \mathbf{g}_n$ satisfies

$$P_{\mathbf{g}_j, \dots, \mathbf{g}_n}(x_1, \dots, x_n) = \prod_i P_{\mathbf{g}_i}(x_i) \quad (148)$$

because, by construction, the rows $\mathbf{g}_1, \dots, \mathbf{g}_n$ of the matrix \mathbf{G}^T are independent.

Conditioning on $\mathbf{g}_1 = l_1, \dots, \mathbf{g}_{i+1} = l_{i+1}$ we obtain

$$\begin{aligned} B_{i+1} - B_i &= \sum_{x_{i+2}, \dots, x_n} P_{\mathbf{g}_{i+2}, \dots, \mathbf{g}_n}(x_{i+2}, \dots, x_n) \\ &\quad \times f([l_1, \dots, l_{i+1}, x_{i+2}, \dots, x_n]) \\ &\quad - \sum_{x_{i+1}, \dots, x_n} P_{\mathbf{g}_{i+1}, \dots, \mathbf{g}_n}(x_{i+1}, \dots, x_n) \\ &\quad \times f([l_1, \dots, l_i, x_{i+1}, \dots, x_n]) \quad (149) \\ &= \sum_{x_{i+1}, \dots, x_n} P_{\mathbf{g}_{i+1}, \dots, \mathbf{g}_n}(x_{i+1}, \dots, x_n) \\ &\quad \times (f([l_1, \dots, l_{i+1}, x_{i+2}, \dots, x_n]) \\ &\quad - f([l_1, \dots, l_i, x_{i+1}, \dots, x_n])). \quad (150) \end{aligned}$$

A change in one column of the \mathbf{G} matrix can change $f(\mathbf{G})$ by at most one. Therefore

$$\begin{aligned} |B_{i+1} - B_i| &\leq \sum_{x_{i+1}, \dots, x_n} P_{\mathbf{g}_{i+1}, \dots, \mathbf{g}_n}(x_{i+1}, \dots, x_n) \\ &\quad \times |f([l_1, \dots, l_{i+1}, x_{i+2}, \dots, x_n]) \\ &\quad - f([l_1, \dots, l_i, x_{i+1}, \dots, x_n])| \quad (151) \\ &\leq \sum_{x_{i+1}, \dots, x_n} P_{\mathbf{g}_{i+1}, \dots, \mathbf{g}_n}(x_{i+1}, \dots, x_n) \quad (152) \\ &= 1 \quad (153) \end{aligned}$$

regardless of l_1, \dots, l_n . Therefore, (146) holds. \square

Note that Lemma 15 was shown for the code construction in [11], however, we cannot use the proof therein due to the difference in code construction.

Lemma 16: For $\mathbf{x} \in \mathcal{E}^n$ and every $\epsilon > 0$ there exists an $n(\epsilon)$ such that for $n > n(\epsilon)$

$$\Lambda(\lambda) - \epsilon < \frac{1}{n} \Lambda_n(\mathbf{x}, \lambda) < \Lambda(\lambda) + \epsilon \quad (154)$$

where \mathcal{E}^n , $\Lambda_n(\mathbf{x}, \lambda)$, and $\Lambda(\lambda)$ are defined in (119), (116), and (118), respectively.

Proof:

$$\begin{aligned} &\frac{1}{n} \sum_{l=1}^n \log \frac{\mathbb{E}[e^{\lambda d(\mathbf{x}[l], \hat{\mathcal{S}}(n))}]}{\mathbb{E}[e^{\lambda d(\mathbf{x}[l], \hat{\mathcal{S}})}]} \\ &= \frac{1}{n} \sum_{l=1}^n \log \left(1 + \frac{\mathbb{E}[e^{\lambda d(\mathbf{x}[l], \hat{\mathcal{S}}(n))}] - \mathbb{E}[e^{\lambda d(\mathbf{x}[l], \hat{\mathcal{S}})}]}{\mathbb{E}[e^{\lambda d(\mathbf{x}[l], \hat{\mathcal{S}})}]} \right). \quad (155) \end{aligned}$$

From Lemma 11, for every $\varsigma > 0$ there exists an $n_1(\varsigma)$ such that for $n > n_1(\varsigma)$

$$\sum_{\hat{a} \in \hat{\mathcal{A}}} |P_{\hat{\mathcal{S}}(n)}(\hat{a}) - P_{\hat{\mathcal{S}}}(\hat{a})| \leq \varsigma. \quad (156)$$

Therefore, for $n > n_1(\varsigma)$

$$\begin{aligned} &\log(1 - \varsigma e^{\lambda d_{\max}}) \\ &< \log \left(1 + \frac{\mathbb{E}[e^{\lambda d(\mathbf{x}[l], \hat{\mathcal{S}}(n))}] - \mathbb{E}[e^{\lambda d(\mathbf{x}[l], \hat{\mathcal{S}})}]}{\mathbb{E}[e^{\lambda d(\mathbf{x}[l], \hat{\mathcal{S}})}]} \right) \quad (157) \\ &< \log(1 + \varsigma e^{\lambda d_{\max}}). \quad (158) \end{aligned}$$

Choosing $\varsigma = \min[\frac{e^{\epsilon'} - 1}{e^{\lambda d_{\max}}}, \frac{1 - e^{-\epsilon'}}{e^{\lambda d_{\max}}}]$, for $n > n_2(\epsilon') \triangleq n_1(\varsigma)$

$$-\epsilon' < \frac{1}{n} \sum_{l=1}^n \log \frac{\mathbb{E}[e^{\lambda d(\mathbf{x}[l], \hat{\mathcal{S}}(n))}]}{\mathbb{E}[e^{\lambda d(\mathbf{x}[l], \hat{\mathcal{S}})}]} < \epsilon'. \quad (159)$$

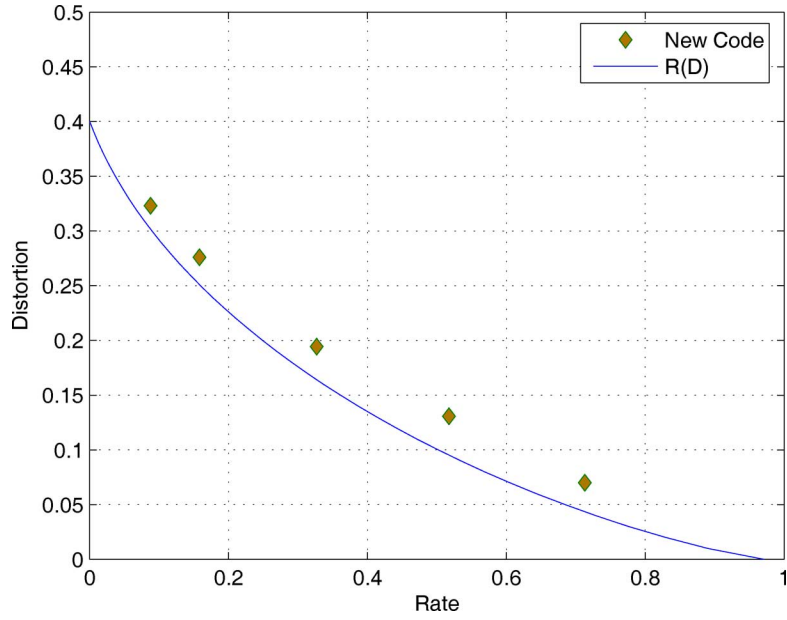


Fig. 6. Empirical performance of the code for compressing the Bernoulli ($p = 0.4$) source with block length $n = 1000$ and bit-error-rate distortion criterion.

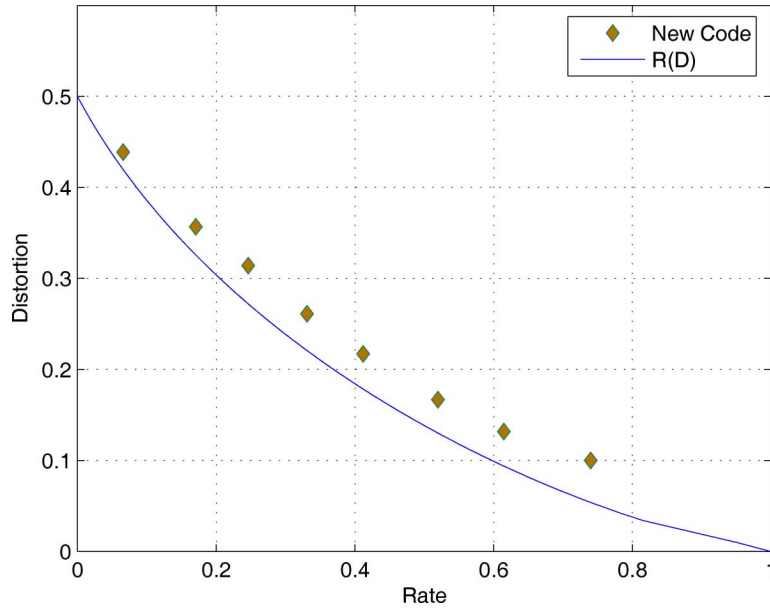


Fig. 7. Empirical performance of the new code for compressing the binary symmetric source with block length $n = 1000$ and the asymmetric distortion criterion in (128).

Substituting $\Lambda_n(\mathbf{x}, \lambda)$ from (116) in (159) and using (119), for $\mathbf{x} \in \mathcal{E}^n$

$$-\epsilon' - \frac{1}{\sqrt[4]{n}} + \Lambda(\lambda) < \frac{1}{n} \Lambda_n(\mathbf{x}, \lambda) < \Lambda(\lambda) + \frac{1}{\sqrt[4]{n}} + \epsilon'. \quad (160)$$

Choosing $\epsilon' = \epsilon/2$ and $n(\epsilon) = \max\{n_2(\epsilon/2), (\epsilon/2)^{-4}\}$ we get the required result. \square

Lemma 17: Let B_i $i \in \{1, 2, \dots, n\}$ be defined as in (114) then

$$|B_{i+1} - B_i| \leq d_{\max}. \quad (161)$$

Proof: For fixed \mathbf{x} let

$$f_{\mathbf{x}}(\mathbf{G}) = \min_{j \in \{1, \dots, M\}} d^n(\mathbf{x}, \mathbf{c}_j). \quad (162)$$

Using (149)–(150) and replacing $f(\cdot)$ by $f_{\mathbf{x}}(\cdot)$, we have

$$|B_{i+1} - B_i| \leq \sum_{x_{i+1}, \dots, x_n} P_{\mathbf{g}_{i+1}, \dots, \mathbf{g}_n}(x_{i+1}, \dots, x_n) \times |f_{\mathbf{x}}([l_1, \dots, l_{i+1}, x_{i+2}, \dots, x_n]) - f_{\mathbf{x}}([l_1, \dots, l_i, x_{i+1}, \dots, x_n])| \quad (163)$$

$$\leq \sum_{x_{i+1}, \dots, x_n} P_{\mathbf{g}_{i+1}, \dots, \mathbf{g}_n}(x_{i+1}, \dots, x_n) d_{\max} \quad (164)$$

$$= d_{\max}. \quad (165)$$

We get (164) because a change in one column of the \mathbf{G} matrix can change $f_{\mathbf{x}}(\mathbf{G})$ by at most d_{\max} . \square

Lemma 18: Let $j_n \in \{1, \dots, M\}$ be such that

$$w_H(\mathbf{u}_{j_n} \oplus \mathbf{u}_1) > \frac{\lceil k\omega \rceil}{\log_2 n} \quad (166)$$

then for all $(a, b) \in \{0, 1\}^2$

$$\lim_{n \rightarrow \infty} P_n(a, b) = \frac{1}{4} \quad (167)$$

where $P_n(a, b) = \mathbb{P}[\mathbf{c}_{j_n}[m] = a, \mathbf{c}_1[m] = b]$, which does not depend on m .

Proof: Let

$$f_n = \frac{w_H(\mathbf{u}_{j_n} \oplus \mathbf{u}_1)}{2\lceil k\omega \rceil}. \quad (168)$$

From (166), $f_n > 1/(2\log_2 n)$. Define $\mathbf{u}'_1 = \mathbf{w}_{1j_n}$, and $\mathbf{u}'_2 = \mathbf{w}_{j_n 1}$ where

$$\mathbf{w}_{ij}[m] = \mathbf{u}_i[m] \cdot \bar{\mathbf{u}}_j[m] \quad (169)$$

and

$$\mathbf{u}'[m] = \mathbf{u}_{j_n}[m] \cdot \mathbf{u}_1[m] \quad (170)$$

where \cdot denotes the logical AND operation and \bar{a} denotes the complement of a . The vectors \mathbf{u}'_1 , \mathbf{u}'_2 , and \mathbf{u}' are nonoverlapping in the sense that

$$\mathbf{u}'_1 \cdot \mathbf{u}'_2 = \mathbf{u}'_1 \cdot \mathbf{u}' = \mathbf{u}'_2 \cdot \mathbf{u}' = \mathbf{0}. \quad (171)$$

Further

$$\mathbf{u}_1[m] = \mathbf{u}'_1[m] \oplus \mathbf{u}'[m] \quad (172)$$

and

$$\mathbf{u}_{j_n}[m] = \mathbf{u}'_2[m] \oplus \mathbf{u}'[m] \quad (173)$$

where \oplus denotes the logical XOR operation. Further

$$w_H[\mathbf{u}'_1] = w_H[\mathbf{u}'_2] = f_n \lceil k\omega \rceil \quad (174)$$

and

$$w_H[\mathbf{u}'] = (1 - f_n) \lceil k\omega \rceil. \quad (175)$$

Denote $\mathbf{G}^T \mathbf{u}'_1$, $\mathbf{G}^T \mathbf{u}'_2$, and $\mathbf{G}^T \mathbf{u}'$ as \mathbf{c}'_1 , \mathbf{c}'_2 , and \mathbf{c}' , respectively. These vectors are mutually independent since \mathbf{u}'_1 , \mathbf{u}'_2 , and \mathbf{u}' are nonoverlapping. Further, $\mathbf{c}'_1[m] = 1$ if and only if the $f_n \lceil k\omega \rceil$ ones in \mathbf{u}'_1 select an odd number of ones in the m th row of \mathbf{G}^T matrix. The probability of this event satisfies (see (20)–(22) substituting $\lceil k\omega \rceil$ by $f_n \lceil k\omega \rceil$)

$$\mathbb{P}[\mathbf{c}'_1[m] = 1] = \frac{1}{2} \left[1 - e^{-\frac{2R \log^2 n}{\log \log n} f_n} \right] + o(1) \quad (176)$$

$$\rightarrow \frac{1}{2} \quad (177)$$

and analogously for \mathbf{c}'_2 . We get (177) from (176) because $\frac{2R \log^2 n}{\log \log n} f_n \rightarrow \infty$, for $f_n > \frac{1}{2\log_2 n}$.

If V, X_1, X_2 are independent binary random variables with

$$\lim_{n \rightarrow \infty} \mathbb{P}[X_i = 0] = \frac{1}{2} \quad (178)$$

then

$$\lim_{n \rightarrow \infty} \mathbb{P}[V \oplus X_1 = a, V \oplus X_2 = b] = \frac{1}{4} \quad (179)$$

for all $(a, b) \in \{0, 1\}^2$. Identifying $X_1 = \mathbf{c}'_1[m]$, $X_2 = \mathbf{c}'_2[m]$, and $V = \mathbf{c}'[m]$ we get (167). \square

Lemma 19: For a sequence of vectors $\mathbf{v}_n, \mathbf{w}_n$, and a sequence of integers j_n satisfying (166)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 = \mathbf{v}_n, \mathbf{c}_{j_n} = \mathbf{w}_n] = -2. \quad (180)$$

Proof: Obviously

$$\min_{(a,b) \in \{0,1\}^2} \log_2 P_n(a, b) \leq \frac{1}{n} \sum_{i=1}^n \log_2 P_n(\mathbf{v}_n[i], \mathbf{w}_n[i]) \quad (181)$$

$$\leq \max_{(a,b) \in \{0,1\}^2} \log_2 P_n(a, b) \quad (182)$$

but for any n , Lemma 3 yields

$$\frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 = \mathbf{v}_n, \mathbf{c}_{j_n} = \mathbf{w}_n] = \frac{1}{n} \sum_{i=1}^n \log_2 P_n(\mathbf{v}_n[i], \mathbf{w}_n[i]). \quad (183)$$

Therefore, (180) follows from Lemma 18. \square

Lemma 20: For any sequence of deterministic vectors \mathbf{v}_n

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 = \mathbf{v}_n] = -1. \quad (184)$$

Proof: The proof is similar to the proof of Lemma 19. Denoting the common distribution of $\{\mathbf{c}_1[m]\}_{m=1}^n$ by P_{W_n} , we have

$$\frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 = \mathbf{v}_n] = \frac{1}{n} \sum_{i=1}^n \log_2 P_{W_n}(\mathbf{v}_n[i]). \quad (185)$$

Therefore

$$\min_{a \in \{0,1\}} \log_2 P_{W_n}(a) \leq \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 = \mathbf{v}_n] \quad (186)$$

$$\leq \max_{a \in \{0,1\}} \log_2 P_{W_n}(a) \quad (187)$$

using Lemma 3

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 = \mathbf{v}_n] = -1. \quad (188)$$

\square

Lemma 21:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 \in \mathcal{B}_n] = \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 |\mathcal{B}_n| - 1 \quad (189)$$

whenever the limits exist.

Proof: Let

$$[\mathbf{v}^{\min}] = \arg \min_{\mathbf{v} \in \mathcal{B}_n} \mathbb{P}[\mathbf{c}_1 = \mathbf{v}] \quad (190)$$

and

$$[\mathbf{v}^{\max}] = \arg \max_{\mathbf{v} \in \mathcal{B}_n} \mathbb{P}[\mathbf{c}_1 = \mathbf{v}]. \quad (191)$$

We have

$$\frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 = \mathbf{v}^{min}] + \frac{1}{n} \log_2 |\mathcal{B}_n| \leq \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 \in \mathcal{B}_n] \tag{192}$$

$$\leq \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 = \mathbf{v}^{max}] + \frac{1}{n} \log_2 |\mathcal{B}_n|. \tag{193}$$

Taking limits and using Lemma 20 the result follows. \square

Lemma 22: For a sequence of sets $\mathcal{B}_n \subset \{0, 1\}^n$ and a sequence j_n satisfying (166)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_{j_n} \in \mathcal{B}_n, \mathbf{c}_1 \in \mathcal{B}_n] = 2 \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 |\mathcal{B}_n| - 2 \tag{194}$$

whenever the limits exist.

Proof: For each n let

$$[\mathbf{v}_n^{min}, \mathbf{w}_n^{min}] = \arg \min_{\mathbf{v} \in \mathcal{B}_n, \mathbf{w} \in \mathcal{B}_n} \mathbb{P}[\mathbf{c}_1 = \mathbf{v}, \mathbf{c}_{j_n} = \mathbf{w}] \tag{195}$$

and

$$[\mathbf{v}_n^{max}, \mathbf{w}_n^{max}] = \arg \max_{\mathbf{v} \in \mathcal{B}_n, \mathbf{w} \in \mathcal{B}_n} \mathbb{P}[\mathbf{c}_1 = \mathbf{v}, \mathbf{c}_{j_n} = \mathbf{w}]. \tag{196}$$

Therefore

$$\begin{aligned} & \frac{2}{n} \log_2 |\mathcal{B}_n| + \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 = \mathbf{v}_n^{min}, \mathbf{c}_{j_n} = \mathbf{w}_n^{min}] \\ & \leq \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_{j_n} \in \mathcal{B}_n, \mathbf{c}_1 \in \mathcal{B}_n] \end{aligned} \tag{197}$$

$$\leq \frac{2}{n} \log_2 |\mathcal{B}_n| + \frac{1}{n} \log_2 \mathbb{P}[\mathbf{c}_1 = \mathbf{v}_n^{max}, \mathbf{c}_{j_n} = \mathbf{w}_n^{max}]. \tag{198}$$

Taking limits and using Lemma 19 we get (194). \square

ACKNOWLEDGMENT

We wish to thank the referees for their help in improving the presentation.

REFERENCES

[1] I. Csiszár and J. Körner, *Information Theory, Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
 [2] G. Caire, S. Shamai (Shitz), and S. Verdú, "Lossless data compression with error correcting codes," in *Advances in Network Information Theory*. Providence, RI: Amer. Math. Soc., 2004, vol. 66, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, pp. 263–284.
 [3] T. Anчета, "Bounds and Techniques for Linear Source Coding," Ph.D. dissertation, Dep. Elec. Eng., Univ. Notre Dame, Notre Dame, IN, 1977.
 [4] J. L. Massey, "Joint source and channel coding," *Commun. Syst. Random Process Theory*, vol. 11, pp. 279–293, 1978.
 [5] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Eaglewood Cliffs, NJ: Prentice-Hall, 1971.
 [6] J. Chen, D. He, and A. Jagmohan, "Achieving the rate-distortion bound with linear codes," in *Proc. 2007 IEEE Information Theory Workshop*, Lake Tahoe, CA, Sep. 2007, pp. 662–667.
 [7] Y. Matsunaga and H. Yamamoto, "A coding theorem for lossy data compression by LDPC codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2225–2229, Sep. 2003.

[8] M. J. Wainwright and E. Maneva, "Lossy source encoding via message passing and decimation over generalized codewords of LDGM codes," in *Proc. 2005 IEEE Int. Symp. Information Theory*, Adelaide, Australia, Sep. 2005, pp. 1493–1497.
 [9] S. Ciliberti, K. Mezard, and R. Zecchina, "Message passing algorithms for non-linear nodes and data compression," *ComplexUs*, vol. 3, pp. 58–65, Aug. 2006.
 [10] A. Braunstein, K. Mezard, and R. Zecchina, "Survey propagation: An algorithm for satisfiability," *Random Structures and Algorithms*, vol. 27, pp. 201–226, Mar. 2005.
 [11] E. Martinian and M. J. Wainwright, "Low-density codes achieve the rate-distortion bound," in *Proc. 2006 Data Compression Conf.*, Snowbird, UT, Mar. 2006, pp. 153–162.
 [12] S. Miyake, "Lossy data compression over Z_q by LDPC code," in *Proc. 2006 IEEE Int. Symp. Information Theory*, Seattle, WA, Jul. 2006, pp. 813–816.
 [13] S. Miyake and J. Muramatsu, "Construction of a lossy source code using LDPC matrices," in *Proc. 2007 IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 1106–1110.
 [14] E. Martinian and M. J. Wainwright, "Analysis of LDGM and compound codes for lossy compression and binning," in *Proc. 2006 Workshop on Information Theory and its Applications*, La Jolla, CA, Feb. 2006.
 [15] S. Kudekar and R. Urbanke, "Lower bounds on the rate-distortion function of individual LDGM codes," in *Proc. 5th Int. Symp. Turbo Codes and Related Topics*, Lausanne, Switzerland, Sep. 2008, pp. 379–384.
 [16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley Interscience, 2006.
 [17] K. Azuma, "Weighted sums of certain dependent random variables," *Tohoku Math. J.*, vol. 19, pp. 357–367, 1967.
 [18] A. Gupta and S. Verdú, "Nonlinear sparse-graph codes for lossy compression of discrete nonredundant sources," in *Proc. 2007 IEEE Information Theory Workshop*, Lake Tahoe, CA, Sep. 2007, pp. 541–546.
 [19] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
 [20] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York: Springer, 2004.
 [21] A. Dembo and I. Kontoyiannis, "Source coding, large deviations and approximate pattern matching," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1590–1615, Jun. 2002.
 [22] E. R. Berlekamp, R. J. McEliece, and H. C. A. van Tilborg, "On the intractability of certain coding problems," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 384–386, May 1978.
 [23] T. M. Cover, "Enumerative source encoding," *IEEE Trans. Inf. Theory*, vol. IT-10, no. 1, pp. 460–473, Jan. 1973.

Ankit Gupta (S'07) received the B.Tech. degree in 2003 from Indian Institute of Technology, Delhi, India, and the M.A. degree in 2006 from Princeton University, Princeton, NJ, both in electrical engineering. He is currently pursuing the Ph.D. degree in electrical engineering at Princeton University.

Sergio Verdú (S'80–M'84–SM'88–F'03) received the Telecommunications Engineering degree from the Universitat Politècnica de Barcelona, Barcelona, Spain, in 1980 and the Ph.D. degree in electrical engineering from the University of Illinois at Urbana-Champaign, Urbana, IL, in 1984. Since 1984, he has been a member of the faculty of Princeton University, Princeton, NJ, where he is the Eugene Higgins Professor of Electrical Engineering. Sergio Verdú is the recipient of the 2007 Claude E. Shannon Award and the 2008 IEEE Richard W. Hamming Medal. He is a member of the National Academy of Engineering and was awarded a Doctorate *Honoris Causa* from the Universitat Politècnica de Catalunya in 2005. He is a recipient of several paper awards from the IEEE: the 1992 Donald Fink Paper Award, the 1998 Information Theory Outstanding Paper Award, an Information Theory Golden Jubilee Paper Award, the 2002 Leonard Abraham Prize Award, and the 2006 Joint Communications/Information Theory Paper Award. In 1998, Cambridge University Press published his book *Multiuser Detection*, for which he received the 2000 Frederick E. Terman Award from the American Society for Engineering Education. He served as President of the IEEE Information Theory Society in 1997 and as Associate Editor for Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY. He is currently Editor-in-Chief of *Foundations and Trends in Communications and Information Theory*.