Rate-Distortion in Near-Linear Time

Ankit Gupta*, Sergio Verdù* and Tsachy Weissman†‡,
*Department of Electrical Engineering, Princeton University, Princeton, NJ 08540, {ankitg,verdu}@princeton.edu
† Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA, tsachy@stanford.edu
‡Department of Electrical Engineering, Technion, Haifa 32000, Israel

***Abstract*—We present two results related to the computational complexity of lossy compression. The first result shows that for a memoryless source P_S with rate-distortion function $R(D)$, the rate-distortion pair $(R(D)+\gamma, D+\epsilon)$ can be achieved with constant decoding time per symbol and encoding time per symbol proportional to $C_1(\gamma)\epsilon^{-C_2(\gamma)}$. The second results establishes that for any given R, there exists a universal lossy compression scheme with $O(ng(n))$ encoding complexity and $O(n)$ decoding complexity, that achieves the point $(R, D(R))$ asymptotically for any ergodic source with distortion-rate function $D(\cdot)$, where $g(n)$ is an arbitrary non-decreasing unbounded function. A computationally feasible implementation of the first scheme outperforms many of the best previously proposed schemes for binary sources with blocklengths of the order of 1000.**

I. Introduction

Rate-distortion theory assures us that for an ergodic source with rate-distortion function $R(D)$, there exists a lossy compression scheme that achieves the rate-distortion pair $(R(D), D)$ asymptotically. However all achievability proofs in the literature rely on schemes with encoding complexity growing at least exponentially in the blocklength n. An interesting question is whether there exists a compression/decompression scheme with low complexity that can achieve the optimal performance asymptotically.

Asymptotically optimal block codes based on sparse matrices have been proposed for the nonredundant source with Hamming distortion criterion [1], [2] and [3]. Unfortunately, these schemes rely on the exponential time ML encoder for asymptotic optimality. Other schemes based on sparse-graph codes and low complexity message-passing encoders are not theoretically optimal, but perform well in practice [4], [5]. Recently, sparse graph codes have been constructed which combine asymptotically optimality for a general discrete memoryless source and an arbitrary separable distortion criterion under ML encoding and suboptimal quadratic encoders with good empirical performance [6]. A universal, asymptotically optimal lossy compression scheme, for memoryless sources is proposed in [7], however the scheme performs far away from $R(D)$, for blocklengths of the order of 1000 for the simple case of compressing the Bernoulli source with Hamming distortion criterion. An asymptotically optimal low (polynomial) complexity compressor with near optimal empirical performance, has not yet been found in the literature.

In this paper we construct lossy compressors by partitioning a source string of length n into words of length l each and use a short block code (with blocklength equal to l) to encode each of these words (Figure 1). The short code has the same rate as the long code (with blocklength equal to n). If the average distortion (over the source realization) achieved by the short code is less than $D+\epsilon$, then the probability that the long code distortion exceeds $D+\epsilon$ vanishes asymptotically. We show that that there exists a short code with rate $R(D)+\gamma$ that achieves average distortion less than $D+\epsilon$ and has blocklength $l \leq g_1(\gamma)\log(1/\epsilon)+g_2(\gamma)$. By concatenating this code (as shown in Figure 1) we obtain a lossy compressor with constant decoding time per symbol and compression time proportional to $C_1(\gamma)\epsilon^{-C_2(\gamma)}$. Using the code construction in Figure 1, we also show the existence of a universal lossy compressor with encoding complexity $O(ng(n))$ (for a non-decreasing unbounded $g(n)$) and decoding complexity $O(n)$, that asymptotically achieves the rate-distortion pair $(R, D(R))$ for any ergodic source with distortion-rate function $D(\cdot)$.

The rest of the paper is organized as follows. In Section II we present the code design. In Section III we use the code design in Section II to obtain a scheme that has rate $R(D)+\gamma$, and achieves distortion less than $D+\epsilon$ asymptotically with encoding time per symbol proportional to $C_1(\gamma)\epsilon^{-C_2(\gamma)}$. In Section IV we show the existence of a universal lossy compression scheme with near-linear encoding complexity and linear decoding complexity that asymptotically achieves the optimal rate-distortion tradeoff for any ergodic source. In Section IV we also show a near-linear complexity extension of the lossy compression scheme in Section III, that asymptotically achieves the optimal rate-distortion tradeoff. In contrast to the universal scheme which is only of theoretical interest, this scheme is also amenable to practical implementation as demonstrated by the empirical results in Section V. In Section V we show empirical results obtained when the scheme in Section III is used to encode binary sources with various distortion criteria and compare its performance to other schemes in the literature.

II. Code construction

The code construction in this section results in a "long" lossy compressor by bootstrapping a "short" lossy compression code $\mathcal{C}_0$ (consisting of 2^{lR} codewords) that encodes a vector of length l into a binary vector of length lR, with average distortion equal to

$$\mathbb{E}[d^l(S_1, S_2, \cdots, S_l; \mathcal{C}_0)] = D_0, \tag{1}$$

where $d^l(s_1, s_2, \cdots, s_l; \mathcal{C}_0)$ is the distance from $(s_1, s_2, \cdots, s_l)$, to its nearest neighbor in $\mathcal{C}_0$. For an

978-1-4244-2571-6/08/$25.00 ©2008 IEEE

integer k, we now construct a lossy compressor for a vector of length $kl = n$ as follows:

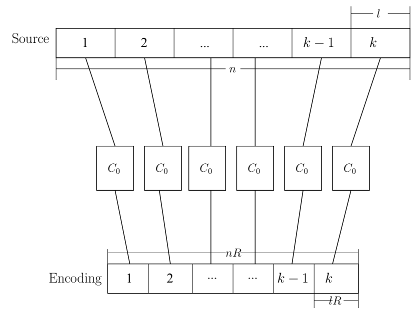

Fig. 1. Encoding procedure

A. Encoding

Given a source realization $\mathbf{s}$, we partition it into k words of length l each. Each word is encoded by the index of the nearest codeword in $\mathcal{C}_0$ (Figure 1), since the code contains 2^{lR} codewords this can be done in $l2^{lR}$ operations. Thus encoding time per symbol is equal to $K_1 2^{lR}$, where K_1 depends only on the underlying computational model. The index of the corresponding codeword is represented by lR bits. The output of the encoder is the sequence of k blocks of lR bits each i.e. nR bits.

B. Decoding

Using the indices corresponding to each block, the decoder outputs the sequence of codewords corresponding to each of the k words using l symbols for each word. Total number of operations required is equal to $lk = n$, thus the decoding time per symbol is a constant K_2 which depends only on the underlying computational model.

III. Asymptotic Optimality and Computational cost per symbol

Theorem 1. *Given a memoryless source, a separable bounded distortion criterion and a point $(R(D), D)$ on the rate-distortion function, for arbitrary $\epsilon, \gamma > 0$ there exists a compression/decompression scheme with blocklength n and rate $R = R(D) + \gamma$ such that the probability that the normalized distortion between the source realization and its reproduction exceeds $D + \epsilon$ vanishes with $n \to \infty$, and*

1) *Encoding time per symbol is proportional to*[1] $C_1(\gamma)\epsilon^{-C_2(\gamma)}$.
2) *Decoding time per symbol is independent of ϵ and γ.*

We will show this theorem using the following result.

Lemma 1. *Given a memoryless source with rate-distortion function $R(D)$, $\gamma > 0$, $\epsilon > 0$ and $D > 0$, there exists a code $\mathcal{C}_0$ with rate $R = R(D) + \gamma$ and blocklength l, such that*

$$\mathbb{E}[d^l(S^l, \mathcal{C}_0)] < D + \epsilon, \tag{2}$$

and

$$l = g_1(\gamma)\log\frac{1}{\epsilon} + g_2(\gamma). \tag{3}$$

[1] $C_1(\gamma)$, $C_2(\gamma)$ do not dependent on ϵ.

Proof. Denoting the distortion-rate function by $R^{-1}(\cdot)$, let $\bar{D} = R^{-1}(R(D) + \gamma/2) = D - g(\gamma)$, Solve the optimization problem

$$R(\bar{D}) = \min_{\substack{P_{\hat{S}|S} \\ \mathbb{E}[d(S,\hat{S})] \le \bar{D}}} I(S; \hat{S}), \tag{4}$$

to obtain $P_{\hat{S}}^{(\bar{D})}$ and $P_{\hat{S},S}^{(\bar{D})}$. For now choose an arbitrary integer l. Let $\mathcal{C}^l$ denote a random codebook consisting of $2^{l(R(D)+\gamma)}$ codewords of length l, such that all symbols are independent and identically distributed with the distribution $P_{\hat{S}}^{(\bar{D})}$. Define the following subsets of $\mathcal{A}^l \times \hat{\mathcal{A}}^l$

$$\mathcal{D} = \{\mathbf{x}, \mathbf{y} : d^l(\mathbf{x}, \mathbf{y}) \le D\}, \tag{5}$$

and

$$\mathcal{T} = \left\{ \mathbf{x}, \mathbf{y} : \frac{1}{l}\log_2 \frac{P_{\hat{S}^l|S^l}^{(\bar{D})}(\mathbf{y}|\mathbf{x})}{P_{\hat{S}^l}^{(\bar{D})}(\mathbf{y})} \le R(D) + \frac{3\gamma}{4} \right\}. \tag{6}$$

The sets $\mathcal{D}^c$ and $\mathcal{T}^c$ are composed of the pairs $(\mathbf{x}, \mathbf{y})$ such that the empirical averages $\frac{1}{l}\sum_{i=1}^{l} d(\mathbf{x}[i], \mathbf{y}[i])$ and $\frac{1}{l}\sum_{i=1}^{l} \log P_{\hat{S}|S}^{(\bar{D})}(\mathbf{y}[i]|\mathbf{x}[i]) / P_{\hat{S}}^{(\bar{D})}(\mathbf{y}[i])$ differ from the expected value by a gap which is a function of γ. Therefore, using the Chernoff bound

$$P_{S^l\hat{S}^l}^{(\bar{D})}(\mathcal{D}^c) \le \exp(-l(f_1(\gamma)), \tag{7}$$

and

$$P_{S^l\hat{S}^l}^{(\bar{D})}(\mathcal{T}^c) \le \exp(-l(f_2(\gamma))), \tag{8}$$

where $f_1(\gamma), f_2(\gamma)$ are large deviations exponents. Define

$$\mathcal{D}(\mathbf{x}) = \{\mathbf{y} \in \hat{\mathcal{A}}^l : (\mathbf{x}, \mathbf{y}) \in \mathcal{D}\}, \tag{9}$$

and

$$\mathcal{T}(\mathbf{x}) = \{\mathbf{y} \in \hat{\mathcal{A}}^l : (\mathbf{x}, \mathbf{y}) \in \mathcal{T}\}. \tag{10}$$

Averaging over source and codebook we have

$$\begin{aligned}
&\mathbb{P}[d^l(S^l, \mathcal{C}^l) > D] && (11)\\
&= \mathbb{E}[(1 - P_{\hat{S}^l}^{(\bar{D})}(\mathcal{D}(S^l)))^{2^{l(R(D)+\gamma)}}] && (12)\\
&\le \mathbb{E}[(1 - P_{\hat{S}^l}^{(\bar{D})}(\mathcal{D}(S^l) \cap \mathcal{T}(S^l)))^{2^{l(R(D)+\gamma)}}] && (13)\\
&\le \mathbb{E}[(1 - P_{\hat{S}^l|S^l}^{(\bar{D})}(\mathcal{D}(S^l) \cap \mathcal{T}(S^l)|S^l) \\
&\qquad 2^{-l(R(D)+3\gamma/4)})^{2^{l(R(D)+\gamma)}}] && (14)\\
&\le \mathbb{E}[1 - P_{\hat{S}^l|S^l}^{(\bar{D})}(\mathcal{D}(S^l) \cap \mathcal{T}(S^l)|S^l) + 2^{-l\gamma/4}] && (15)\\
&= 1 - P_{S^l\hat{S}^l}^{(\bar{D})}(\mathcal{D} \cap \mathcal{T}) + 2^{-l\gamma/4} && (16)\\
&= P_{S^l\hat{S}^l}^{(\bar{D})}(\mathcal{D}^c \cup \mathcal{T}^c) + 2^{-l\gamma/4} && (17)\\
&\le P_{S^l\hat{S}^l}^{(\bar{D})}(\mathcal{D}^c) + P_{S^l\hat{S}^l}^{(\bar{D})}(\mathcal{T}^c) + 2^{-l\gamma/4} && (18)\\
&\le 3(2^{-lf^*(\gamma)}), && (19)
\end{aligned}$$

where (14) follows from the definition in (10) and (15) follows from

$$\left(1-\frac{p\beta}{M}\right)^M \le 1-p+\frac{1}{\beta}, \tag{20}$$

for $0 \le p \le 1$, and $\beta > 0$, and

$$f^*(\gamma) = \min\{f_1(\gamma), f_2(\gamma), \gamma/4\}. \tag{21}$$

Using (19) and averaging over code and source,

$$\begin{aligned} \mathbb{E}[d^l(S^l, \mathcal{C}^l)] &\le \mathbb{P}[d^l(S^l, \mathcal{C}^l) \le D]D \\ &\quad + \mathbb{P}[d^l(S^l, \mathcal{C}^l) > D]D_{max} \quad (22)\\ &\le 3(D_{max}-D)2^{-l(f^*(\gamma))} + D, \quad (23) \end{aligned}$$

We finally choose the blocklength l :

$$l = \left\lceil \frac{1}{f^*(\gamma)} \log_2\left(\frac{3D_{max}}{\epsilon}\right)\right\rceil. \tag{24}$$

From (24) and (23) it follows that $\mathbb{E}[d^l(S^l, \mathcal{C}^l)] < D + \epsilon$, which implies that there exists at least one codebook $\mathcal{C}_0$ for which (2) is satisfied, with the blocklength given in (24). □

Lemma 1 may be compared to Theorem 1 in [8], where it is shown that there exists a sequence of codes with rate $R(D)$ and blocklength n that achieve average distortion D_n such that

$$D_n - D = -D'(R(D))\frac{\log n}{2n} + o\left(\frac{\log n}{n}\right), \tag{25}$$

where $D'(R)$ is the derivative of the distortion-rate function at R. In other words if we allow no slack in the rate, to achieve performance within ϵ of the optimal distortion the scheme in [8] requires blocklength growing at least as $(1/\epsilon)\log(1/\epsilon)$, whereas Lemma 1 shows that if we allow a slack of γ in the rate, then $n \propto \log(1/\epsilon)$ is enough to achieve a distortion $D+\epsilon$. Note that (19) is reminiscent of Marton's reliability function result [9]: There exist a sequence of block codes $\mathcal{C}(l)$ with blocklength l and rate $R(D)+\gamma$, such that $d^l(S^l, \mathcal{C}(l))$ satisfies

$$\lim_{l\to\infty} \frac{1}{l} \log \frac{1}{P[d^l(S^l, \mathcal{C}(l)) > D]} = e(\gamma), \tag{26}$$

where

$$e(\gamma) = \min_{\substack{P_{\bar{S}} \\ R(P_{\bar{S}}, D) \ge R(D)+\gamma}} D(P_{\bar{S}}||P_S), \tag{27}$$

where $R(P_{\bar{S}}, D)$ denotes the rate-distortion function of the source $P_{\bar{S}}$. However Marton's result is not sufficient to prove Lemma 1 for which we need a nonasymptotic bound on $\mathbb{P}[d^l(S^l, \mathcal{C}(l)) > D]$.

Proof of Theorem 1. Choose l from (3), construct a code for compressing a source with blocklength $n = kl$, using $\mathcal{C}_0$ as described in Section II. Let $D'(n)$ be the random variable that denotes the normalized distortion between the source realization and its reproduction obtained with this scheme, and let D_i be the normalized distortion achieved in the i^{th} word, $i = 1, 2, \cdots, k$. Then

$$\lim_{n\to\infty} \mathbb{P}[D'(n) > D+\epsilon] = \lim_{k\to\infty} \mathbb{P}\left[\frac{1}{k}\sum_{i=1}^k D_i > D+\epsilon\right] = 0. \tag{28}$$

where (28) follows from the law of large numbers, because D_i are i.i.d. random variables with $E(D_i) < D+\epsilon$. From Section II-A, the encoding time per symbol is proportional to $2^{l(R(D)+\gamma)} = C_1(\gamma)\epsilon^{-C_2(\gamma)}$ where

$$C_1(\gamma) = 2^{g_2(\gamma)(R(D)+\gamma)}, \tag{29}$$

and

$$C_2(\gamma) = g_1(\gamma)(R(D)+\gamma). \tag{30}$$

The decoding complexity per symbol is a constant as demonstrated in Section II-B. □

Theorem 1 establishes that for a given source P_S, $R = R(P_S, D) + \gamma$ and arbitrary $\epsilon > 0$, there exists a lossy compression scheme, with rate R that achieves distortion $D+\epsilon$ asymptotically almost surely, with encoding time per symbol $\propto C_1(\gamma)\epsilon^{-C_2(\gamma)}$. However if

$$R > \max_{P_S} R(P_S, D) \tag{31}$$

then there exists a lossy compression scheme, with rate R that achieves guaranteed distortion less than D for any source on that alphabet. This scheme is described below:

Marton [9] shows that if R satisfies (31), then there exists l_0 such that if $l \ge l_0$ there exists a block-code $\mathcal{C}(l)$ with rate R and blocklength l, such that for any sequence $s^l \in \mathcal{A}^l$, $d^l(s^l, \mathcal{C}(l)) < D$. Thus for any $n = kl$, $l \ge l_0$ we can obtain a lossy compressor by concatenating k copies of $\mathcal{C}(l)$ as in Section II. This compressor has guaranteed distortion less than D for any source realization and any blocklength $n = kl$ with encoding time per symbol proportional to 2^{lR}. In particular if $P_{\bar{S}}$ achieves the maximum in (31) then for any rate close to (but slightly larger) than $R(P_{\bar{S}}, D)$, there exists a linear-time lossy compression scheme that guarantees distortion less than D.

IV. Almost linear-complexity asymptotically optimal universal lossy compressor

We present a sequence of universal encoding/decoding schemes indexed by blocklength n. These schemes have linear decoding complexity, $O(ng(n))$ encoding complexity (for an arbitrary non-decreasing and unbounded function $g(n)$) and asymptotic rate equal to R. Further, when used to compress an ergodic source with distortion-rate function $D(R)$, the probability that the distortion between the source and its reproduction exceeds $D(R)$ vanishes with n. A constructive proof of the existence of such schemes is given below:

Theorem 2. *For a given R and an arbitrary non-decreasing and unbounded function $g(n)$, there exists a sequence of universal encoding/decoding schemes indexed by n with asymptotic rate equal to R, encoding and decoding complexities*

$O(ng(n))$ and $O(n)$ respectively, such that

$$\lim_{n\to\infty} \mathbb{P}[d^n(S^n, \hat{S}^n) > D(R) + \epsilon] = 0, \tag{32}$$

for any $\epsilon > 0$, where $D(\cdot)$ is the distortion-rate function for the ergodic source S^n.

Proof. We show this result assuming that $g(\cdot)$ is sub-exponential, it then follows a-fortiori for all other $g(\cdot)$.

Encoder: Let

$$l(n) = \left\lfloor \frac{\log_2 \log_2 g(n)}{(R+1)} \right\rfloor, \tag{33}$$

and let $\mathcal{C}^{l(n)}(R)$, denote the set of all codebooks with rate R, and blocklength $l(n)$ with symbols chosen from the reproduction alphabet $\hat{\mathcal{A}}$. Given a source realization s^n choose a codebook from $\mathcal{C}^{l(n)}(R)$ that achieves the minimum distortion when used to encode the sequence s^n as shown in Figure 1. Describe the selected codebook and the encoding with $nR + \log_2 |\mathcal{C}^{l(n)}(R)|$ bits. An upper bound on the number of codebooks is given as

$$|\mathcal{C}^{l(n)}(R)| \le |\hat{\mathcal{A}}|^{l(n)\times 2^{l(n)R}} \tag{34}$$

Using (34) and (33) and the fact that $g(n)$ is sub-exponential

$$\lim_{n\to\infty} R_n \le R + \log_2 |\hat{\mathcal{A}}| \lim_{n\to\infty} \frac{l(n)\times 2^{l(n)R}}{n} = R. \tag{35}$$

Encoding takes $n2^{l(n)R}|\mathcal{C}^{l(n)}(R)|$ operations. From (34) it can be shown that $n2^{l(n)R}|\mathcal{C}^{l(n)}(R)| < n2^{2^{l(n)(R+1)}}$ for all sufficiently large n. Thus from (33) the encoding complexity is given as $O(ng(n))$.

Decoder: Identify the codebook using $\log_2 |\mathcal{C}^{l(n)}(R)|$ bits and obtain the reproduction as described in Section II-B. Decoding takes n operations thus the decoding complexity is $O(n)$.

Optimality See [10]. □

A. An extension to Theorem 1

We now modify the lossy compressor in Section III, such that it achieves the rate-distortion point $(R(D), D)$ asymptotically, with $O(n)$ decoding and $O(ng(n))$ encoding complexity. In contrast to the universal scheme which is only of theoretical interest, this compressor is also amenable to practical implementation as suggested by the results in Section V.

Code Construction: Choose arbitrary $c > 0$. Let

$$l(n) = \left\lceil \frac{\log_2 g(n)}{R(D) + c} \right\rceil \tag{36}$$

and

$$\gamma(n) = f^{*-1}\left(\frac{\log l(n)}{l(n)}\right), \tag{37}$$

where $f^*(\cdot)$ is defined in (21) and

$$f^{*-1}(x) = \inf\{y \ge 0 : f^*(y) \ge x\}. \tag{38}$$

Choose a codebook $\mathcal{C}_0^{l(n)}$ with rate $R(D) + \gamma(n)$, and blocklength $l(n)$, that satisfies

$$\mathbb{P}[d^l(S^{l(n)}, \mathcal{C}_0^{l(n)}) \ge D] \le 3(2^{-l(n)f^*(\gamma(n))}), \tag{39}$$

such a codebook is guaranteed to exist from (19). The lossy compressor is obtained by concatenating $k = n/l(n)$ copies of $\mathcal{C}_0^{l(n)}$ as shown in Figure 1. It can be shown that $\lim_{n\to\infty} \gamma(n) = 0$ (see [10]), thus the asymptotic rate is equal to $R(D)$. From Section II-A encoding complexity is given as $O(n2^{(R(D)+\gamma(n))l(n)})$, choosing n large enough such that $\gamma(n) < c$ the encoding complexity is upper bounded by $O(ng(n))$. The decoding complexity depends neither on $l(n)$ nor on $\gamma(n)$ and is equal to $O(n)$.

Optimality: Let D_i denote the distortion in the i^{th} block. It can be shown that (see [10])

$$\lim_{n\to\infty} \mathbb{P}[D_i > D] \le \lim_{n\to\infty} 3(2^{-l(n)f^*(\gamma(n))}) = 0. \tag{40}$$

Therefore for any $\delta > 0$ there exists $n(\delta)$ such that for $n > n(\delta)$, $\mathbb{P}[D_i > D] < \delta$. Defining the binary random variables $B_i = 1\{D_i > D\}$, and $\epsilon' = \epsilon/D_{max}$ we have

$$P\left[D_n > D + \epsilon\right] \le P\left[\frac{1}{k}\sum_{i=1}^{k} B_i > \epsilon'\right]. \tag{41}$$

Let $0 < \delta < \epsilon'$. For $n > n(\delta)$, from (41) and the Chernoff bound we have

$$\mathbb{P}[D_n > D + \epsilon] \le P\left[\frac{1}{k}\sum_{i=1}^{k} B_i > \epsilon'\right] \le 2^{-\frac{n}{l(n)}d(\epsilon'||\delta)}. \tag{42}$$

Therefore

$$\lim_{n\to\infty} \mathbb{P}[D_n > D + \epsilon] = 0. \tag{43}$$

V. Experiments

The scheme in Section III, with fixed blocklength l bootstraps a short code with rate $R(D) + \gamma$ and average distortion $D+\epsilon$ to obtain a long code, such that when a source realization is compressed using this code, the probability that the resulting distortion exceeds $D + \epsilon$ vanishes asymptotically. Thus it can achieve the rate-distortion function in the strong excess-distortion sense. To further highlight this notion of guaranteed distortion we plot the percentile distortion as a function of rate. Where a p percentile distortion implies that in p percent of empirical runs the distortion is less than the plotted value. Empirically for blocklength of the order of thousands, this scheme outperforms existing schemes under the percentile distortion criterion. In each of these experiments we randomly generate and fix a short random code with blocklength $l = \lceil 22/(R(D) + \gamma)\rceil$. Thus, the encoding scheme has linear complexity in the blocklength n. Figure 2 shows the distortion-rate performance of our scheme compared with the message passing based encoder applied to a regular sparse graph code described in [4] (with $O(n^2)$ complexity) when used to compress a fair coin source with blocklength $n = 1000$. Each point is obtained by running 1000 trials of encoding randomly generated source vectors for each rate, and choosing the bit error rate level such that 95% of the trials result in smaller bit error rate. In Figure 3 we compare the performance of our scheme when used for compressing the Bernoulli(0.4) source with bit error rate-distortion criterion. It is noted in [11] that

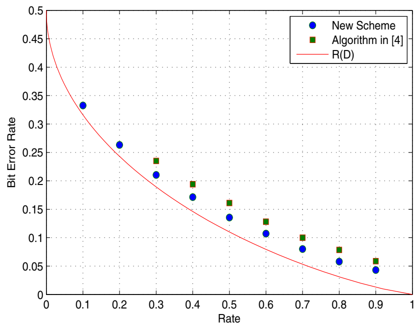

Fig. 2. Bit Error Rate (95% percentile criterion) for the fair coin source with blocklength=1000.

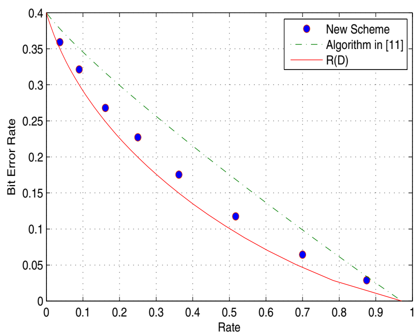

Fig. 3. Bit Error Rate (95% percentile criterion) for the biased coin (bias=0.4) source and blocklength=1000. Average Bit Error Rate shown for the scheme in [11].

for this problem, even to beat the straight line connecting the points $(0, h(0.4))$ and $(0.4, 0)$ on the $(D, R(D))$ curve is a challenge for any encoder, and no constructive schemes are known that perform near the optimal distortion-rate function. The compressor proposed in [11] (with $O(n)$ complexity) is slightly better than the straight line, and is one of the best implementable schemes for this problem in the literature. The asymptotic ($n \to \infty$) average distortion obtained with the scheme in [11] is compared to our scheme with $n = 1000$ in Figure 4. It should be noted that while the curve for [11] represents the average bit error rate, the curve for our scheme shows the stricter 95% percentile distortion criterion. In Figure 4 we show results for compressing fair coin flips with the asymmetric distortion criterion:

$$d(0,1) = 2d(1,0) = 2. \tag{44}$$

This problem was also addressed in [6] using sparse graph codes. The new scheme also outperforms the scheme given in [6] (which has a time complexity of $O(n^2 \log^3 n)$).

In Figures 3 and 4 with biased coin and asymmetric distortion, respectively, we use a short random code with blocklength $l = \lceil 22/(R(D)+\gamma) \rceil$, which is obtained in the following fashion. Choose a desired operating point $(D, R(D))$ and approximate the optimal binary reproduction distribution $P^D_{\hat{S}}$ with a q-type ($\bar{P}^D_{\hat{S}}$) over $\{0, 1\}$. First construct a random

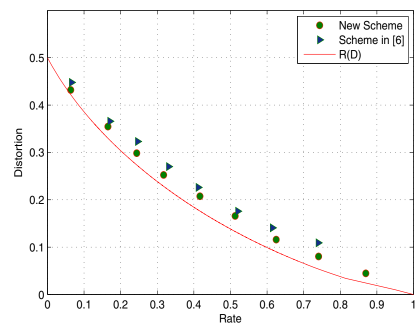

Fig. 4. Asymmetric distortion (44) (95% percentile criterion) for the fair coin source with blocklength=1000.

code over the ring $\{0, 1, \cdots, q-1\}$, using a $nR \times n$ matrix $\mathbf{G}$, such that each of its elements is chosen from the set $\{0, 1, \cdots, q-1\}$ uniformly at random. The codewords in the code are obtained by $\mathbf{c}_i = \mathbf{G}^T \mathbf{u}_i$, where $\mathbf{u}_i$ is a binary vector and addition is modulo q. now apply a mapping $Q : \{0, 1, \cdots, q-1\} \to \{0, 1\}$ to each $\mathbf{c}_i$ such that Q maps the uniform probability distribution over $\{0, 1, \cdots, q-1\}$ to $\bar{P}^D_{\hat{S}}$ over $\{0, 1\}$.

Increasing blocklength while maintaining the same short code results in a further reduction in the probability of excess distortion. Narrowing the gap with the distortion-rate curve requires increasing the blocklength of the short code.

These experimental results establish that the code construction in Section III is not only theoretically interesting but with small (computationally-feasible) wordlength outperforms existing constructive schemes for blocklengths of the order 1000.

REFERENCES

[1] Y. Matsunaga and H. Yamamoto, "A coding theorem for lossy data compression by LDPC codes," *IEEE Transactions on Information Theory*, vol. 49, pp. 2225– 2229, Sep 2003.

[2] S. Miyake, "Lossy data compression over Z_q by LDPC code," *IEEE International Symposium on Information Theory, Seattle USA*, july 2006.

[3] E. Martinian and M. J. Wainwright, "Low-density codes achieve the rate-distortion bound," *Data Compression Conference, Utah*, 2006.

[4] E. Maneva and M. J. Wainwright, "Lossy source encoding via messagepassing and decimation over generalized codewords of LDGM codes," *IEEE International Symposium on Information Theory, Adelaide, Austrailia*, September 2005.

[5] S. Ciliberti, K. Mezard, and R. Zecchina, "Message passing algorithms for non-linear nodes and data compression," *arxiv*. [Online]. Available: http://arxiv.org/abs/cond-mat/0508723

[6] A. Gupta and S. Verdú, "Nonlinear sparse-graph codes for lossy compression," *submitted to: IEEE Transactions on Information Theory*.

[7] I. Kontoiannis, "An Implementable Lossy Version of the Lempel-Ziv Algorithm Part I: Optimality for Memoryless Sources," *IEEE Transactions on Information Theory*, vol. 45, pp. 2293–2305, Nov 1999.

[8] Z. Zhang, E. Yang, and V. Wei, "The redundancy of source coding with a fidelity criterion. 1. Known statistics," *IEEE Transactions on Information Theory*, vol. 43, no. 1, pp. 71–91, June 1997.

[9] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 197–199, March 1974.

[10] A. Gupta, S. Verdu, and T. Weissman, "Almost-Linear lossy compression for memoryless sources," *Draft*, 2008.

[11] J. Rissanen and I. Tabus, "Rate-distortion without random codebooks," *Workshop on Information Theory and Applications (ITA),UCSD*, 2006.