

Achieving the Rate-Distortion Bound with Low-Density Generator Matrix Codes

Zhibin Sun, Mingkai Shao, Jun Chen, *Member, IEEE*, Kon Max Wong, *Fellow, IEEE*,
and Xiaolin Wu, *Senior Member, IEEE*

Abstract—It is shown that binary low-density generator matrix codes can achieve the rate-distortion bound of discrete memoryless sources with general distortion measure via multilevel quantization. A practical encoding scheme based on the survey-propagation algorithm is proposed. The effectiveness of the proposed scheme is verified through simulation.

Index Terms—Rate-distortion bound, linear code, low-density generator matrix, message-passing algorithm.

I. INTRODUCTION

ACHIEVING the rate-distortion bound of discrete memoryless sources has been one of the grand challenges in information theory. Though Shannon [1] has shown that the bound can be achieved asymptotically by using random block codes, the exponential encoding complexity makes such codes impractical. Coincidentally, another grand challenge in information theory is to design channel codes for discrete memoryless channels that have reasonable decoding complexity. To meet this challenge, research in channel coding has been almost exclusively focused on linear codes whose linear structure can be utilized to achieve efficient encoding and decoding. The focus on linear codes is further justified by the well-known result that linear codes can achieve the capacity of symmetric discrete memoryless channels under maximum likelihood decoding [2], [3]. The focused effort proves to be successful. Recent breakthroughs have shown that linear codes, e.g., Turbo codes and low-density parity check (LDPC) codes, used in conjunction with various message-passing decoding algorithms, can indeed approach the capacity of discrete memoryless channels.

Recall that in lossy source coding the main design problem is to reduce the exponential encoding complexity resulting from the search for a reconstruction sequence in an exponentially large sequence space. Similarly, the main design problem in channel coding is to reduce the exponential decoding

complexity resulting from the search for a codeword in an exponentially large space. The similar nature of the two design problems implies that the techniques developed for channel coding might be useful for lossy source coding as well. Indeed, lossy source codes like tree codes [4] and trellis codes [5] all have their origins in channel coding. Note that these two types of codes are known to be capable of achieving the rate-distortion bound. However, while these codes get the strengths from their counterparts in channel coding, they also inherit their counterparts' drawbacks. For example, saturating the rate-distortion bound with trellis codes requires taking the constraint length to infinity, which incurs exponential complexity.

In view of the recent success of LDPC codes and message-passing algorithms for channel coding, it is natural to ask whether similar approaches are also applicable to lossy source coding. Indeed, motivated by this line of thinking, it has been proposed to use low-density generator matrix (LDGM) codes (which can be viewed as the dual of LDPC codes) and various message-passing algorithms for lossy source coding. As a consequence, the following questions naturally arise.

- Q1: Can LDGM codes achieve the rate-distortion bound of discrete memoryless sources with general distortion measure?
- Q2: If the answer to Q1 is affirmative, can LDGM codes, used in conjunction with certain message-passing algorithms, approach the rate-distortion bound of discrete memoryless sources with general distortion measure?

For the binary erasure quantization problem, Q1 and Q2 have been successfully answered in [6]. For uniform binary sources with Hamming distortion measure, Q1 has been addressed in [7]; it is shown that LDGM codes¹ (under optimal encoding) can saturate the rate-distortion bound in this case. Moreover, for uniform binary sources with Hamming distortion measure, good empirical answers to Q2 have been given in [8]–[11], where LDGM codes are used in conjunction with variants of message-passing algorithms for efficient encoding. Beyond the aforementioned examples, however, the answers to Q1 and Q2 are still unknown.

The purpose of this paper is to settle Q1 in the affirmative for general discrete memoryless sources and distortion measure. Specifically, we show that binary LDGM codes can

Paper approved by Z. Xiong, the Editor for Distributed Coding and Processing of the IEEE Communications Society. Manuscript received April 17, 2009; revised September 14, 2009.

Z. Sun was with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. He is now with Hydro One Networks Inc., Toronto, ON, Canada (e-mail: zhibin.sun@hydroone.com).

M. Shao, J. Chen, K. M. Wong, and X. Wu are with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada (e-mail: shaom2@mcmaster.ca, {junchen, wong, xwu}@ece.mcmaster.ca).

The material in this paper was presented in part at the IEEE Information Theory Workshop, Lake Tahoe, CA, USA, Sep. 2007, and the 24th Queen's Biennial Symposium on Communications, Kingston, ON, Canada, June 2008.

Digital Object Identifier 10.1109/TCOMM.2010.06.090203

¹More precisely, a hybrid LDPC-LDGM based construction is used in [7]. Nevertheless, the proof technique developed in [7] works for LDGM codes as well.

achieve the rate-distortion bound via multilevel quantization. Moreover, our simulation results indicate that binary LDGM codes, used in conjunction with the survey-propagation algorithm [12]–[14], provide a promising answer to Q2.

Note that linear lossy source codes should not be confused with linear lossy source encoding. Indeed, although we shall use linear codebooks, the encoding operation (i.e., the mapping from the source sequence to the codeword index) is intrinsically nonlinear. In fact, it is known that linear encoding generally cannot achieve the rate-distortion bound [15], albeit it suffices for lossless data compression [16]. It should also be emphasized that research on practical lossy source coding is by no means restricted to the aforementioned ones. For example, there has been a significant amount of work devoted to designing lossy source coding schemes based on the existing lossless data compression algorithms [17]–[19]. Moreover, several promising new directions [20]–[23] have emerged during the writing of this paper. In particular, it is shown in [21] that nonlinear sparse-graph codes can achieve the rate-distortion bound of discrete memoryless sources with general distortion measure.

The remainder of this paper is organized as follows. In Section II, we discuss the fundamental limit of linear codes in channel coding and lossy source coding. In Section III, we show that binary LDGM codes can saturate the rate-distortion bound via multilevel quantization. A practical encoding scheme based on the survey-propagation algorithm is proposed in Section IV. The effectiveness of the proposed scheme is verified through simulation. Finally, we close with some concluding remarks in Section V.

II. LINEAR CODES FOR CHANNEL CODING AND LOSSY SOURCE CODING

Consider the lossy source coding problem for an i.i.d. process $\{X_i\}_{i=1}^{\infty}$ with marginal probability distribution P_X on \mathcal{X} . Let $d(\cdot, \cdot) : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow [0, d_{\max}]$ be a distortion measure, where $\hat{\mathcal{X}}$ is the reconstruction alphabet. Throughout this paper, \mathcal{X} and $\hat{\mathcal{X}}$ are assumed to be finite. The rate-distortion function $R(D)$ for source X with distortion measure $d(\cdot, \cdot)$ is given by [24]

$$R(D) = \min_{P_{\hat{X}|X}} I(X; \hat{X}), \quad (1)$$

where the minimization is over all test channels $P_{\hat{X}|X} : \mathcal{X} \rightarrow \hat{\mathcal{X}}$ subject to the constraint $\mathbb{E}[d(X, \hat{X})] \leq D$. The rate-distortion bound is known to be achievable with random block codes. However, the encoding complexity of such codes grows exponentially with the block length since there is no structure to exploit. As a consequence, it is of considerable interest to see whether structured codes, such as linear codes, suffice to achieve the rate-distortion bound.

First we shall give a definition of linear codes. A binary linear code \mathcal{C}_n of length n and rate $\frac{k}{n}$ is a k -dimensional linear subspace of the vector space \mathbb{F}_2^n , where \mathbb{F}_2 is the binary field with elements 0 and 1. Moreover, we can write $\mathcal{C}_n = \{c^n : c^n = u^k G, u^k \in \mathbb{F}_2^k\}$, where G , referred to as the generator matrix of \mathcal{C}_n , is a $k \times n$ matrix with rows being the basis codewords of \mathcal{C}_n .

To see whether linear codes are capable of achieving the rate-distortion bound, it is instructive to examine a similar problem in channel coding. Let $P_{V|U} : \mathcal{U} \rightarrow \mathcal{V}$ be a discrete memoryless channel. For illustrative purpose, we shall temporarily assume $\mathcal{U} = \{0, 1\}$. It is well known that the capacity (i.e., the maximum reliable communication rate) of channel $P_{V|U}$ is given by

$$C = \max_{P_U} I(U; V), \quad (2)$$

where the maximization is over all probability distributions on \mathcal{U} . However, the channel capacity is not always directly achievable using binary linear codes. Specifically, it has been shown [3] that the maximum reliable communication rate for binary linear codes over channel $P_{V|U}$ is given by

$$C_L \triangleq I(U; V)|_{P_U=P^*}, \quad (3)$$

where P^* is the uniform distribution on \mathcal{U} . Since the role of reconstruction variable \hat{X} in lossy source coding is similar to that of channel input U in channel coding, while the role of generic source variable X is similar to that of channel output V , it is natural to restrict \hat{X} to be uniformly distributed if linear codes are used for lossy source coding. Specifically, for $\hat{\mathcal{X}} = \{0, 1\}$, one might conjecture that the minimum achievable rate in lossy source coding using binary linear codes is given by

$$R_L(D) \triangleq \min_{P_{\hat{X}|X}} I(X; \hat{X}), \quad (4)$$

where the minimization is over all test channels $P_{\hat{X}|X} : \mathcal{X} \rightarrow \hat{\mathcal{X}}$ subject to the constraints $\mathbb{E}[d(X, \hat{X})] \leq D$ and $P_{\hat{X}}(0) = P_{\hat{X}}(1) = \frac{1}{2}$; in particular, if there is no test channel $P_{\hat{X}|X} : \mathcal{X} \rightarrow \hat{\mathcal{X}}$ satisfying all the constraints, then $R_L(D)$ is set to be equal to infinity, which essentially asserts that no binary linear code can meet the distortion constraint D . This conjecture turns out to be false as shown by the following example. Let $\mathcal{X} = \hat{\mathcal{X}} = \{0, 1\}$, $P_X(0) = p$ with $p \in (0, \frac{1}{2})$, and $d(\cdot, \cdot)$ be the Hamming distortion measure. It is easy to verify that $R_L(D) = \infty$ when $D < \frac{1}{2}(1 - 2p)$. However, since the linear span of every binary block code is a binary linear code, it is always possible to find a binary linear code to meet the distortion constraint D (even when $D < \frac{1}{2}(1 - 2p)$). Therefore, this example indicates that in lossy source coding one might not need to restrict the output of the test channel to be uniformly distributed when linear codes are used.

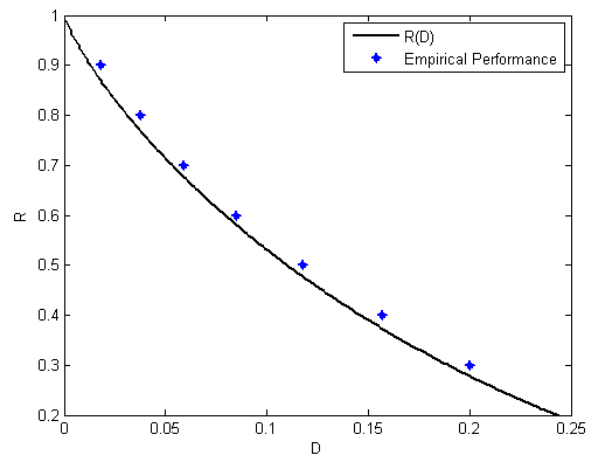
To better understand the subtle difference between channel coding and lossy source coding in terms of the fundamental limit of linear codes, we shall have a close look at the reasoning behind (3). First we need a few definitions. For any sequence $a^n = (a_1, a_2, \dots, a_n)$ with $a_i \in \mathcal{A}$, $i = 1, 2, \dots, n$, the empirical distribution P_{a^n} of a^n is defined as $(P_{a^n}(a))_{a \in \mathcal{A}}$ with $P_{a^n}(a) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(a_i = a)$, where $\mathbb{I}(\cdot)$ is the indicator function. We define the empirical distribution of a block code as the average of the empirical distributions of all its codewords. In particular, for a binary linear code \mathcal{C}_n of length n , the empirical distribution $P_{\mathcal{C}_n}$ of a codeword $c^n \in \mathcal{C}_n$ is given by $\frac{1}{n}(n - w(c^n), w(c^n))$, where $w(\cdot)$ is the weight function that counts the number of 1's in a sequence; moreover, the empirical distribution of \mathcal{C}_n is given by $\frac{1}{|\mathcal{C}_n|} \sum_{c^n \in \mathcal{C}_n} P_{c^n}$. It is easy to verify that the empirical distribution of an arbitrary

binary linear code is the uniform distribution on $\{0, 1\}$ unless the code contains coordinates in which every codeword is zero (such coordinates can be deleted as they carry no information). Now one can readily prove (3) by invoking Fano's inequality, the data processing inequality, and a convexity argument (cf. [25, Problem 5, pp. 112-113]). However, in contrast to the fact that each codeword in a channel code is transmitted with the same probability, the codewords in a lossy source code may not be used with the same frequency². As a consequence, the empirical distribution of a code does not play much of a role in lossy source coding (see [26] for a more in-depth discussion); indeed, this is the reason why (4) should not be viewed as the fundamental limit of binary linear lossy source codes. On the other hand, the rate-distortion function (1) does impose certain constraints on the empirical distributions of individual codewords of a good lossy source code. Let (R, D) be an arbitrary point on the rate-distortion curve. Let $P_{\hat{X}^*|X}$ be the optimal test channel associated with this rate-distortion point and $P_{\hat{X}^*}$ be the output distribution induced by P_X and $P_{\hat{X}^*|X}$. It can be shown using the result in [27] that a good lossy source code, in the sense that its rate-distortion performance is very close to the given (R, D) point, must contain an exponentially non-negligible fraction of codewords whose empirical distributions are close to $P_{\hat{X}^*}$. Note that $P_{\hat{X}^*}$ is in general not a uniform distribution except for certain special cases such as the uniform binary source with Hamming distortion measure. However, it is well known that the codeword empirical distributions of many commonly used binary random linear code ensembles concentrate (on the exponential scale) around the uniform distribution (see, e.g., [28]). Actually it can be shown under mild conditions that a binary linear code must have an exponentially non-negligible fraction of codewords with approximately the uniform empirical distribution. Although this does not exclude the possibility that the code also contains an exponentially non-negligible fraction of codewords with empirical distributions close to a non-uniform $P_{\hat{X}^*}$, it does indicate that such linear codes are difficult, if not impossible, to find³. As a consequence, if $P_{\hat{X}^*}$ associated with the given (R, D) on the rate-distortion curve is not a uniform distribution, then it is very hard to use linear codes to directly achieve this rate-distortion point due to the mismatch between the output distribution of the optimal test channel and the dominant codeword empirical distribution of linear codes⁴. To illustrate this point, we shall use the binary lossy source coding problem with Hamming distortion measure as an example. It can be seen from Fig. 1 that the performance of the scheme proposed in [9] (which is based on linear codes) degrades significantly when applied to a non-uniform source although it works well for the uniform

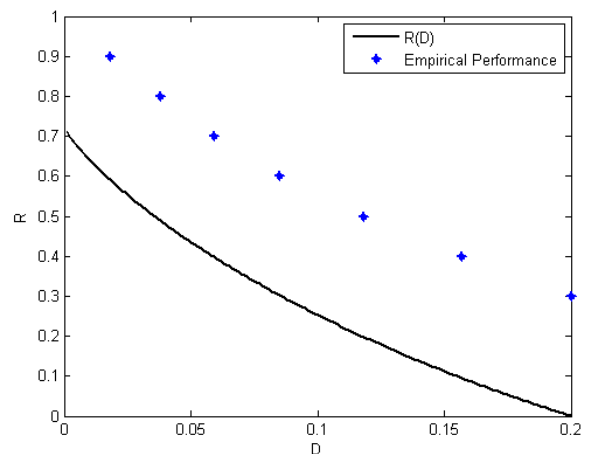
²The frequency that a codeword gets used depends on the encoding function and the source distribution.

³Since it is easy to generate linear codes for which the empirical distributions of most of the codewords concentrate around the uniform distribution, one can construct linear codes which contain an exponentially non-negligible fraction of codewords whose empirical distributions are close to a certain non-uniform distribution by adding coordinates in which every codeword is zero. However, the resulting linear codes are not very useful due to the inefficient use of the available freedom.

⁴A related discussion of mismatched codebooks in lossy source coding can be found in [29].



(a) $P_X(0) = 0.5$



(b) $P_X(0) = 0.2$

Fig. 1. Performance degradation for non-uniform sources.

source. This is because when the source is not uniform, the corresponding $P_{\hat{X}^*}$ is also not uniform. Therefore, the performance degradation is not solely due to the limitation of the encoding algorithm; it is also a consequence of the intrinsic property of linear codes.

III. MULTILEVEL QUANTIZATION

We shall show that (4) is actually an upperbound on the minimum achievable rate of binary linear lossy source codes. However, in order to prove the sufficiency of binary linear codes for lossy source coding, one has to deal with the following scenarios:

- 1) The output distribution $P_{\hat{X}^*}$ induced by P_X and the optimal test channel $P_{\hat{X}^*|X}$ is non-uniform;
- 2) The reconstruction alphabet is non-binary.

It turns out that both scenarios can be handled by leveraging a method called multilevel quantization. The multilevel quantization method has a natural counterpart in channel coding called multilevel coding, which was first proposed by Gallager [30]. The idea can be illustrated by a simple example. Suppose the optimal input distribution P_U for a binary-input memoryless channel $P_{V|U}$ of capacity C is given

by $(P_U(0), P_U(1)) = (\frac{1}{4}, \frac{3}{4})$. Now construct a random variable \tilde{U} that is uniformly distributed over \mathbb{F}_2^2 . Define a deterministic mapping $U = f(\tilde{U})$ such that $U = 0$ if $\tilde{U} = (0, 0)$ and $U = 1$ otherwise. Clearly, the induced distribution of U is exactly the capacity-achieving input distribution, and we have $I(\tilde{U}; V) = I(U; V) = C$. Here the essential idea is that, by incorporating a deterministic mapping, one can convert a channel with non-uniform capacity-achieving input distribution to a channel of the same capacity whose optimal input distribution is uniform. In general, we can approximate an arbitrary input distribution by a uniform distribution over sufficiently large alphabet.

Now we proceed to adopt this idea in lossy source coding. Specifically, we shall approximate the output distribution $P_{\hat{X}^*}$ by a uniform distribution over \mathbb{F}_2^m through a deterministic mapping. The rationale underlying such an approximation is the following. Loosely speaking, it can be assumed that the value of a binary linear code at any m coordinates is approximately uniformly distributed over \mathbb{F}_2^m (which will be justified later). Therefore, by mapping every m coordinates into a reconstruction symbol through a certain function, one can convert a binary linear code into a lossy source code whose dominant codeword empirical distribution matches $P_{\hat{X}^*}$. It is clear that in order to do so, the length of the binary linear code should be m times that of the length of the source sequence.

In the rest of this section we shall provide a rigorous treatment of this multilevel quantization method. For each positive integer m , we define

$$R_{L,m}(D) = \min_{P_{\hat{X}|X}} I(X; \tilde{X}),$$

where the minimization is over all $P_{\hat{X}|X} : \mathcal{X} \rightarrow \mathbb{F}_2^m$ such that the output distribution $P_{\hat{X}}$ induced by P_X and $P_{\hat{X}|X}$ is uniform over \mathbb{F}_2^m , and there exists a function $f : \mathbb{F}_2^m \rightarrow \hat{\mathcal{X}}$ satisfying the distortion constraint $\mathbb{E}[d(X, f(\tilde{X}))] \leq D$. We let $R_{L,m}(D) = \infty$ if there does not exist any $P_{\hat{X}|X}$ satisfying the constraints. Equivalently, we can define

$$R_{L,m}(D) = \min_{P_{\hat{X}|X}} I(X; \hat{X}),$$

where the minimization is over all $P_{\hat{X}|X} : \mathcal{X} \rightarrow \hat{\mathcal{X}}$ such that the output distribution $P_{\hat{X}}$ induced by P_X and $P_{\hat{X}|X}$ is in \mathcal{P}_m , and the distortion constraint $\mathbb{E}[d(X, \hat{X})] \leq D$ is satisfied. Here \mathcal{P}_m is the set of probability distributions $P_{\hat{X}}$ over $\hat{\mathcal{X}}$ satisfying $2^m P_{\hat{X}}(\hat{x}) \in \mathbb{Z}$ for all $\hat{x} \in \hat{\mathcal{X}}$.

Theorem 1: $\lim_{m \rightarrow \infty} R_{L,m}(D) = R(D)$ for $D > \sum_{x \in \mathcal{X}} P_X(x) \min_{\hat{x} \in \hat{\mathcal{X}}} d(x, \hat{x})$.

Proof: See Appendix A. ■

Now we proceed to show that $R_{L,m}(D)$ is achievable using binary LDGM codes. A binary LDGM code is a binary linear code associated with a sparse generator matrix (i.e., a generator matrix in which the number of 1's is significantly less than the number of 0's). We shall construct a binary random LDGM code ensemble by specifying a random generator matrix \mathbf{G} . Let \mathbf{G} be a $k \times mn$ matrix with entries selected independently from \mathbb{F}_2 using a Bernoulli distribution with parameter p_n (i.e., $\text{Ber}(p_n)$). We shall choose p_n such that $p_n \rightarrow 0$ and $np_n \rightarrow \infty$ as $n \rightarrow \infty$. Note that for such p_n , the realization of \mathbf{G} is a sparse generator matrix with high

probability as $n \rightarrow \infty$. Define $Y_i(u^k) = u^k \mathbf{g}_i$, $u^k \in \mathbb{F}_2^k$, where \mathbf{g}_i is the i -th column of \mathbf{G} , $i = 1, 2, \dots, mn$. Partition $\{1, 2, \dots, mn\}$ into n subsets $\mathcal{N}_1, \mathcal{N}_2, \dots, \mathcal{N}_n$, each of size m . We assume the elements in \mathcal{N}_i , $i = 1, 2, \dots, n$, are ordered. For each \mathcal{N}_i (say, $\mathcal{N}_i = (n_1, \dots, n_m)$), we define $\tilde{X}_i(u^k) = (Y_{n_1}(u^k), \dots, Y_{n_m}(u^k))$. For any deterministic mapping $f : \mathbb{F}_2^m \rightarrow \hat{\mathcal{X}}$, let $\mathcal{C}(\mathbf{G}, f) = \{\hat{X}^n(u^k) : u^k \in \mathbb{F}_2^k\}$, where $\hat{X}^n(u^k) = (f(\tilde{X}_1(u^k)), f(\tilde{X}_2(u^k)), \dots, f(\tilde{X}_n(u^k)))$. Define

$$R(n, k) = \frac{k}{n},$$

$$D(n, k, m, f) = \mathbb{E}[d_n(X^n, \mathcal{C}(\mathbf{G}, f))],$$

where

$$d_n(x^n, \mathcal{C}(\mathbf{G}, f)) = \min_{u^k \in \mathbb{F}_2^k} \frac{1}{n} \sum_{i=1}^n d(x_i, \hat{X}_i(u^k)), \quad x^n \in \mathcal{X}^n.$$

Note that $R(n, k)$ is not the rate of the linear code generated by \mathbf{G} ; instead, it should be interpreted as the rate⁵ of the lossy source code $\mathcal{C}(\mathbf{G}, f)$ induced by \mathbf{G} and $f(\cdot)$.

Now we are ready to state the main result of this section.

Theorem 2: Let $P_{\hat{X}|X} : \mathcal{X} \rightarrow \mathbb{F}_2^m$ be a transition probability distribution such that the output distribution $P_{\hat{X}}$ induced by P_X and $P_{\hat{X}|X}$ is uniform over \mathbb{F}_2^m . Let $f(\cdot)$ be a deterministic mapping from \mathbb{F}_2^m to $\hat{\mathcal{X}}$. For any $\epsilon > 0$, we have

$$R(n, k) \leq I(X; \tilde{X}) + \epsilon,$$

$$D(n, k, m, f) \leq \mathbb{E}[d(X, f(\tilde{X}))] + \epsilon$$

for some sufficiently large n and k .

Remark: In view of the definition of $R_{L,m}(D)$ and Theorem 1, one can readily show that binary LDGM codes can achieve the rate-distortion bound via multilevel quantization by optimizing $(P_{\hat{X}|X}, f(\cdot))$ and letting $m \rightarrow \infty$.

Proof: See Appendix B. ■

It has been shown recently [31] that for uniform binary sources with Hamming distortion measure, a necessary condition for achieving the rate-distortion bound using LDGM codes is that the generator matrix contains rows with the number of 1's growing unboundedly as $n \rightarrow \infty$ (see [32] for a weak version of this result). Note that in our construction, the condition $np_n \rightarrow \infty$ can be interpreted as the requirement that the number of 1's in each row of the generator matrix grows unboundedly with n ; therefore, our result indicates that the aforementioned necessary condition is essentially sufficient.

IV. A PRACTICAL ENCODING SCHEME

In this section, we shall propose a low-complexity encoding scheme based on the survey-propagation algorithm.

Let G be a $k \times mn$ low-density generator matrix. It is well known that one can represent G by a factor graph with k variable nodes and mn check nodes (see Fig. 2). Label these variable nodes and check nodes by $\{V_1, V_2, \dots, V_k\}$ and $\{C_1, C_2, \dots, C_{mn}\}$, respectively. Let there be an edge connecting V_i and C_j if and only if the (i, j) -entry of G is one. Let \mathcal{A}_k be the set of indices of check nodes connected to

⁵Strictly speaking, $R(n, k)$ is an upper bound on the rate of $\mathcal{C}(\mathbf{G}, f)$ since it is possible that $\hat{X}^n(u^k) = \hat{X}^n(v^k)$ for some $u^k \neq v^k$.

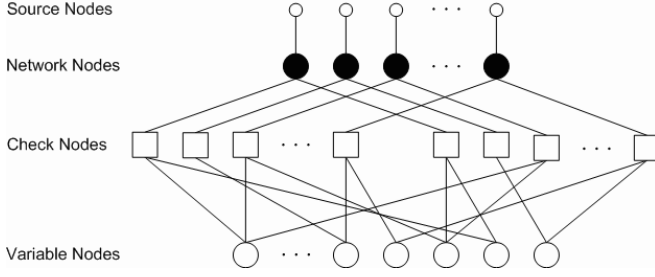


Fig. 2. Factor graph of an LDGM code.

V_i and \mathcal{B}_j be the set of indices of variable nodes connected to C_j . Each variable node V_i is associated with its corresponding information bit u_i . The value of C_j can be computed using the values of variable nodes in \mathcal{B}_j through modulo-2 addition.

We further introduce n source nodes and n network nodes, denoted by $\{S_1, S_2, \dots, S_n\}$ and $\{N_1, N_2, \dots, N_n\}$, respectively. Each source node S_l is connected with its corresponding network node N_l ; moreover, we let S_l and N_l be associated with source symbol x_l and its corresponding reconstruction symbol \hat{x}_l , respectively.

We partition $\{C_1, C_2, \dots, C_{mn}\}$ into n disjoint subsets, each containing m check nodes. Each subset is connected to a unique network node. Specifically, we denote the index of the network node connected with C_j as $\mu(j)$ and denote the set of indices of check nodes connected to N_l as \mathcal{N}_l . Label each element in \mathcal{N}_l with a unique number in $\{1, \dots, m\}$. Note that such a labeling induces a mapping $\nu(\cdot)$ from $\{1, 2, \dots, mn\}$ to $\{1, \dots, m\}$. The value of N_l (i.e., \hat{x}_l) is determined by the check nodes in \mathcal{N}_l through a mapping $f: \mathbb{F}_2^m \rightarrow \hat{\mathcal{X}}$.

Our goal is to find an efficient way to determine the values of the variable nodes so that the distortion between the resulting reconstruction sequence \hat{x}^n and the given source sequence x^n is made reasonably small.

To this end, we propose a message-passing algorithm tailored to the aforementioned factor graph. Specifically, messages are passed among different types of nodes with each message representing a probability distribution. They are specified as follows.

- From V_i to C_j , $j \in \mathcal{A}_i$: $M_{V_i \rightarrow C_j}^0$, $M_{V_i \rightarrow C_j}^1$, and $M_{V_i \rightarrow C_j}^*$.
- From C_j to V_i , $i \in \mathcal{B}_j$: $M_{C_j \rightarrow V_i}^0$, $M_{C_j \rightarrow V_i}^1$, and $M_{C_j \rightarrow V_i}^*$.
- From C_j to N_l , $l = \mu(j)$: $M_{C_j \rightarrow N_l}^0$, $M_{C_j \rightarrow N_l}^1$, and $M_{C_j \rightarrow N_l}^*$.
- From S_l to N_l : $M_{S_l \rightarrow N_l}^x$, $x \in \mathcal{X}$.
- From N_l to C_j , $j \in \mathcal{N}_l$: $M_{N_l \rightarrow C_j}^0$ and $M_{N_l \rightarrow C_j}^1$.

Remark: The introduction of the free state $*$ is one of the main features of the survey-propagation algorithm.

Here is a description of the message-passing routine (also see Fig. 3).

- 1) Each check node sends messages to the adjacent variable nodes (cf. (6)) and network node (cf. (7) and (8)). The initial message is set to be $(0.5, 0.5, 0)$.
- 2) Each variable node computes the new messages based on those from the adjacent check nodes, then sends the messages back to the adjacent check nodes (cf. (5)). Similarly, each network node computes the new messages based on those from the adjacent check nodes

and source node (cf. (9)), then sends the messages back to the adjacent check nodes (cf. (10)).

- 3) Each check node computes the new messages based on those from the adjacent variable nodes and network node. If the messages converge at all check nodes or the maximum number of iterations (e.g., 100) is reached, then go to 4), otherwise go to 1).
- 4) Each variable node (say, V_i) computes $(M_{V_i}^0, M_{V_i}^1, M_{V_i}^*)$, where

$$\begin{aligned} M_{V_i}^0 &= \prod_{j \in \mathcal{A}_i} (1 - M_{C_j \rightarrow V_i}^1) \\ &\quad - \prod_{j \in \mathcal{A}_i} (1 - M_{C_j \rightarrow V_i}^1 - M_{C_j \rightarrow V_i}^0), \\ M_{V_i}^1 &= \prod_{j \in \mathcal{A}_i} (1 - M_{C_j \rightarrow V_i}^0) \\ &\quad - \prod_{j \in \mathcal{A}_i} (1 - M_{C_j \rightarrow V_i}^1 - M_{C_j \rightarrow V_i}^0), \\ M_{V_i}^* &= \prod_{j \in \mathcal{A}_i} (1 - M_{C_j \rightarrow V_i}^1 - M_{C_j \rightarrow V_i}^0). \end{aligned}$$

Find those variable nodes whose bias value (i.e., $|M_{V_k}^0 - M_{V_k}^1|$) is greater than a certain threshold and fix those variables accordingly. If there is no variable node whose bias value is greater than the threshold, then we fix the value of the variable node with the maximum bias value.

- 5) Remove those fixed variable nodes from the factor graph. If all the variable nodes are fixed, then go to 6), otherwise go to 1).
- 6) Calculate the value of each check node through modulo-2 addition of the values of its adjacent variable nodes. Determine the value of each network node using the values of its adjacent check nodes through the deterministic mapping $f(\cdot)$.

Remark: All the messages need to be normalized.

Now we proceed to apply the proposed message-passing algorithm to various sources and distortion measures to verify its effectiveness.

A. Non-Uniform Binary Source With Hamming Distortion Measure

Let $\mathcal{X} = \hat{\mathcal{X}} = \{0, 1\}$ and $d(\cdot, \cdot)$ be the Hamming distortion measure, i.e., $d(x, \hat{x}) = 0$ if $x = \hat{x}$ and $d(x, \hat{x}) = 1$ otherwise. Consider a binary source X with $P_X(0) = 0.25$. Let $P_{\hat{X}^*}$ be the output distribution induced by P_X and the optimal test channel $P_{\hat{X}^*|X}$. Note that $P_{\hat{X}^*}$ is rate-dependent. Let $m = 4$ and \tilde{X} be uniformly distributed over \mathbb{F}_2^4 . A deterministic mapping $f: \mathbb{F}_2^4 \rightarrow \mathcal{X}$ is chosen so that $P_{f(\tilde{X})}$ is a good approximation of $P_{\hat{X}^*}$. Table I lists the optimal $P_{f(\tilde{X})}$ for different rates. Note that although the optimal $P_{f(\tilde{X})}$ for a given rate is in general unique, the deterministic mapping $f(\cdot)$ that induces this distribution is clearly not unique.

The message-passing equations are given in Fig. 3. In (9), we choose $\alpha_{x_l, x'} = 1$ if $x' = x_l$ and $\alpha_{x_l, x'} = \alpha$ otherwise. This choice is tailored to the Hamming distortion measure.

Variable node to check node	$M_{V_i \rightarrow C_j}^0 = \prod_{q \in \mathcal{A}_i \setminus \{j\}} (1 - M_{C_q \rightarrow V_i}^1) - \prod_{q \in \mathcal{A}_i \setminus \{j\}} (1 - M_{C_q \rightarrow V_i}^1 - M_{C_q \rightarrow V_i}^0)$ $M_{V_i \rightarrow C_j}^1 = \prod_{q \in \mathcal{A}_i \setminus \{j\}} (1 - M_{C_q \rightarrow V_i}^0) - \prod_{q \in \mathcal{A}_i \setminus \{j\}} (1 - M_{C_q \rightarrow V_i}^1 - M_{C_q \rightarrow V_i}^0)$ $M_{V_i \rightarrow C_j}^* = \prod_{q \in \mathcal{A}_i \setminus \{j\}} (1 - M_{C_q \rightarrow V_i}^1 - M_{C_q \rightarrow V_i}^0)$	(5)
Check node to variable node	$M_{C_j \rightarrow V_i}^0 = \frac{1}{2} (M_{N_{\mu(j)} \rightarrow C_j}^0 + M_{N_{\mu(j)} \rightarrow C_j}^1) \prod_{q \in \mathcal{B}_j \setminus \{i\}} (M_{V_q \rightarrow C_j}^0 + M_{V_q \rightarrow C_j}^1)$ $+ \frac{1}{2} (M_{N_{\mu(j)} \rightarrow C_j}^0 - M_{N_{\mu(j)} \rightarrow C_j}^1) \prod_{q \in \mathcal{B}_j \setminus \{i\}} (M_{V_q \rightarrow C_j}^0 - M_{V_q \rightarrow C_j}^1)$ $M_{C_j \rightarrow V_i}^1 = \frac{1}{2} (M_{N_{\mu(j)} \rightarrow C_j}^0 + M_{N_{\mu(j)} \rightarrow C_j}^1) \prod_{q \in \mathcal{B}_j \setminus \{i\}} (M_{V_q \rightarrow C_j}^0 + M_{V_q \rightarrow C_j}^1)$ $- \frac{1}{2} (M_{N_{\mu(j)} \rightarrow C_j}^0 - M_{N_{\mu(j)} \rightarrow C_j}^1) \prod_{q \in \mathcal{B}_j \setminus \{i\}} (M_{V_q \rightarrow C_j}^0 - M_{V_q \rightarrow C_j}^1)$ $M_{C_j \rightarrow V_i}^* = 1 - M_{C_j \rightarrow V_i}^0 - M_{C_j \rightarrow V_i}^1$	(6)
Check node to network node	$M_{C_j \rightarrow N_l}^0 = \frac{1}{2} \prod_{q \in \mathcal{B}_j} (M_{V_q \rightarrow C_j}^0 + M_{V_q \rightarrow C_j}^1) + \frac{1}{2} \prod_{q \in \mathcal{B}_j} (M_{V_q \rightarrow C_j}^0 - M_{V_q \rightarrow C_j}^1)$ $M_{C_j \rightarrow N_l}^1 = \frac{1}{2} \prod_{q \in \mathcal{B}_j} (M_{V_q \rightarrow C_j}^0 + M_{V_q \rightarrow C_j}^1) - \frac{1}{2} \prod_{q \in \mathcal{B}_j} (M_{V_q \rightarrow C_j}^0 - M_{V_q \rightarrow C_j}^1)$ $M_{C_j \rightarrow N_l}^* = 1 - M_{C_j \rightarrow N_l}^0 - M_{C_j \rightarrow N_l}^1$	(7)
If all $V_q \in \mathcal{B}_j$ are fixed, then	$M_{C_j \rightarrow N_l}^0 = \frac{(1 - c_j) \exp(\delta) + c_j \exp(-\delta)}{\exp(\delta) + \exp(-\delta) + W_{sou}}$ $M_{C_j \rightarrow N_l}^1 = \frac{c_j \exp(\delta) + (1 - c_j) \exp(-\delta)}{\exp(\delta) + \exp(-\delta) + W_{sou}}$ $M_{C_j \rightarrow N_l}^ = \frac{W_{sou}}{\exp(\delta) + \exp(-\delta) + W_{sou}}$	(8)
Source node to network node	$M_{S_l \rightarrow N_l}^x = \frac{\alpha_{x_l, x_l} \exp(\gamma)}{\alpha_{x_l, x_l} \exp(\gamma) + \sum_{x' \in \mathcal{X}: x' \neq x_l} \alpha_{x_l, x'} \exp(-\gamma)} \quad x = x_l$ $M_{S_l \rightarrow N_l}^x = \frac{\alpha_{x_l, x} \exp(-\gamma)}{\alpha_{x_l, x_l} \exp(\gamma) + \sum_{x' \in \mathcal{X}: x' \neq x_l} \alpha_{x_l, x'} \exp(-\gamma)} \quad x \neq x_l$	(9)
Network node to check node	$M_{N_l \rightarrow C_j}^0 = \sum_{(k_1, \dots, k_m): k_{\nu(j)}=0} (M_{S_l \rightarrow N_l}^r \prod_{q \in \mathcal{N}_l \setminus \{j\}} M_{C_q \rightarrow N_l}^{k_{\nu(q)}}) \quad r = f(k_1, \dots, k_m)$ $M_{N_l \rightarrow C_j}^1 = \sum_{(k_1, \dots, k_m): k_{\nu(j)}=1} (M_{S_l \rightarrow N_l}^r \prod_{q \in \mathcal{N}_l \setminus \{j\}} M_{C_q \rightarrow N_l}^{k_{\nu(q)}}) \quad r = f(k_1, \dots, k_m)$	(10)

Fig. 3. Message-passing equations. Here c_j is the modulo-2 addition of the values of the variable nodes in \mathcal{B}_j while δ , W_{sou} , $\alpha_{x_l, x'}$, and γ are the parameters that can be adjusted.

B. Uniform Ternary Source With Hamming Distortion Measure

Let $\mathcal{X} = \hat{\mathcal{X}} = \{0, 1, 2\}$ and $d(\cdot, \cdot)$ be the Hamming distortion measure. Consider a ternary source X with $P_X(0) = P_X(1) = P_X(2) = \frac{1}{3}$. Let $m = 4$. It is easy to verify that we have $P_{\hat{X}^*}(0) = P_{\hat{X}^*}(1) = P_{\hat{X}^*}(2) = \frac{1}{3}$ for all rates. Correspondingly, we choose $f: \mathbb{F}_2^4 \rightarrow \mathcal{X}$ such that $(P_{f(\hat{X})}(0), P_{f(\hat{X})}(1), P_{f(\hat{X})}(2)) = (\frac{5}{16}, \frac{5}{16}, \frac{6}{16})$ for all rates.

C. Uniform Ternary Source With ℓ_1 Distortion Measure

Let $\mathcal{X} = \hat{\mathcal{X}} = \{0, 1, 2\}$ and $d(\cdot, \cdot)$ be the ℓ_1 distortion measure, i.e., $d(x, \hat{x}) = |x - \hat{x}|$. Consider a ternary source X with $P_X(0) = P_X(1) = P_X(2) = \frac{1}{3}$. Let $m = 4$. Table II lists the optimal $P_{f(\hat{X})}$ for different rates.

The message-passing equations are given in Fig. 3. In (9), we choose $\alpha_{x_l, x'} = 1$ if $x' = x_l$, $\alpha_{x_l, x'} = \alpha$ if $|x_l - x'| = 1$, and $\alpha_{x_l, x'} = \beta$ if $|x_l - x'| = 2$. This choice is tailored to the ℓ_1 distortion measure. Since smaller distortion is preferable, we always set α to be larger than β .

TABLE I
THE OPTIMAL $P_{f(\tilde{X})}$ FOR DIFFERENT RATES.

R	$(P_{f(\tilde{X})}(0), P_{f(\tilde{X})}(1))$
0.479 - 1.000	$(\frac{4}{16}, \frac{12}{16})$
0.228 - 0.478	$(\frac{3}{16}, \frac{13}{16})$
0.096 - 0.227	$(\frac{2}{16}, \frac{14}{16})$
0.000 - 0.095	$(\frac{1}{16}, \frac{15}{16})$

TABLE II
THE OPTIMAL $P_{f(\tilde{X})}$ FOR DIFFERENT RATES.

R	$(P_{f(\tilde{X})}(0), P_{f(\tilde{X})}(1), P_{f(\tilde{X})}(2))$
0.43 - 1.58	$(\frac{5}{16}, \frac{6}{16}, \frac{5}{16})$
0.22 - 0.43	$(\frac{4}{16}, \frac{8}{16}, \frac{4}{16})$
0.13 - 0.22	$(\frac{3}{16}, \frac{10}{16}, \frac{3}{16})$
0 - 0.13	$(\frac{2}{16}, \frac{12}{16}, \frac{2}{16})$

TABLE III
THE OPTIMAL $P_{f(\tilde{X})}$ FOR DIFFERENT RATES.

D. Simulation Results

Now we examine the performance of the proposed scheme for the aforementioned three cases. The algorithm is implemented using C. The degree distributions of the LDGM codes are obtained from [33], which are optimized for the AWGN channel.

The simulation results are shown in Fig. 4, Fig. 5, and Fig. 6, respectively. Source sequences of length 1000 and 10000 are tested. Since $m = 4$, the corresponding LDGM codes are of length 4000 and 40000, respectively. The damping method [11] is used in the message-passing algorithm if the messages do not converge after 30 iterations. The decimation threshold is set to be 0.9 and the maximum number of iterations is set to be 100. Each plotted point is obtained by averaging over 1000 source sequences. It can be seen that the resulting distortions are close to the theoretical lower bound in all these three cases.

Note that although the degrees of variable nodes need to grow unboundedly in n to achieve the rate-distortion bound, simulation results indicate that relatively small degrees suffice for practical purposes. Moreover, it is worth mentioning that for a fixed n , the performance of the proposed message-passing algorithm may not improve with increasing degrees, and may even degrade.

V. CONCLUSION

We have shown that binary LDGM codes can achieve the rate-distortion bound of discrete memoryless sources with general distortion measure via multilevel quantization. Moreover, a practical encoding scheme based on the survey-propagation algorithm has been proposed. The simulation results validate the effectiveness of the proposed scheme. It is of considerable interest to generalize the results in the present work to the

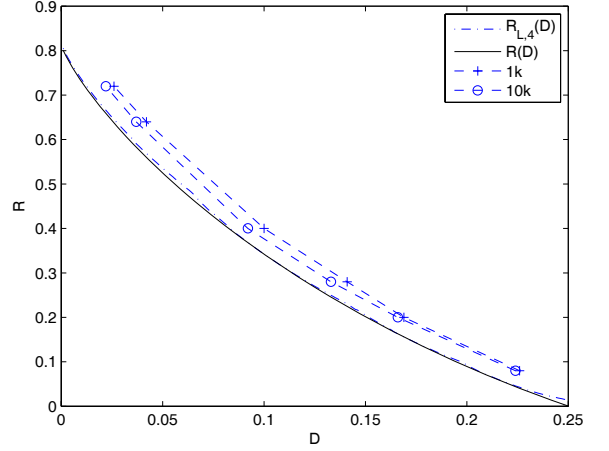


Fig. 4. Non-uniform binary source with Hamming distortion measure.

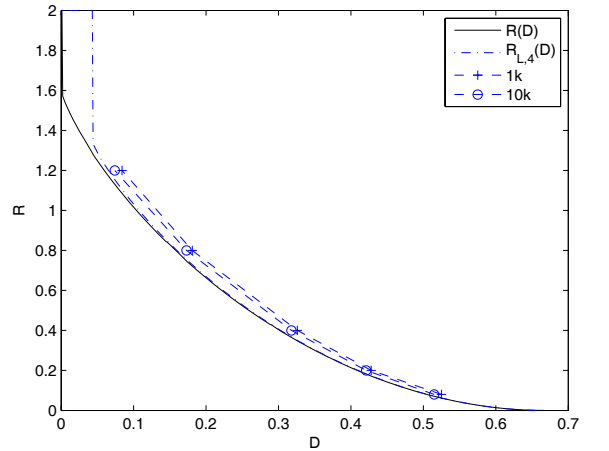


Fig. 5. Uniform ternary source with Hamming distortion measure.

multi-user scenarios such as distributed source coding and multiple description coding.

APPENDIX A
PROOF OF THEOREM 1

It is clear that $R_{L,m}(D)$ is lower-bounded by $R(D)$ for all m . Therefore, it suffices to show that $\limsup_{m \rightarrow \infty} R_{L,m}(D) \leq R(D)$.

Let $P_{\hat{X}^*|X}$ be the optimal test channel associated with the rate-distortion pair $(R(D), D)$ and $P_{\hat{X}^*}$ be the output distribution induced by P_X and $P_{\hat{X}^*|X}$. We shall construct a random variable \tilde{X} over \mathbb{F}_2^m such that $X - \hat{X}^* - \tilde{X}$ form a Markov chain. Let $k_{\hat{x}}$ be the greatest integer satisfying $\frac{k_{\hat{x}}}{2^m} \leq P_{\hat{X}^*}(\hat{x})$, $\hat{x} \in \hat{\mathcal{X}}$. Let $\mathcal{K}_{\hat{x}}$, $\hat{x} \in \hat{\mathcal{X}}$, be disjoint subsets of \mathbb{F}_2^m with $|\mathcal{K}_{\hat{x}}| = k_{\hat{x}}$. Let $\mathcal{K}^* = \mathbb{F}_2^m \setminus \cup_{\hat{x} \in \hat{\mathcal{X}}} \mathcal{K}_{\hat{x}}$. Now construct the transition probability distribution $P_{\tilde{X}|\hat{X}^*}$ as follows:

$$P_{\tilde{X}|\hat{X}^*}(\tilde{x}|\hat{x}) = \begin{cases} \frac{1}{2^m P_{\hat{X}^*}(\hat{x})}, & \tilde{x} \in \mathcal{K}_{\hat{x}} \\ \frac{1}{|\mathcal{K}^*|} \left(1 - \frac{k_{\hat{x}}}{2^m P_{\hat{X}^*}(\hat{x})}\right), & \tilde{x} \in \mathcal{K}^* \\ 0, & \text{otherwise} \end{cases}$$

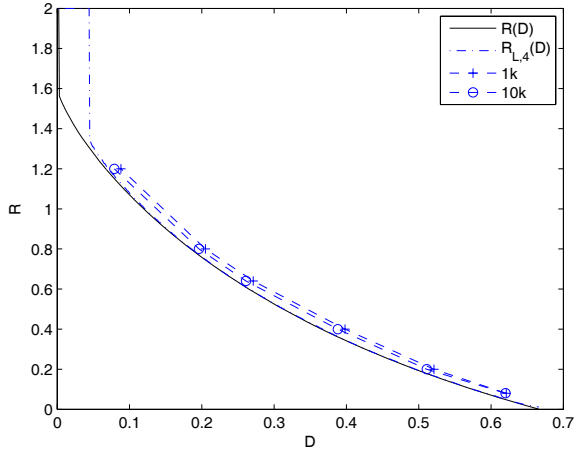


Fig. 6. Uniform ternary source with ℓ_1 distortion measure.

Without loss of generality, here we only define $P_{\hat{X}|\hat{X}^*}(\cdot|\hat{x})$ for $\hat{x} \in \hat{\mathcal{X}}_+$, where $\hat{\mathcal{X}}_+ = \{\hat{x} \in \hat{\mathcal{X}} : P_{\hat{X}^*}(\hat{x}) > 0\}$. It can be verified that

$$P_{\hat{X}}(\tilde{x}) = \sum_{\hat{x} \in \hat{\mathcal{X}}_+} P_{\hat{X}^*}(\hat{x}) P_{\hat{X}|\hat{X}^*}(\tilde{x}|\hat{x}) = \frac{1}{2^m}$$

if $\tilde{x} \in \mathcal{K}_{\hat{x}}$ for some $\hat{x} \in \hat{\mathcal{X}}$, and

$$\begin{aligned} P_{\hat{X}}(\tilde{x}) &= \sum_{\hat{x} \in \hat{\mathcal{X}}_+} P_{\hat{X}^*}(\hat{x}) P_{\hat{X}|\hat{X}^*}(\tilde{x}|\hat{x}) \\ &= \frac{1}{|\mathcal{K}^*|} \sum_{\hat{x} \in \hat{\mathcal{X}}_+} \left(P_{\hat{X}^*}(\hat{x}) - \frac{k_{\hat{x}}}{2^m} \right) = \frac{1}{2^m} \end{aligned}$$

if $\tilde{x} \in \mathcal{K}^*$. Therefore, the constructed \tilde{X} is uniformly distributed over \mathbb{F}_2^m .

Define a deterministic mapping $f : \mathbb{F}_2^m \rightarrow \hat{\mathcal{X}}$ such that $f(\tilde{x}) = \hat{x}$ if $\tilde{x} \in \mathcal{K}_{\hat{x}}$ for some $\hat{x} \in \hat{\mathcal{X}}$, and $f(\tilde{x})$ is set to be an arbitrary reconstruction symbol if $\tilde{x} \in \mathcal{K}^*$. By the data processing inequality, we have

$$R(D) = I(X; \hat{X}^*) \geq I(X; \tilde{X}) \geq I(X; f(\tilde{X})). \quad (11)$$

Moreover, it is easy to verify that $\lim_{m \rightarrow \infty} P_{X, f(\tilde{X})}(x, \hat{x}) = P_{X, \hat{X}^*}(x, \hat{x})$ for all $x \in \mathcal{X}$ and $\hat{x} \in \hat{\mathcal{X}}$. Therefore, we have

$$\lim_{m \rightarrow \infty} \mathbb{E}[d(X, f(\tilde{X}))] = \mathbb{E}[d(X, \hat{X}^*)] \leq D. \quad (12)$$

By (11), (12), and the continuity of the rate-distortion function, it can be readily shown that $\limsup_{m \rightarrow \infty} R_{L,m}(D) \leq R(D)$. The proof is complete.

APPENDIX B PROOF OF THEOREM 2

Let $\mathcal{T}_{P_X}^{\delta,n}$ denote the δ -strongly typical set of sequences x^n with respect to the probability distribution P_X . Similarly, let $\mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n}$ denote the δ -strongly typical set of sequences (x^n, \tilde{x}^n) with respect to $P_{X,\tilde{X}}$, where $P_{X,\tilde{X}}$ is the probability distribution induced by P_X and $P_{\tilde{X}|X}$. Note that if $(x^n, \tilde{x}^n) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n}$, then $x^n \in \mathcal{T}_{P_X}^{\delta,n}$. The definition of strongly typical set and its properties can be found in [24].

For each $x^n \in \mathcal{X}^n$, define

$$\phi_n(x^n, \mathbf{G}) = \sum_{u^k \in \mathbb{F}_2^k} \mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n}),$$

$$\begin{aligned} \psi_n(x^n, \mathbf{G}) &= \sum_{u^k \in \mathbb{F}_2^k} \mathbb{I}\left(\frac{1}{n} \sum_{i=1}^n d(x_i, \hat{X}_i(u^k)) \leq \mathbb{E}[d(X, f(\tilde{X}))] + \frac{\epsilon}{2}\right). \end{aligned}$$

Let δ be small enough so that $\frac{1}{n} \sum_{i=1}^n d(x_i, f(\tilde{x}_i)) \leq \mathbb{E}[d(X, f(\tilde{X}))] + \frac{\epsilon}{2}$ whenever $(x^n, \tilde{x}^n) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n}$. As a consequence, we have

$$\phi_n(x^n, \mathbf{G}) \leq \psi_n(x^n, \mathbf{G}) \quad (13)$$

for all $x^n \in \mathcal{X}^n$.

Let $\{w_n\}$ be a sequence such that $w_n p_n \rightarrow \infty$ and $\frac{w_n}{n} \rightarrow 0$ as $n \rightarrow \infty$. Since $Y_i(u^k) = u^k \mathbf{g}_i$, $i = 1, 2, \dots, mn$, it is easy to see that $Y_1(u^k), Y_2(u^k), \dots, Y_{mn}(u^k)$ are independent for any fixed $u^k \in \mathbb{F}_2^k$. Moreover, in view of the fact that

$$\Pr\{u^k \mathbf{g}_i = 0\} = \frac{1 + (1 - 2p_n)^{w(u^k)}}{2},$$

we have

$$\Pr\{Y_i(u^k) = 0\} \rightarrow \frac{1}{2} \quad (14)$$

as $n \rightarrow \infty$, where the convergence in (14) is uniform for all $i \in \{1, 2, \dots, mn\}$ and $u^k \in \mathbb{F}_2^k$ with $w(u^k) \geq w_n$. Therefore, it can be shown using [24, Lemma 13.6.2, p. 359] and a continuity argument that for any $\delta' > 0$, one can choose a sufficiently small δ such that

$$\begin{aligned} 2^{-n(I(X; f(\tilde{X})) - \delta')} &\geq \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \\ &\geq 2^{-n(I(X; f(\tilde{X})) + \delta')} \end{aligned} \quad (15)$$

for all $x^n \in \mathcal{T}_{P_X}^{\delta,n}$ and $u^k \in \mathbb{F}_2^k$ with $w(u^k) \geq w_n$ when n is sufficiently large. Note that

$$\begin{aligned} &\sum_{u^k \in \mathbb{F}_2^k} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \\ &= \sum_{w(u^k) \geq w_n} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \\ &\quad + \sum_{w(u^k) < w_n} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})]. \end{aligned}$$

Let $R \triangleq R(n, k) = \frac{k}{n}$. In view of (15) and the fact that $2^{nR} \geq |\{u^k \in \mathbb{F}_2^k : w(u^k) \geq w_n\}| \geq 2^{n(R - \delta')}$ and $|\{u^k \in \mathbb{F}_2^k : w(u^k) < w_n\}| \leq 2^{n\delta'}$ for sufficiently large n , we have

$$\begin{aligned} &2^{n(R - I(X; f(\tilde{X})) + \delta')} + 2^{n\delta'} \\ &\geq \sum_{u^k \in \mathbb{F}_2^k} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \\ &\geq 2^{n(R - I(X; f(\tilde{X})) - 2\delta')} \end{aligned} \quad (16)$$

for all $x^n \in \mathcal{T}_{P_X}^{\delta,n}$ when n is sufficiently large.

Now we shall derive a lower bound on $\mathbb{E}[\phi_n(X^n, \mathbf{G})]$. Note that

$$\mathbb{E}[\phi_n(X^n, \mathbf{G})]$$

$$\begin{aligned}
 &= \sum_{x^n \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n}} \Pr\{X^n = x^n\} \sum_{u^k \in \mathbb{F}_2^k} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \\
 &+ \sum_{x^n \notin \mathcal{T}_{P_X}^{\delta,n}} \Pr\{X^n = x^n\} \sum_{u^k \in \mathbb{F}_2^k} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \\
 &= \sum_{x^n \in \mathcal{T}_{P_X}^{\delta,n}} \Pr\{X^n = x^n\} \sum_{u^k \in \mathbb{F}_2^k} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})],
 \end{aligned}$$

where the last equality is because $\mathbb{I}((x^n, \tilde{x}^n) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n}) = 0$ for all \tilde{x}^n if $x^n \notin \mathcal{T}_{P_X}^{\delta,n}$. In view of (16) and the fact that $\Pr\{X^n \in \mathcal{T}_{P_X}^{\delta,n}\} \geq 1 - \delta'$ for all sufficiently large n , we have

$$\mathbb{E}[\phi_n(X^n, \mathbf{G})] \geq (1 - \delta')2^{n(R-I(X;f(\tilde{X}))-2\delta')} \quad (17)$$

when n is sufficiently large.

Now we proceed to derive an upper bound on $\mathbb{E}[(\phi_n(X^n, \mathbf{G}))^2]$. Note that

$$\begin{aligned}
 &\mathbb{E}[(\phi_n(X^n, \mathbf{G}))^2] \\
 &= \mathbb{E}\left[\sum_{u^k, v^k \in \mathbb{F}_2^k} \mathbb{I}((X^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n}) \mathbb{I}((X^n, \tilde{X}^n(v^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})\right] \\
 &= \sum_{x^n \in \mathcal{T}_{P_X}^{\delta,n}} \Pr\{X^n = x^n\} \mathbb{E}\left[\sum_{u^k, v^k \in \mathbb{F}_2^k} \mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n}) \mathbb{I}((x^n, \tilde{X}^n(v^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})\right] \\
 &= \sum_{x^n \in \mathcal{T}_{P_X}^{\delta,n}} \Pr\{X^n = x^n\} \mathbb{E}\left[\sum_{w(u^k \oplus v^k) \geq 2w_n} \mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n}) \mathbb{I}((x^n, \tilde{X}^n(v^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})\right] \\
 &+ \sum_{w(u^k \oplus v^k) < 2w_n} \mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n}) \mathbb{I}((x^n, \tilde{X}^n(v^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})], \quad (18)
 \end{aligned}$$

where \oplus is the modulo-2 addition. Define $(u \wedge v)^k = ((u \wedge v)_1, (u \wedge v)_2, \dots, (u \wedge v)_k)$ with $(u \wedge v)_i = 1$ if $u_i = v_i = 1$ and $(u \wedge v)_i = 0$ otherwise. Define $(u \setminus v)^k = u^k \oplus (u \wedge v)^k$ and $(v \setminus u)^k = v^k \oplus (u \wedge v)^k$. Note that $w(u^k \oplus v^k) \geq 2w_n$ implies $w((u \setminus v)^k) \geq w_n$ or $w((v \setminus u)^k) \geq w_n$. Without loss of generality, we shall assume $w((v \setminus u)^k) \geq w_n$, which further implies $w(v^k) \geq w_n$. In view of (14), we have

$$\begin{aligned}
 \Pr\{\tilde{X}^n(v^k) = \tilde{x}^n\} &\geq 2^{-mn} 2^{-\frac{n\delta'}{2}}, \\
 \Pr\{\tilde{X}^n((v \setminus u)^k) = \tilde{x}^n\} &\leq 2^{-mn} 2^{\frac{n\delta'}{2}}
 \end{aligned}$$

for all \tilde{x}^n when n is sufficiently large. Therefore, if $w(u^k \oplus v^k) \geq 2w_n$, then

$$\begin{aligned}
 &\Pr\{\tilde{X}^n(u^k) = \tilde{a}^n, \tilde{X}^n(v^k) = \tilde{b}^n\} \\
 &= \sum_{\tilde{c}^n} \Pr\{\tilde{X}^n((u \setminus v)^k) = \tilde{a}^n \oplus \tilde{c}^n, \\
 &\quad \tilde{X}^n((v \setminus u)^k) = \tilde{b}^n \oplus \tilde{c}^n, \tilde{X}^n((u \wedge v)^k) = \tilde{c}^n\} \\
 &= \sum_{\tilde{c}^n} \Pr\{\tilde{X}^n((u \setminus v)^k) = \tilde{a}^n \oplus \tilde{c}^n, \tilde{X}^n((u \wedge v)^k) = \tilde{c}^n\} \\
 &\quad \times \Pr\{\tilde{X}^n((v \setminus u)^k) = \tilde{b}^n \oplus \tilde{c}^n\}
 \end{aligned}$$

$$\begin{aligned}
 &\leq 2^{-mn} 2^{\frac{n\delta'}{2}} \sum_{\tilde{c}^n} \Pr\{\tilde{X}^n((u \setminus v)^k) = \tilde{a}^n \oplus \tilde{c}^n, \\
 &\quad \tilde{X}^n((u \wedge v)^k) = \tilde{c}^n\}
 \end{aligned}$$

$$\begin{aligned}
 &= 2^{-mn} 2^{\frac{n\delta'}{2}} \Pr\{\tilde{X}^n(u^k) = \tilde{a}^n\} \\
 &\leq 2^{n\delta'} \Pr\{\tilde{X}^n(u^k) = \tilde{a}^n\} \Pr\{\tilde{X}^n(v^k) = \tilde{b}^n\} \quad (19)
 \end{aligned}$$

uniformly for all \tilde{a}^n and \tilde{b}^n when n is sufficiently large. Continuing from (18),

$$\begin{aligned}
 &\mathbb{E}[(\phi_n(X^n, \mathbf{G}))^2] \\
 &\leq \sum_{x^n \in \mathcal{T}_{P_X}^{\delta,n}} \Pr\{X^n = x^n\} \left[\sum_{w(u^k \oplus v^k) \geq 2w_n} 2^{n\delta'} \right. \\
 &\quad \times \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(v^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \\
 &\quad + \sum_{w(u^k \oplus v^k) < 2w_n} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \\
 &\quad \left. \times \mathbb{I}((x^n, \tilde{X}^n(v^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n}) \right] \quad (20)
 \end{aligned}$$

$$\begin{aligned}
 &\leq \sum_{x^n \in \mathcal{T}_{P_X}^{\delta,n}} \Pr\{X^n = x^n\} \left[\sum_{u^k, v^k \in \mathbb{F}_2^k} 2^{n\delta'} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(v^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \right. \\
 &\quad + \sum_{w(u^k \oplus v^k) < 2w_n} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \left. \right] \\
 &\leq \sum_{x^n \in \mathcal{T}_{P_X}^{\delta,n}} \Pr\{X^n = x^n\} \left[\sum_{u^k, v^k \in \mathbb{F}_2^k} 2^{n\delta'} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(v^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \right. \\
 &\quad + 2^{n\delta'} \sum_{u^k \in \mathbb{F}_2^k} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \left. \right] \quad (21)
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{x^n \in \mathcal{T}_{P_X}^{\delta,n}} \Pr\{X^n = x^n\} \left[2^{n\delta'} \left(\sum_{u^k \in \mathbb{F}_2^k} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \right)^2 \right. \\
 &\quad + 2^{n\delta'} \sum_{u^k \in \mathbb{F}_2^k} \mathbb{E}[\mathbb{I}((x^n, \tilde{X}^n(u^k)) \in \mathcal{T}_{P_{X,\tilde{X}}}^{\delta,n})] \left. \right] \\
 &\leq 2^{n\delta'} (2^{n(R-I(X;f(\tilde{X}))+\delta')} + 2^{n\delta'})^2 \\
 &\quad + 2^{n\delta'} (2^{n(R-I(X;f(\tilde{X}))+\delta')} + 2^{n\delta'}), \quad (22)
 \end{aligned}$$

where (20) and (22) follow from (19) and (16), respectively, while (21) is due to the fact that for any $u^k \in \mathbb{F}_2^k$, $|\{v^k \in \mathbb{F}_2^k : w(u^k \oplus v^k) < 2w_n\}| \leq 2^{n\delta'}$ when n is sufficiently large.

By [7, Lemma 2],

$$\Pr\{\phi_n(X^n, \mathbf{G}) > 0\} \geq \frac{(\mathbb{E}[\phi_n(X^n, \mathbf{G})])^2}{\mathbb{E}[(\phi_n(X^n, \mathbf{G}))^2]}.$$

Set $R = I(X; f(\tilde{X})) + \epsilon$. In view of (13), (17), and (22), we have

$$\begin{aligned}
 &\liminf_{n \rightarrow \infty} \frac{1}{n} \log \Pr\{\psi_n(X^n, \mathbf{G}) > 0\} \\
 &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \Pr\{\phi_n(X^n, \mathbf{G}) > 0\} \\
 &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \frac{(\mathbb{E}[\phi_n(X^n, \mathbf{G})])^2}{\mathbb{E}[(\phi_n(X^n, \mathbf{G}))^2]} \\
 &\geq -7\delta'.
 \end{aligned}$$

Since $\delta' > 0$ is arbitrary, it follows that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \Pr\{\psi_n(X^n, \mathbf{G}) > 0\} \geq 0.$$

Now invoking [7, Lemma 1], one can readily show that

$$\begin{aligned} D(n, k, m, f) &\leq \mathbb{E}[d(X, f(\tilde{X}))] + \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \mathbb{E}[d(X, f(\tilde{X}))] + \epsilon \end{aligned}$$

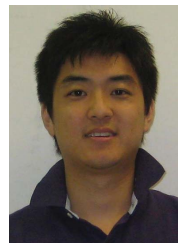
for some sufficiently large n . The proof is complete.

ACKNOWLEDGMENT

The authors would like to thank Dr. Da-ke He and Dr. Ashish Jagmohan for their contribution in the early stages of this work. They also wish to thank Prof. Alexander Barg and Prof. Prakash Narayan of the University of Maryland for helpful discussions.

REFERENCES

- [1] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec., Part 4*, pp. 142-163, 1959.
- [2] P. Elias, "Coding for noisy channels," *IRE Conv. Rec., Part 4*, pp. 37-46, 1955.
- [3] E. M. Gabidulin, "Limits for the decoding error probability when linear codes are used in memoryless channel," *Probl. Inf. Transm.*, pp. 43-48, 1967, translated from *Probl. Pered. Inform.*.
- [4] F. Jelinek, "Tree encoding of memoryless time-discrete sources with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 5 pp. 584-590, Sep. 1969.
- [5] A. Viterbi and J. Omura, "Trellis encoding of memoryless discrete-time sources with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 3, pp. 325-332, May 1974.
- [6] E. Martinian and J. S. Yedidia, "Iterative quantization using codes on graphs," *41th Annual Allerton Conf. Commun., Control, Comput., Monticello, IL*, Oct. 2003.
- [7] M. J. Wainwright and E. Martinian, "Low-density graph codes that are optimal for binning and coding with side information," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1061-1079, Mar. 2009.
- [8] T. Murayama, "Thouless-Anderson-Palmer approach for lossy compression," *Physical Review E*, vol. 69, no. 035105, pp. 1-4, 2004.
- [9] M. J. Wainwright and E. Maneva, "Lossy source coding via message-passing and decimation over generalized codewords of LDGM codes," *IEEE International Symp. Inf. Theory, Adelaide, Australia*, Sept. 2005.
- [10] S. Cilibertia, M. Mézard, and R. Zecchina, "Lossy data compression with random gates," *Phys. Rev. Lett.*, vol. 95, no. 038701, 2005.
- [11] T. Filler and J. Fridrich, "Binary quantization using belief propagation with decimation over factor graphs of LDGM codes," in *Proc. Allerton Conf. Commun., Control, Comput., Monticello, IL*, Sep. 2007.
- [12] M. Mézard, G. Parisi, and R. Zecchina, "Analytic and algorithmic solution of random satisfiability problems," *Science*, vol. 297, pp. 812-815, June 2002.
- [13] E. Maneva, E. Mossel, and M. J. Wainwright, "A new look at survey propagation and its generalizations," in *Proc. 16th Annual Symp. Discrete Algorithms (SODA)*, pp. 1089-1098, Jan. 2005.
- [14] A. Braunstein, M. Mézard, and R. Zecchina, "Survey propagation: an algorithm for satisfiability," *Random Structures Algorithms*, vol. 27, no. 2, pp. 201-226, Mar. 2005.
- [15] J. L. Massey, "Joint source and channel coding," *Commun. Syst. Random Process Theory*, vol. 11, pp. 279-293, 1978.
- [16] T. C. Anчета, Jr., "Syndrome-source-coding and its universal generalizations," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 4, pp. 432-436, July 1976.
- [17] E.-H. Yang and J. Kieffer, "Simple universal lossy data compression schemes derived from the Lempel-Ziv algorithm," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 239-245, Jan. 1996.
- [18] E.-H. Yang, Z. Zhang, and T. Berger, "Fixed-slope universal lossy data compression," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1465-1479, Sep. 1996.
- [19] I. Kontoyiannis, "An implementable lossy version of the Lempel-Ziv algorithm—part I: optimality for memoryless sources," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2293-2305, Nov. 1999.
- [20] A. Gupta, S. Verdù, and T. Weissman, "Rate-distortion in near-linear time," *IEEE International Symp. Inf. Theory*, Toronto, Canada, July 2008.
- [21] A. Gupta and S. Verdù, "Nonlinear sparse-graph codes for lossy compression," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 1961-1975, May 2009.
- [22] S. Jalali and T. Weissman, "Rate-distortion via Markov chain Monte Carlo," *IEEE Trans. Inf. Theory*, submitted for publication. [Online]. Available: <http://arxiv.org/abs/0808.4156>.
- [23] S. B. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1751-1768, Apr. 2010.
- [24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [25] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
- [26] T. Weissman and E. Ordentlich, "The empirical distribution of rate-constrained source codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3718-3733, Nov. 2005.
- [27] A. Kanlis, S. Khudanpur, and P. Narayan, "Typicality of a good rate-distortion code" (in Russian), *Probl. Pered. Inform. (Probl. Inf. Transm.)*, vol. 32, no. 1, pp. 96-103, 1996.
- [28] A. Barg and G. D. Forney, Jr., "Random codes: minimum distances and error exponents," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2568-2573, Sep. 2002.
- [29] A. Dembo and I. Kontoyiannis, "Source coding, large deviations, and approximate pattern matching," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1590-1615, June 2002.
- [30] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [31] S. Kudekar and R. Urbanke, "Lower bounds on the rate-distortion function of individual LDGM codes," *5th International Symp. Turbo Codes Related Topics*, Lausanne, Switzerland, pp. 379-384, Sep. 2008.
- [32] A. G. Dimakis, M. J. Wainwright, and K. Ramchandran, "Lower bounds on the rate-distortion function of LDGM codes," *IEEE Inf. Theory Workshop*, Lake Tahoe, CA, Sep. 2007.
- [33] [Online]. Available: <http://lthcwww.epfl.ch/research/ldpccpt>



Zhibin Sun received the B.Eng. and M.A.Sc. degrees in electrical engineering from McMaster University, Hamilton, ON, Canada, in 2007 and 2009, respectively. He joined the Department of Standards and New Technology Engineering at Hydro One Networks Inc., Toronto, ON, Canada, in 2009, where he is now an Engineering Trainee.



Mingkai Shao received the B.Eng. degree in computer engineering from Huazhong University of Science and Technology, China, in 2004, and the M.Sc. degree in computer science from Wuhan University, China, in 2006. He is currently pursuing the Ph.D. degree in computer engineering at McMaster University, Hamilton, ON, Canada. His research interests include information theory and multimedia communications.



Jun Chen (S'03-M'06) received the B.E. degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2004 and 2006, respectively.

He was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, from 2005 to 2006, and a Josef Raviv Memorial Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from 2006 to 2007. He is currently an Assistant Professor of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada. He holds the Barber-Gennum Chair in Information Technology. His research interests include information theory, wireless communications, and signal processing.



Kon Max Wong (SM'81-F'02) received his BSc(Eng), DIC, Ph.D., and DSc(Eng) degrees, all in electrical engineering, from the University of London, England, in 1969, 1972, 1974 and 1995, respectively. He started working at the Transmission Division of Plessey Telecommunications Research Ltd., England, in 1969. In October 1970 he was on leave from Plessey pursuing postgraduate studies and research at Imperial College of Science and Technology, London. In 1972, he rejoined Plessey as a research engineer and worked on digital signal

processing and signal transmission. In 1976, he joined the Department of Electrical Engineering at the Technical University of Nova Scotia, Canada, and in 1981, moved to McMaster University, Hamilton, Canada, where he has been a Professor since 1985 and served as Chairman of the Department of Electrical and Computer Engineering in 1986-87, 1988-94 and 2003-08. Professor Wong was on leave as Visiting Professor at the Department of Electronic Engineering of the Chinese University of Hong Kong from 1997 to 1999. At present, he holds the Canada Research Chair in Signal Processing at McMaster University. His research interest is in signal processing and communication theory and has published over 250 papers in the area.

Professor Wong was the recipient of the IEE Overseas Premium for the best paper in 1989, and is also the co-author of the papers that received the IEEE Signal Processing Society "Best Young Author" awards of 2006 and 2008. In 2009, he was awarded the Royal Academy of Engineering Distinguished Visiting Fellowship, and he has also been named a recipient of the Alexander von Humboldt Research Award for 2010. He is a Fellow of IEEE, a Fellow of the Institution of Electrical Engineers, a Fellow of the Royal Statistical Society, and a Fellow of the Institute of Physics. More recently, he has also

been elected as Fellow of the Canadian Academy of Engineering as well as Fellow of the Royal Society of Canada. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, 1996-98 and served as Chair of the Sensor Array and Multi-channel Signal Processing Technical Committee of the IEEE Signal Processing Society in 2002-04. Professor Wong was the recipient of a medal presented by the International Biographical Centre, Cambridge, England, for his "outstanding contributions to the research and education in signal processing" in May 2000, and was honoured with the inclusion of his biography in the two books: *Outstanding People of the 20th Century* and *2000 Outstanding Intellectuals of the 20th Century*, published by IBC to celebrate the arrival of the new millennium.



Xiaolin Wu (M'89-SM'96) got his B.Sc. from Wuhan University, China in 1982, and Ph.D. from the University of Calgary, Canada, in 1988, both in computer science. Dr. Wu started his academic career in 1988, and has since been on the faculty of the University of Western Ontario, New York Polytechnic University, and currently McMaster University, where he is a professor at the Department of Electrical and Computer Engineering and holds the NSERC-DALSA industrial research chair in Digital Cinema. His research interests include multimedia

signal compression, joint source-channel coding, multiple description coding, network-aware visual communication and image processing. He has published over two hundred research papers and holds two patents in these fields. Dr. Wu is an associate editor of IEEE TRANSACTIONS ON MULTIMEDIA and an associate editor of IEEE TRANSACTIONS ON IMAGE PROCESSING.