Robust Multiresolution Coding

Jun Chen, *Member, IEEE*, Sorina Dumitrescu, *Member, IEEE*, Ying Zhang, Jia Wang

Abstract

In multiresolution coding a source sequence is encoded into a base layer and a refinement layer. The refinement layer, constructed using a conditional codebook, is in general not decodable without the correct reception of the base layer. By relating multiresolution coding with multiple description coding, we show that it is in fact possible to construct multiresolution codes in certain ways so that the refinement layer alone can be used to reconstruct the source to achieve a nontrivial distortion. As a consequence, one can improve the robustness of the existing multiresolution coding schemes without sacrificing the efficiency. Specifically, we obtain an explicit expression of the minimum distortion achievable by the refinement layer for arbitrary finite alphabet sources with Hamming distortion measure. Experimental results show that the information-theoretic limits can be approached using a practical robust multiresolution coding scheme based on low-density generator matrix codes.

Index Terms

Low-density generator matrix, message-passing algorithm, multiple description coding, multiresolution coding, successive refinement.

I. Introduction

Many important applications require multicast delivery of data from a single user to multiple receivers with diverse characteristics in terms of bandwidth resources, computational capabilities,

Jun Chen, Sorina Dumitrescu, and Ying Zhang are with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (email: {junchen,sorina}@mail.ece.mcmaster.ca, yingzhang@grads.ece.mcmaster.ca).

Jia Wang is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (email: jiawang@sjtu.edu.cn).

The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Austin, Texas, USA, Jun. 13-18, 2010.

and fidelity requirements. It is desirable that the reconstruction quality at each receiver is commensurate with its own demand and capability. As a promising solution to this problem, multiresolution coding has received significant attention in recent years [1]–[8].

In multiresolution coding a source sequence is encoded into a base layer and a refinement layer; a coarse reconstruction of the source is possible based on the base layer while the two layers together can lead to better reconstruction quality. Although it is commonly assumed that the role of the refinement layer is simply to improve the reconstruction precision upon that achieved by the base layer, it is of considerable interest to know whether the refinement layer alone can be used to reconstruct the source. Unfortunately, for most existing multiresolution coding schemes, the refinement layer is constructed using a conditional codebook, thus is undecodable without the correct reception of the base layer or is essentially useless for producing any nontrivial reconstruction. By interpreting multiresolution coding as a special case of multiple description coding and leveraging relevant multiple description code constructions, we shall show that it is in fact possible to design multiresolution codes in certain ways so that the refinement layer alone can be (partially) decoded to produce a nontrivial reconstruction of the source. This is certainly a desirable feature since it improves the robustness of multiresolution codes without sacrificing the efficiency.

The remainder of this paper is organized as follows. In Section II, we discuss the connection between multiresolution coding and multiple description coding. Some existing results on these two coding problems are reviewed. In Section III, we derive an explicit expression of the minimum distortion achievable by the refinement layer for arbitrary finite alphabet sources with Hamming distortion measure. A practical robust multiresolution coding scheme based on low-density generator matrix (LDGM) codes is proposed in Section IV. The effectiveness of the proposed scheme is verified in Section V. Finally, we conclude the paper in Section VI.

II. Multiresolution Coding and Multiple Description Coding

We shall first review some basic definitions and results regarding multiresolution coding and multiple description coding. It will be seen that a key step toward understanding the role of refinement layer in multiresolution coding is to interpret multiresolution coding as a special form of multiple description coding.

A. Multiple Description Coding

In the multiple description problem, a source sequence is encoded into two descriptions, which are constructed in such a way that an adequate reconstruction of the source is possible based on each description while the two descriptions together can lead to better reconstruction quality. A fundamental problem of multiple description coding is to characterize the rate-distortion region, which determines the information-theoretic limits of multiple description coding.

Consider an i.i.d. process $\{X(l)\}_{l=1}^{\infty}$ with marginal distribution p_X on source alphabet $\mathcal{X}$. Let $d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$ be a distortion measure, where $\hat{\mathcal{X}}$ is the reconstruction alphabet. We assume that $\mathcal{X}$ and $\hat{\mathcal{X}}$ are finite sets.

Definition 1: The quintuple $(R_1, R_2, D_0, D_1, D_2)$ is said achievable, if for all sufficiently large n, there exist encoding functions

$$f_i^{(n)} : \mathcal{X}^n \to \{1, 2, \cdots, \lfloor 2^{nR_i} \rfloor\}, \quad i = 1, 2,$$

and decoding functions

$$g_0^{(n)} : \{1, 2, \cdots, \lfloor 2^{nR_1} \rfloor\} \times \{1, 2, \cdots, \lfloor 2^{nR_2} \rfloor\} \to \hat{\mathcal{X}}^n,$$
$$g_i^{(n)} : \{1, 2, \cdots, \lfloor 2^{nR_i} \rfloor\} \to \hat{\mathcal{X}}^n, \quad i = 1, 2,$$

such that

$$\mathbb{E}\left[\frac{1}{n}\sum_{l=1}^{n} d(X(l), \hat{X}_i(l))\right] \leq D_i, \quad i = 0, 1, 2,$$

where $\hat{X}_0^n = g_0^{(n)}(f_1^{(n)}(X^n), f_2^{(n)}(X^n))$ and $\hat{X}_i^n = g_i^{(n)}(f_i^{(n)}(X^n))$, $i = 1, 2$. The multiple description rate-distortion region, denoted by $\mathcal{RD}_{\text{MD}}$, is the closure of the set of all achievable quintuples $(R_1, R_2, D_0, D_1, D_2)$.

While a computable characterization of $\mathcal{RD}_{\text{MD}}$ is still unknown, several inner bounds of $\mathcal{RD}_{\text{MD}}$ can be found in the literature [9]–[11], among which the EGC inner bound [9] is the one that is particularly relevant to our setting. Specifically, the EGC inner bound $\mathcal{RD}_{\text{EGC}}$ is the convex hull of the set of quintuples $(R_1, R_2, D_0, D_1, D_2)$ for which there exist auxiliary random variables X_i, $i = 0, 1, 2$, jointly distributed with the generic source variable X, such that

$$R_i \geq I(X; X_i), \quad i = 1, 2,$$

$$R_1 + R_2 \geq I(X; X_0, X_1, X_2) + I(X_1; X_2),$$

$$D_i \geq \mathbb{E}[d(X, X_i)], \quad i = 0, 1, 2.$$

The coding scheme associated with the EGC inner bound can be roughly understood as follows. Generate codebook 1 and codebook 2 using marginal distributions p_{X_1} and p_{X_2}, respectively. For each pair of codewords, one from codebook 1 and the other from codebook 2, generate a codebook using the conditional distribution $p_{X_0|X_1X_2}$; such a codebook will be referred to as a conditional codebook. The source X is encoded into two descriptions, where description 1 contains an index specifying a codeword X_1 in codebook 1 and a portion of index specifying a codeword X_0 in the conditional codebook while description 2 contains an index specifying a codeword X_2 in codebook 2 and the remaining portion of index for X_0. Here the conditional codebook itself is specified by X_1 and X_2 (or equivalently, the indices of X_1 and X_2). Given a single description, say, description i, one can decode X_i and use it as the reconstruction of X. If both descriptions are received, then one can decode X_0 and use it as the reconstruction. Note that given a single description, it is in general impossible to (even partially) decode X_0 since the available information is not enough to specify the conditional codebook from which X_0 is picked; moreover, if such a description only contains a partial index for X_0, then the position of X_0 in the conditional codebook is also ambiguous.

B. Multiresolution Coding

It is instructive to view multiresolution coding as a special form of multiple description coding in which the distortion constraint on the second description (i.e., D_2) is not imposed. In this scenario it is common to refer to the first description as the base layer and the second description as the refinement layer.

Definition 2: The multiresolution coding rate-distortion region $\mathcal{RD}_{\text{MR}}$ is given by

$$\mathcal{RD}_{\text{MR}} = \{(R_1, R_2, D_0, D_1) : (R_1, R_2, D_0, D_1, \infty) \in \mathcal{RD}_{\text{MD}}\}.$$

As shown in [12], $\mathcal{RD}_{\text{MR}}$ is equal to the set of quadruples (R_1, R_2, D_0, D_1) for which there exist auxiliary random variables X_0 and X_1, jointly distributed with the generic source variable

X, such that

$$\begin{aligned} &R_1 \geq I(X; X_1), \\ &R_1 + R_2 \geq I(X; X_0, X_1), \\ &D_i \geq \mathbb{E}[d(X, X_i)], \quad i = 0, 1. \end{aligned}$$

It is easy to see that $\mathcal{RD}_{\text{MR}}$ is equivalent to $\mathcal{RD}_{\text{EGC}}$ with X_2 set to be a constant. Such a connection is well understood. In fact, most existing multiresolution code constructions are based on this interpretation of $\mathcal{RD}_{\text{MR}}$. However, this interpretation has the following implicit consequence on the resulting constructions, that is, the refinement layer alone is in general useless for reconstructing the source. Indeed, in the aforedescribed EGC scheme, if one sets X_2 to be a constant, then the second description alone is in general not (even partially) decodable since it only contains a portion of index specifying a codeword X_0 in the conditional codebook. It will be seen that there is an alternative way to design the refinement layer based on a deeper connection between $\mathcal{RD}_{\text{MR}}$ and $\mathcal{RD}_{\text{EGC}}$.

C. Connection

In the most general formulation, if one simply imposes the requirement that the refinement layer alone can be used to produce a nontrivial reconstruction of the source, then multiresolution coding becomes equivalent to multiple description coding. In practice, multiresolution coding often has a more restricted meaning: loosely speaking, the base layer and the refinement layer should be constructed in a greedy manner to achieve the minimum distortion at each reconstruction step. This is the case where multiresolution coding is most interesting. Indeed, such a greedy property can even be viewed as the essential feature of multiresolution coding. We shall show that in this natural setting it is possible to determine the minimum distortion achievable by the refinement layer of a multiresolution code.

Let $R(D)$ denote the rate-distortion function, i.e., $R(D) = \min_{p_{\hat{X}|X}} I(X; \hat{X})$, where the minimization is over $p_{\hat{X}|X}$ subject to the constraint $\mathbb{E}[d(X, \hat{X})] \leq D$. Define

$$R(R_1, D_0, D_1) = \min\{R_1 + R_2 : (R_1, R_2, D_0, D_1) \in \mathcal{RD}_{\text{MR}}\}.$$

It can be shown [12] that

$$R(R_1, D_0, D_1) = \max\left\{R_1, \min_{p_{X_0X_1|X}} I(X; X_0, X_1)\right\},$$

where the minimization is over $p_{X_0X_1|X}$ subject to the constraints $\mathbb{E}[d(X, X_0)] \le D_0$, $I(X; X_1) \le R_1$, and $\mathbb{E}[d(X, X_1)] \le D_1$. Define

$$D_2^*(D_0, D_1) = \min_{\substack{R_1 = R(D_1) \\ R_1 + R_2 = R(R_1, D_0, D_1) \\ (R_1, R_2, D_0, D_1, D_2) \in \mathcal{RD}_{\mathrm{MD}}}} D_2.$$

Note that $D_2^*(D_0, D_1)$ can be interpreted as the minimum distortion achievable by the refinement layer in the case where $R_1 = R(D_1)$ and $R_1 + R_2 = R(R(D_1), D_0, D_1)$. The following result is a simple consequence of [13, Lemma 3].

Theorem 1: Let $\mathcal{Q}$ denote the convex hull of the set of quintuples $(R_1, R_2, D_0, D_1, D_2)$ for which there exist auxiliary random variables X_i, $i = 0, 1, 2$, jointly distributed with the generic source variable X, such that

$$I(X_1; X_2) = 0,$$

$$R_i \ge I(X; X_i), \quad i = 1, 2,$$

$$R_1 + R_2 \ge I(X; X_0, X_1, X_2),$$

$$D_i \ge \mathbb{E}[d(X, X_i)], \quad i = 0, 1, 2.$$

We have

$$D_2^*(D_1, D_0) = \min_{\substack{R_1 = R(D_1) \\ R_1 + R_2 = R(R_1, D_0, D_1) \\ (R_1, R_2, D_0, D_1, D_2) \in \mathcal{Q}}} D_2. \tag{1}$$

Remark: This result can also be proved by invoking [14, Theorem 1] if $R(R_1, D_1, D_0) = R(D_0)$.

Note that one can obtain $\mathcal{Q}$ from $\mathcal{RD}_{\mathrm{EGC}}$ by imposing an additional constraint $I(X_1; X_2) = 0$ (i.e., X_1 and X_2 are independent). This reveals a new perspective on multiresolution coding. Roughly speaking, to obtain a multiresolution coding scheme from an EGC scheme, one just needs to let X_2 be independent of X_1, instead of setting X_2 to be a constant. In this way, the refinement layer alone is still useful since one can decode X_2 and use it as the reconstruction of the source.

In principle it is possible to compute $D_2^*(D_0, D_1)$ by solving the minimization problem in (1) via numerical methods. In the next section we shall derive an explicit expression of $D_2^*(D_0, D_1)$ for arbitrary finite alphabet sources with Hamming distortion measure.

III. Finite Alphabet Source With Hamming Distortion Measure

Let $\mathcal{X} = \hat{\mathcal{X}} = \{0, 1, \cdots, m\}$ for some positive integer m. Let $d = d_H : \mathcal{X} \times \hat{\mathcal{X}} \to \{0, 1\}$ be the Hamming distortion measure, i.e, $d_H(x, \hat{x}) = 0$ if $x = \hat{x}$ and $d_H(x, \hat{x}) = 1$ if $x \neq \hat{x}$. Note that $\mathbb{E}[d_H(X, \hat{X})] = \mathbb{P}(X \neq \hat{X})$. Without loss of generality, we shall assume $p_X(0) \geq p_X(1) \geq \cdots \geq p_X(m) > 0$.

Let $R(D)$ and $D(R)$ denote the rate-distortion function and the distortion-rate function, respectively, of source X with Hamming distortion measure. The following facts are well known [15]–[17].

F1) $D(R)|_{R=0} = 1 - \max_{x \in \mathcal{X}} p_X(x) = 1 - p_X(0)$.

F2) $R(D)$ is a strictly convex function of D for $D \in [0, D(R)|_{R=0}]$.

F3) For $0 \leq D_0 \leq D_1 < D(R)|_{R=0}$, we have

$$
\begin{aligned}
&I(X; X_1) = R(D_1), \\
&I(X; X_0, X_1) = R(D_0), \\
&\mathbb{P}(X \neq X_i) = D_i, \quad i = 0, 1,
\end{aligned}
$$

if and only if $X - X_0 - X_1$ form a Markov chain, and $p_{XX_0X_1}$ is specified by

$$
p_{X_i}(x) = \frac{(p_X(x) - \lambda_i)^+}{\sum_{x' \in \mathcal{X}} (p_X(x') - \lambda_i)^+}, \quad x \in \mathcal{X}, i = 0, 1, \tag{2}
$$

$$
p_{X|X_0}(x|x_0) = \begin{cases} 1 - D_0, & x = x_0 \in \mathcal{X}_0^+ \\ \lambda_0, & x \neq x_0, x \in \mathcal{X}_0^+, x_0 \in \mathcal{X}_0^+ \\ p_X(x) & x \notin \mathcal{X}_0^+, x_0 \in \mathcal{X}_0^+ \end{cases}, \tag{3}
$$

$p_{X_0|X_1}(x_0|x_1)$

$$= \begin{cases} \frac{1-D_1-\lambda_0}{1-D_0-\lambda_0}, & x_0 = x_1 \in \mathcal{X}_1^+ \\ \frac{\lambda_1-\lambda_0}{1-D_0-\lambda_0}, & x_0 \neq x_1, x_0 \in \mathcal{X}_1^+, x_1 \in \mathcal{X}_1^+ \\ \frac{p_X(x_0)-\lambda_0}{1-D_0-\lambda_0} & x_0 \in \mathcal{X}_0^+, x_0 \notin \mathcal{X}_1^+, x_1 \in \mathcal{X}_1^+ \end{cases}, \tag{4}$$

where $\lambda_0 \in [0, \lambda_1]$ and $\lambda_1 \in [0, p_X(1))$ are uniquely determined by

$$\sum_{x_i \in \mathcal{X}_i^+} p_{X_i}(x_i) p_{X|X_i}(x|x_i) = p_X(x), \quad x \in \mathcal{X}, i = 0, 1,$$

and $\mathcal{X}_i^+ = \{x \in \mathcal{X} : p_X(x) - \lambda_i > 0\}$, $i = 0, 1$.

F4) $R(R(D_1), D_0, D_1) = R(D_0)$ for $0 \leq D_0 \leq D_1 \leq D(R)|_{R=0}$.

The following theorem is the main result of this section.

Theorem 2: $D_2^*(D_1, D_0) = D(R)|_{R=0} + D_0 - D_1$ for $0 \leq D_0 \leq D_1 \leq D(R)|_{R=0}$.

Remark: For the special case of a binary symmetric source with Hamming distortion measure, Theorem 2 reduces to [13, Theorem 5]. Moreover, it is interesting to note that Theorem 2 also holds for the quadratic Gaussian case [18].

We shall prove Theorem 2 by establishing a series of lemmas. It is clear that Theorem 2 is true when $D_1 = D(R)|_{R=0}$. Therefore, we shall only consider the case $0 \leq D_0 \leq D_1 < D(R)|_{R=0}$. For $0 \leq D_0 \leq D_1 < D(R)|_{R=0}$, let $p_{XX_0X_1}$ be the probability distribution specified by (2)-(4) and the Markov chain constraint $X - X_0 - X_1$; define $\tilde{D}_2(D_0, D_1) = \min_{p_{X_2|XX_0X_1}} \mathbb{P}(X \neq X_2)$, where the minimization is over $p_{X_2|XX_0X_1}$ subject to the constraints $I(X; X_1, X_2|X_0) = 0$ (i.e., $X - X_0 - (X_1, X_2)$ form a Markov chain) and $I(X_1; X_2) = 0$ (i.e., X_1 and X_2 are independent).

Lemma 1: $D_2^*(D_0, D_1) = \tilde{D}_2(D_0, D_1)$ for $0 \leq D_0 \leq D_1 < D(R)|_{R=0}$.

Proof: For any X_2 such that $I(X; X_1, X_2|X_0) = 0$ and $I(X_1; X_2) = 0$, let

$$R_1 = I(X; X_1),$$

$$R_2 = I(X, X_1; X_2) + I(X; X_0|X_1, X_2),$$

$$D_i = \mathbb{P}(X \neq X_i), \quad i = 0, 1, 2.$$

Note that $R_2 \geq I(X; X_2)$ and

$$\begin{aligned} R_1 + R_2 &= I(X; X_1) + I(X, X_1; X_2) + I(X; X_0|X_1, X_2) \\ &= I(X; X_1) + I(X_1; X_2) + I(X; X_2|X_1) \end{aligned}$$

$$
\begin{aligned}
&+ I(X; X_0|X_1, X_2) \\
&= I(X; X_1) + I(X; X_2|X_1) + I(X; X_0|X_1, X_2) \\
&= I(X; X_0, X_1, X_2).
\end{aligned}
$$

Therefore, we have $(R_1, R_2, D_0, D_1, D_2) \in \mathcal{Q}$. Moreover, since

$$
\begin{aligned}
&R_1 = I(X; X_1) = R(D_1), \\
&R_1 + R_2 = I(X; X_0, X_1, X_2) = I(X; X_0, X_1) = R(D_0), \\
&D_i = \mathbb{P}(X \neq X_i), \quad i = 0, 1,
\end{aligned}
$$

it follows from F4) and Theorem 1 that $D_2^*(D_0, D_1) \leq \mathbb{P}(X \neq X_2)$, which further implies that $D_2^*(D_0, D_1) \leq \tilde{D}_2(D_0, D_1)$.

Now we proceed to show that $D_2^*(D_0, D_1) \geq \tilde{D}_2(D_0, D_1)$. In view of F4) and Theorem 1, we have $(R(D_1), R(D_0) - R(D_1), D_0, D_1, D_2^*(D_0, D_1)) \in \mathcal{Q}$. By the definition of $\mathcal{Q}$, there exist $p_{XX_0^{(j)}X_1^{(j)}X_2^{(j)}}$ and $\mu_j > 0$, $j = 1, 2, \cdots, r$, for some positive integer r such that

$$
I(X_1^{(j)}; X_2^{(j)}) = 0, \quad j = 1, 2, \cdots, r, \tag{5}
$$

$$
\sum_{j=1}^{r} \mu_j = 1, \tag{6}
$$

$$
\sum_{j=1}^{r} \mu_j I(X; X_1^{(j)}) \leq R(D_1), \tag{7}
$$

$$
\sum_{j=1}^{r} \mu_j I(X; X_0^{(j)}, X_1^{(j)}, X_2^{(j)}) \leq R(D_0), \tag{8}
$$

$$
\sum_{j=1}^{r} \mu_j \mathbb{P}(X \neq X_i^{(j)}) \leq D_i, \quad i = 0, 1, \tag{9}
$$

$$
\sum_{j=1}^{r} \mu_j \mathbb{P}(X \neq X_2^{(j)}) \leq D_2^*(D_0, D_1). \tag{10}
$$

It can be shown by leveraging F2) that

$$
I(X; X_i^{(j)}) = R(D_i), \quad i = 0, 1, \quad j = 1, 2, \cdots, r, \tag{11}
$$

$$
\mathbb{P}(X \neq X_i^{(j)}) = D_i, \quad i = 0, 1, \quad j = 1, 2, \cdots, r. \tag{12}
$$

By (8) and (11), we must have

$$I(X; X_0^{(j)}, X_1^{(j)}, X_2^{(j)}) = I(X; X_0^{(j)}, X_1^{(j)}) = I(X; X_0^{(j)}),$$
$$j = 1, 2, \cdots, r, \tag{13}$$

i.e., $I(X; X_1^{(j)}, X_2^{(j)}|X_0^{(j)}) = 0$, $j = 1, 2, \cdots, r$. In view of (11), (12), and (13), one can readily show by invoking F3) that $p_{XX_0^{(j)}X_1^{(j)}} = p_{XX_0X_1}$, $j = 1, 2, \cdots, r$. Therefore, it follows from (10) and the definition of $\tilde{D}_2(D_0, D_1)$ that

$$D_2^*(D_0, D_1) \geq \min_{j \in \{1,2,\cdots,r\}} \mathbb{P}(X \neq X_2^{(j)}) \geq \tilde{D}_2(D_0, D_1).$$

The proof is complete. ■

It is easy to see from the definition of $\mathcal{X}_0^+$ and $\mathcal{X}_1^+$ that $\mathcal{X}_i^+ = \{0, 1, \cdots, m_i\}$, $i = 0, 1$, for some positive integers m_0 and m_1; moreover, we have $\mathcal{X}_1^+ \subseteq \mathcal{X}_0^+$ (i.e., $m_1 \leq m_0$). Let $\mathcal{P}(\mathcal{X}_0^+)$ denote the set of probability distributions defined on $\mathcal{X}_0^+$.

Lemma 2: With no loss of generality one can assume $p_{X_2} \in \mathcal{P}(\mathcal{X}_0^+)$ in the definition of $\tilde{D}_2(D_0, D_1)$.

Proof: See Appendix A. ■

Lemma 3: For any X_2 such that $p_{X_2} \in \mathcal{P}(\mathcal{X}_0^+)$ and $I(X; X_2|X_0) = 0$, we have

$$\mathbb{P}(X \neq X_2) = D_0 + (1 - D_0 - \lambda_0)\mathbb{P}(X_0 \neq X_2).$$

Proof: See Appendix B. ■

Define $\bar{D}_2(D_0, D_1) = \min_{p_{X_2|X_0X_1}} \mathbb{P}(X_0 \neq X_2)$, where the minimization is over $p_{X_2|X_0X_1}$ subject to the constraints $p_{X_2} \in \mathcal{P}(\mathcal{X}_0^+)$ and $I(X_1; X_2) = 0$. It is obvious that $\bar{D}_2(D_0, D_1)$ is unaffected if the constraint $p_{X_2} \in \mathcal{P}(\mathcal{X}_0^+)$ is removed.

Lemma 4: $\tilde{D}_2(D_0, D_1) = D_0 + (1 - D_0 - \lambda_0)\bar{D}_2(D_0, D_1)$.

Proof: In view of the fact that $D_0 \leq D(R)|_{R=0} = 1 - p_X(0)$ (see F1)) and $\lambda_0 \leq \lambda_1 \leq p_X(1) \leq p_X(0)$ (see F3)), we have $1 - D_0 - \lambda_0 \geq 0$. Therefore, this result is a direct consequence of Lemma 2 and Lemma 3. ■

Lemma 5: $\bar{D}_2(D_0, D_1) = 1 - p_{X_0}(0) - \mathbb{P}(X_0 \neq X_1)$.

Proof: See Appendix C. ■

Combining Lemmas 1, 4, and 5, we have

$$
\begin{aligned}
&D_2^*(D_0, D_1)\\
&= D_0 + (1 - D_0 - \lambda_0)(1 - p_{X_0}(0) - \mathbb{P}(X_0 \neq X_1))\\
&= D_0 + (1 - D_0 - \lambda_0)\left(1 - p_{X_0}(0) - 1 + \frac{1 - D_1 - \lambda_0}{1 - D_0 - \lambda_0}\right) && (14)\\
&= 1 - (1 - D_0)p_{X_0}(0) - \lambda_0(1 - p_{X_0}(0)) + D_0 - D_1\\
&= 1 - p_X(0) + D_0 - D_1 && (15)\\
&= D(R)|_{R=0} + D_0 - D_1, && (16)
\end{aligned}
$$

where (14), (15), and (16) follow from (4), (3), and F1), respectively. This completes the proof of Theorem 2.

IV. Practical Robust Multiresolution Coding Scheme

We shall present a robust multiresolution coding scheme based on LDGM codes.

It is instructive to first explain the underlying ideas using random codes and joint typicality encoding.

1) Codebook Generation: Generate two random codebooks $\mathcal{C}_1 = \{x^n_{1,k_1}\}_{k_1=1}^{2^{n(I(X;X_1)+\epsilon_1)}}$ and $\mathcal{C}_2 = \{x^n_{2,k_2}\}_{k_2=1}^{2^{n(I(X,X_1;X_2)+\epsilon_2)}}$ according to $\prod_{l=1}^n p_{X_1}(\cdot)$ and $\prod_{l=1}^n p_{X_2}(\cdot)$, respectively. For each pair of codewords $x^n_{1,k_1} \in \mathcal{C}_1$ and $x^n_{2,k_2} \in \mathcal{C}_2$, generate a random codebook $\mathcal{C}_0(k_1, k_2) = \{x^n_{0,k_0,k_1,k_2}\}_{k_0=1}^{n(I(X;X_0|X_1,X_2)+\epsilon_0)}$ according to $\prod_{l=1}^n p_{X_0|X_1X_2}(\cdot|x_{1,k_1}(l), x_{2,k_2}(l))$.
2) Encoding: Given the source sequence x^n, first find k_1^* such that x^n_{1,k_1^*} is jointly strongly typical with x^n, then find k_2^* such that x^n_{2,k_2^*} is jointly strongly typical with (x^n, x^n_{1,k_1^*}), finally find k_0^* such that $x^n_{0,k_0^*,k_1^*,k_2^*}$ is jointly strongly typical with $(x^n, x^n_{1,k_1^*}, x^n_{2,k_2^*})$. The base layer of the multiresolution code contains the index k_1^* while the refinement layer contains k_2^* and k_0^*.
3) Decoding: The decoder uses x^n_{1,k_1^*} as the reconstruction if the base layer is received, uses x^n_{2,k_2^*} as the reconstruction if the refinement layer is received, and uses $x^n_{0,k_0^*,k_1^*,k_2^*}$ as the reconstruction if both layers are received.

Notice that the encoder can be regarded as the cascade of three encoders $\mathcal{E}_1$, $\mathcal{E}_2$, and $\mathcal{E}_3$ outputting indices k_1^*, k_2^*, and k_0^*, respectively. Following the above theoretical coding system

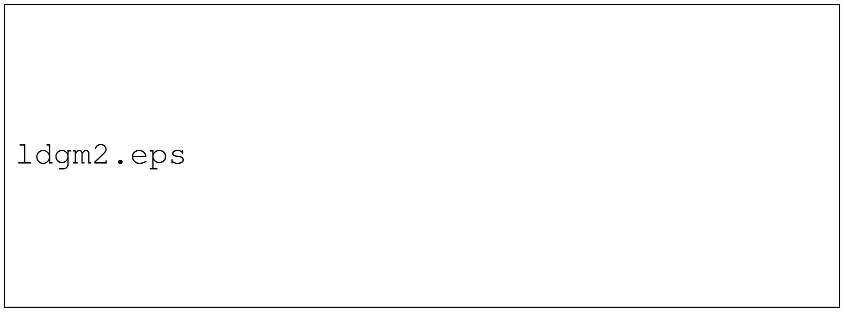

Fig. 1. Factor graph of a multilevel LDGM code.

we propose a practical coding scheme, which employs a multilevel LDGM code to generate the codebook, in conjunction with a message passing algorithm, at each component encoder. Multilevel LDGM codes were introduced in [19] to generate codebooks with codewords of non-uniform empirical distribution. They were shown to achieve the rate-distortion bound for single description coding of general finite alphabet sources, when used with the strong typicality encoding rule.

We mention that a three-stage successive coding scheme based on LDGM codes for the multiple description problem was proposed in [20]. However, the purpose of the three component encoders in [20] differs from our case. Precisely, in [20], the first stage generates a common part of the two descriptions, while the second and third stages produce the remaining part of description 1, respectively description 2.

Next we briefly review the design of multilevel LDGM codes under a uniform framework and clarify the specifics for each encoding stage. Notice that the design requirements for each codebook specify the output alphabet $\mathcal{Z}$ of the codewords, the size 2^{nR} of the codebook and n probability distributions $p_1(\cdot), \cdots, p_n(\cdot)$ over $\mathcal{Z}$. The requirement for the codebook is to be randomly generated according to $\prod_{l=1}^{n} p_l(\cdot)$. Precisely, we have $p_l(\cdot) = p_{X_1}(\cdot)$ for $\mathcal{E}_1$, $p_l(\cdot) = p_{X_2}(\cdot)$ for $\mathcal{E}_2$, and $p_l(\cdot) = p_{X_0|X_1X_2}(\cdot|x_{1,k_1^*}(l), x_{2,k_2^*}(l))$ for $\mathcal{E}_3$, $1 \le l \le n$.

To approximately satisfy the above requirement, we select an integer $\omega > 0$ and n mappings $\phi_l : \{0,1\}^\omega \to \mathcal{Z}$ such that $|\phi_l^{-1}(z)| \approx 2^\omega p_l(z)$ for all $1 \le l \le n$ and $z \in \mathcal{Z}$. Based on these mappings, the function $\Phi : \{0,1\}^{n\omega} \to \mathcal{Z}^n$ is defined as follows: the l-th symbol of $\Phi(c^{n\omega})$ equals $\phi_l(c(l), c(n+l), \cdots, c(n(\omega-1)+l))$ for all $c^{n\omega} \in \{0,1\}^{n\omega}$. Further, a low-density generator matrix G of dimension $n\omega \times m$, over the binary field $GF(2)$ is chosen, where $m = nR$. Then the codebook generated by the multilevel LDGM code is defined as

$$\mathcal{C} = \{z^n \in \mathcal{Z}^n | z^n = \Phi(Gv^m), v^m \in \{0,1\}^m\},$$

where the matrix multiplication is performed over $GF(2)$.

The multilevel LDGM code is associated with a factor graph as illustrated in Figure 1. The graph consists of n source nodes $\{S_1, \cdots, S_n\}$, corresponding to the sequence input to the encoder, m variable nodes $\{V_1, \cdots, V_m\}$, $n\omega$ check nodes $\{C_1, \cdots, C_{n\omega}\}$, and n network nodes $\{N_1, \cdots, N_n\}$. Each variable node V_k is associated with information bit $v(k)$ and is connected by an edge to every check node C_q such that $G(q,k) = 1$. Every check node C_q is assigned a bit value $c(q)$ computed as the modulo 2 summation of the bit values at adjacent variables nodes. Finally, each network node N_l is connected by an edge to check nodes $C_l, C_{l+n}, \cdots, C_{l+(\omega-1)n}$, and to the source node S_l. N_l is associated with the l-th symbol $z(l)$ of the codeword, computed by applying the mapping $\phi_l(\cdot)$ to the bit values at the adjacent check nodes. The construction of the mapping $\phi_l(\cdot)$ ensures that the marginal distribution of symbol $z(l)$ approximates $p_l(\cdot)$.

Notice that encoder $\mathcal{E}_3$ needs multiple conditional codebooks. However, by choosing a common value of the integer ω and a common low-density generator matrix G for all these codes, the associated graphs become identical. What differs from one codebook to another are only the functions $\phi_l(\cdot)$. Since the number of different mappings $\phi_l(\cdot)$ is small, the storage space needed at $\mathcal{E}_3$ is comparable with that for a single LDGM code.

Each encoder is associated with a pair of random variables Y and Z jointly distributed over the alphabets $\mathcal{Y}$ and $\mathcal{Z}$, respectively. The encoder requirement is, given the input sequence y^n, to select a codeword z^n jointly strongly typical with y^n. The input sequence y^n is the source sequence x^n combined with the codeword(s) selected by previous encoder(s), if any. Thus, the alphabet $\mathcal{Y}$ is the cartesian product of $\mathcal{X}$ and the codeword alphabets of previous encoder(s), if any. Precisely, for $\mathcal{E}_1$ we have $(Y,Z) = (X, X_1)$ and $y^n = x^n$. For $\mathcal{E}_2$ we have $Y = (X, X_1)$, $Z = X_2$, and $y(l) = (x(l), x_{1,k_1^*}(l))$, $1 \leq l \leq n$. Finally, for $\mathcal{E}_3$, we have $Y = (X, X_1, X_2)$, $Z = X_0$, and $y(l) = (x(l), x_{1,k_1^*}(l), x_{2,k_2^*}(l))$, $1 \leq l \leq n$.

As in prior work on LDGM-based coding [19]–[22], we use a message passing algorithm over the associated factor graph as a heuristic to solve the encoder problem. Our algorithm of choice is belief propagation with decimation. It proceeds in a series of rounds. Each round consists of a message passing phase where messages are transmitted between every adjacent nodes in a series of iterations, followed by a decimation phase where some variable nodes are fixed and removed from the factor graph. The algorithm stops when all variable nodes are fixed.

At each message passing iteration, every node A passes a message to each adjacent non-source node B. If A is not a source node then the message consists of two components: $M_{A\to B}(0)$ and $M_{A\to B}(1)$. If A is a source node, the message consists of $|\mathcal{Z}|$ components: $M_{A\to B}(z)$, $z \in \mathcal{Z}$.

$$
\begin{aligned}
M_{S_l\to N_l}(z) =& \exp(-\lambda(y(l), z)),\\
&\text{for all } z \in \mathcal{Z}, 1 \le l \le n,\\
M_{N_l\to C_{l+sn}}(b) =& \sum_{z\in\mathcal{Z}} M_{S_l\to N_l}(z) \sum_{\substack{b^\omega\in\{0,1\}^\omega\\ b(s+1)=b\\ \phi_l(b^\omega)=z}} \prod_{\substack{j=0\\ j\neq s}}^{\omega-1} M_{C_{l+jn}\to N_l}(b(j))\\
M_{C_{l+sn}\to N_l}(b) =& \frac{1}{2} + \frac{(-1)^b}{2} \prod_{k\in\mathcal{B}_v(l+sn)} (M_{V_k\to C_{l+sn}}(0) - M_{V_k\to C_{l+sn}}(1))\\
&\text{for all } b \in \{0,1\}, 1 \le l \le n, 0 \le s \le \omega-1,\\
M_{C_{l+sn}\to V_k}(b) =& \frac{1}{2} + \frac{(-1)^b}{2} (M_{N_l\to C_{l+sn}}(0) - M_{N_l\to C_{l+sn}}(1)) \prod_{i\in\mathcal{B}_v(l+sn)\setminus\{k\}} (M_{V_i\to C_{l+sn}}(0) - M_{V_i\to C_{l+sn}}(1))\\
M_{V_k\to C_{l+sn}}(b) =& \prod_{q\in\mathcal{A}_c(k)\setminus\{l+sn\}} M_{C_q\to V_k}(b)\\
&\text{for all } b \in \{0,1\}, k \in \mathcal{B}_v(l+sn), 1 \le l \le n, 0 \le s \le \omega-1.
\end{aligned}
$$

Fig. 2. Message passing equations. After applying these equations, the components of each message are normalized to sum up to 1.

During the first iteration in the first round, only the source nodes and check nodes pass messages, the messages sent by check nodes being $(0.5, 0.5)$. After that, at each iteration the schedule of message transmission is: $1)$ from network nodes and variable nodes to check nodes; $2)$ from source nodes and check nodes to their adjacent nodes. Every non-source node computes the message to pass along an edge using the messages received along its other adjacent edges at the previous iteration. The equations to calculate the messages are presented in Figure 2. We have denoted by $\mathcal{A}_c(k)$ the set of indices q such that C_q is adjacent to node V_k, and by $\mathcal{B}_v(q)$ the set of of indices k such that V_k is adjacent to C_q. The quantities $\lambda(y, z) \ge 0$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$,

P_i	$\mathcal{E}_1$	$\mathcal{E}_2$	$\mathcal{E}_3$
P_1	0.7	1.6	1.6
P_2	0.7	1.6	1.6
P_3	0.7	4	4

TABLE I
VALUES OF PARAMETER δ USED IN SIMULATIONS.

used in defining the source messages, are parameters tuned during code design phase based on simulations. As a guideline for selecting these parameters we use the intuition that $\lambda(y, z)$ should be (roughly speaking) inverse proportional to $p_{Z|Y}(z|y)$.

The message passing phase ends when all the messages $M_{V_k \to C_q}(0)$ converge or a maximum number of iterations, typically 100, is reached. At the decimation phase, the marginal distributions at variable nodes are computed and the variable nodes whose bias is greater than a threshold $\eta > 0$ are fixed. If no such variable node exists then the one with highest bias is fixed. After that the fixed variables are removed from the graph. The check node whose all adjacent variable nodes are fixed will send to the adjacent network node the following message: $M(0) = \frac{1}{\gamma}((1-c)\exp(\delta) + c\exp(-\delta))$ and $M(1) = \frac{1}{\gamma}(c\exp(\delta) + (1-c)\exp(-\delta))$, where c equals to the modulo-2 summation of the values of all adjacent variable nodes. Finally, after all variable nodes are fixed, the output codeword is determined on network nodes by mapping the connected check nodes values.

Notice that the design of the proposed scheme does not depend on the distortion measure, but only on the joint distribution $p_{XX_1X_2X_0}$. Therefore, although we have tested this scheme only for Hamming distortion measure, we hypothesize that it is applicable to any distortion function. To support this claim it is worth mentioning that the simulation results in [19] show very good performance of multilevel LDGM codes in the case of single description source coding with a non-Hamming distortion measure.

V. EXPERIMENTAL RESULTS

We have tested the proposed robust multiresolution coding scheme for the binary uniform source with Hamming distortion measure, targeting three distortion triples $(D_0, D_1, D_2^*(D_1, D_0))$: $P_1 = (0.1, 0.3, 0.3)$, $P_2 = (0.05, 0.3, 0.25)$ and $P_3 = (0, 0.3, 0.2)$. In all three cases, $R_1 = R(D_1)$

x, x_1, x_2	P_1	P_2	P_3
$0,0,0$	0	0.6	0
$0,0,1$	0	0	0
$0,1,0$	0	0	0
$0,1,1$	2.0	2.2	8
$1,0,0$	2.8	2.8	8
$1,0,1$	0	0	0
$1,1,0$	2.8	2.8	9
$1,1,1$	0	0	0

TABLE II
VALUES OF PARAMETERS $\lambda((x, x_1), x_2)$ AT ENCODER $\mathcal{E}_2$.

x, x_1, x_2, x_0	P_1	P_2	P_3
$0,0,1,0$	0	0	0
$0,0,1,1$	2.8	2.8	10
$1,0,1,0$	1.8	2.6	10
$1,0,1,1$	0	0	0

TABLE III
VALUES OF PARAMETERS $\lambda((x, x_1, x_2), x_0)$ AT ENCODER $\mathcal{E}_3$.

and $R_2 = R(D_0) - R(D_1)$ hold. The degree distributions of the LDGM codes used in our tests are taken from the website (http://lthcwww.epfl.ch.research/ldpcopt) or obtained by implementing the algorithm in [23]. We use damping as in [20], [22] in our message passing algorithm, if the messages do not converge after 30 iterations.

(R_1, R_2)	$(D_0, D_1, D_2^*(D_1, D_0))$	$\hat{D}_0$	$\hat{D}_1$	$\hat{D}_2$
$(0.1187, 0.4122)$	$(0.10, 0.30, 0.30)$	0.109	0.308	0.304
$(0.1187, 0.5955)$	$(0.05, 0.30, 0.25)$	0.066	0.308	0.258
$(0.1187, 0.8813)$	$(0.00, 0.30, 0.20)$	0.012	0.309	0.206

TABLE IV
TEST RESULTS: $(D_0, D_1, D_2^*(D_1, D_0))$ IS A TARGET DISTORTION TRIPLE; R_1 AND R_2 ARE THE RATES OF THE BASE, RESPECTIVELY, REFINEMENT LAYER; $\hat{D}_0$, $\hat{D}_1$, AND $\hat{D}_2$ ARE THE EMPIRICAL DISTORTIONS.

The length of the input sequences in our tests is $n = 10,000$. We use $\eta = 0.9$ and $\omega = 4$. The values of parameter δ are listed in Table I. To define the source messages for $\mathcal{E}_1$, we use $\lambda(0,0) = \lambda(1,1) = 0$ and $\lambda(0,1) = \lambda(1,0) = 0.7$. The values of $\lambda(y,z)$ for encoders $\mathcal{E}_2$, and $\mathcal{E}_3$ are presented in Tables II and III, respectively. It is worth mentioning that for all three cases P_1, P_2, P_3, variable X_0 is a deterministic function of X_1 and X_2 when $(X_1, X_2) \neq (0,1)$. Thus, at encoder $\mathcal{E}_3$, for l such that $(x_{1,k_1^*}(l), x_{2,k_2^*}(l)) \neq (0,1)$, the network node N_l always sends the uniform message $(0.5, 0.5)$, irrespective of the message received from S_l. Therefore, we set $\lambda((x, x_1, x_2), x_0) = 0$ for all binary quadruples (x, x_1, x_2, x_0) with $(x_1, x_2) \neq (0,1)$.

Table IV presents the experimental results. The first column contains the rates R_1, R_2 of the base, respectively, refinement layer. The second column contains the target distortion triple (D_0, D_1, D_2). The remaining three columns present the empirical values of the three distortions, respectively, averaged over 100 runs. As observed from Table IV, the distortions are very close to the theoretical limits.

VI. Conclusion

This work derives an explicit expression of the minimum distortion achievable by the refinement layer of a multiresolution code for arbitrary finite alphabet sources with Hamming distortion measure. A practical robust multiresolution coding scheme based on LDGM codes is proposed, which shows promising performance.

Appendix A
Proof of Lemma 2

For any X_2 such that $I(X; X_1, X_2|X_0) = 0$ and $I(X_1; X_2) = 0$, define $\tilde{X}_2 = X_2$ if $X_2 \in \mathcal{X}_0^+$ and $\tilde{X}_2 = 0$ if $X_2 \notin \mathcal{X}_0^+$. It is clear that $I(X; X_1, \tilde{X}_2|X_0) = 0$ and $I(X_1; \tilde{X}_2) = 0$. Note that for $x_0 \in \mathcal{X}_0^+$

$$
\begin{aligned}
p_{X|X_0}(0|x_0) &\geq \min(1 - D_0, \lambda_0) \\
&\geq \min(1 - D(R)|_{R=0}, \lambda_0) \\
&= \min(p_X(0), \lambda_0) \\
&= \lambda_0
\end{aligned}
$$

$$\geq \max_{x \notin \mathcal{X}_0^+} p_{X|X_0}(x|x_0),$$

where the first equality follows from F1), and the second equality follows from the fact that $\lambda_0 \leq \lambda_1 \leq p_X(1) \leq p_X(0)$ (see F3)). Therefore, we have

$$\begin{aligned}
\mathbb{P}(X = X_2) &= \sum_{x_0 \in \mathcal{X}_0^+} \sum_{x_2 \in \mathcal{X}} p_{XX_0X_2}(x_2, x_0, x_2) \\
&= \sum_{x_0 \in \mathcal{X}_0^+} \sum_{x_2 \in \mathcal{X}} p_{X|X_0}(x_2|x_0) p_{X_0X_2}(x_0, x_2) \\
&= \sum_{x_0 \in \mathcal{X}_0^+} \Bigg(\sum_{x_2 \in \mathcal{X}_0^+} p_{X|X_0}(x_2|x_0) p_{X_0X_2}(x_0, x_2) \\
&\qquad + \sum_{x_2 \notin \mathcal{X}_0^+} p_{X|X_0}(x_2|x_0) p_{X_0X_2}(x_0, x_2) \Bigg) \\
&\leq \sum_{x_0 \in \mathcal{X}_0^+} \Bigg(\sum_{x_2 \in \mathcal{X}_0^+} p_{X|X_0}(x_2|x_0) p_{X_0X_2}(x_0, x_2) \\
&\qquad + p_{X|X_0}(0|x_0) \sum_{x_2 \notin \mathcal{X}_0^+} p_{X_0X_2}(x_0, x_2) \Bigg) \\
&= \sum_{x_0 \in \mathcal{X}_0^+} \Bigg(p_{X|X_0}(0|x_0) \Bigg(p_{X_0X_2}(x_0, 0) \\
&\qquad + \sum_{x_2 \notin \mathcal{X}_0^+} p_{X_0X_2}(x_0, x_2) \Bigg) \\
&\qquad + \sum_{x_2 \in \mathcal{X}_0^+, x_2 \neq 0} p_{X|X_0}(x_2|x_0) p_{X_0X_2}(x_0, x_2) \Bigg) \\
&= \sum_{x_0 \in \mathcal{X}_0^+} \sum_{\tilde{x}_2 \in \mathcal{X}_0^+} p_{X|X_0}(\tilde{x}_2|x_0) p_{X_0\tilde{X}_2}(x_0, \tilde{x}_2) \\
&= \mathbb{P}(X = \tilde{X}_2).
\end{aligned}$$

Appendix B
Proof of Lemma 3

Note that

$$\mathbb{P}(X = X_2)$$

$$
\begin{aligned}
&= \sum_{x_0 \in \mathcal{X}_0^+} \sum_{x_2 \in \mathcal{X}_0^+} p_{XX_0X_2}(x_2, x_0, x_2) \\
&= \sum_{x_0 \in \mathcal{X}_0^+} \sum_{x_2 \in \mathcal{X}_0^+} p_{X|X_0}(x_2|x_0) p_{X_0X_2}(x_0, x_2) \\
&= \sum_{x_0 \in \mathcal{X}_0^+} \Bigg(p_{X|X_0}(x_0|x_0) p_{X_0X_2}(x_0, x_0) \\
&\qquad + \sum_{x_2 \in \mathcal{X}_0^+, x_2 \neq x_0} p_{X|X_0}(x_2|x_0) p_{X_0X_2}(x_0, x_2) \Bigg) \\
&= \sum_{x_0 \in \mathcal{X}_0^+} \Bigg((1 - D_0) p_{X_0X_2}(x_0, x_0) \\
&\qquad + \sum_{x_2 \in \mathcal{X}_0^+, x_2 \neq x_0} \lambda_0 p_{X_0X_2}(x_0, x_2) \Bigg) \qquad (17) \\
&= (1 - D_0)\mathbb{P}(X_0 = X_2) + \lambda_0 \mathbb{P}(X_0 \neq X_2),
\end{aligned}
$$

where (17) is due to (3). Therefore, we have

$$
\begin{aligned}
\mathbb{P}(X \neq X_2) &= 1 - \mathbb{P}(X = X_2) \\
&= 1 - (1 - D_0)\mathbb{P}(X_0 = X_2) - \lambda_0 \mathbb{P}(X_0 \neq X_2) \\
&= D_0 + (1 - D_0 - \lambda_0)\mathbb{P}(X_0 \neq X_2).
\end{aligned}
$$

APPENDIX C
PROOF OF LEMMA 5

It is easy to see that

$$
\begin{aligned}
&\bar{D}_2(D_0, D_1) \\
&= \min_{p_{X_0X_1X_2}} 1 - \sum_{x_0 \in \mathcal{X}_0^+} \sum_{x_1 \in \mathcal{X}_1^+} p_{X_0X_1X_2}(x_0, x_1, x_0) \qquad (18)
\end{aligned}
$$

subject to the constraints

$$
\begin{aligned}
&- p_{X_0X_1X_2}(x_0, x_1, x_2) \leq 0, \\
&\qquad\qquad x_0 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+, x_2 \in \mathcal{X}_0^+, \qquad (19)
\end{aligned}
$$

$$\sum_{x_2 \in \mathcal{X}_0^+} p_{X_0X_1X_2}(x_0, x_1, x_2) = p_{X_0X_1}(x_0, x_1),$$

$$x_0 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+, \tag{20}$$

$$p_{X_1}(x_1) \sum_{x_0 \in \mathcal{X}_0^+} p_{X_0X_1X_2}(x_0, 0, x_2)$$
$$- p_{X_1}(0) \sum_{x_0 \in \mathcal{X}_0^+} p_{X_0X_1X_2}(x_0, x_1, x_2) = 0,$$
$$x_1 \in \mathcal{X}_1^+ \setminus \{0\}, x_2 = \mathcal{X}_0^+ \setminus \{0\}, \tag{21}$$

where (19) and (20) are due to the fact that $p_{X_0X_1X_2}$ is a probability distribution and that $p_{X_0X_1}$ is fixed while (21) is due to the independence of X_1 and X_2.

Since (18) is a linear programming problem, the Karush-Kuhn-Tucker conditions are sufficient for global optimality. Now introduce Lagrangian multipliers $\mu = (\mu_{x_0,x_1,x_2})_{x_0 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+, x_2 \in \mathcal{X}_0^+}$, $\alpha = (\alpha_{x_0,x_1})_{x_0 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+}$, and $\beta = (\beta_{x_1,x_2})_{x_1 \in \mathcal{X}_1^+ \setminus \{0\}, x_2 \in \mathcal{X}_0^+ \setminus \{0\}}$ for (19), (20), and (21), respectively. Define

$$\begin{aligned}
&G(p_{X_0X_1X_2}, \mu, \alpha, \beta) \\
&= 1 - \sum_{x_0 \in \mathcal{X}_0^+} \sum_{x_1 \in \mathcal{X}_1^+} p_{X_0X_1X_2}(x_0, x_1, x_0) \\
&\quad - \sum_{x_0 \in \mathcal{X}_0^+} \sum_{x_1 \in \mathcal{X}_1^+} \sum_{x_0 \in \mathcal{X}_0^+} \mu_{x_0,x_1,x_2} p_{X_0X_1X_2}(x_0, x_1, x_2) \\
&\quad + \sum_{x_0 \in \mathcal{X}_0^+} \sum_{x_1 \in \mathcal{X}_1^+} \alpha_{x_0,x_1} \sum_{x_2 \in \mathcal{X}_0^+} p_{X_0X_1X_2}(x_0, x_1, x_2) \\
&\quad + \sum_{x_1 \in \mathcal{X}_1^+ \setminus \{0\}} \sum_{x_2 \in \mathcal{X}_0^+ \setminus \{0\}} \beta_{x_1,x_2} \sum_{x_0 \in \mathcal{X}_0^+} (p_{X_1}(x_1) \\
&\qquad \times p_{X_0X_1X_2}(x_0, 0, x_2) - p_{X_1}(0) p_{X_0X_1X_2}(x_0, x_1, x_2)).
\end{aligned}$$

The Karush-Kuhn-Tucker conditions are given by

$$\frac{\partial G(p_{X_0X_1X_2}, \mu, \alpha, \beta)}{\partial p_{X_0X_1X_2}(x_0, x_1, x_2)} = 0, \quad x_0 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+, x_2 \in \mathcal{X}_0^+, \tag{22}$$

$$- p_{X_0X_1X_2}(x_0, x_1, x_2) \leq 0, \quad x_0 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+, x_2 \in \mathcal{X}_0^+,$$

$$\sum_{x_2\in\mathcal{X}_0^+} p_{X_0X_1X_2}(x_0,x_1,x_2) = p_{X_0X_1}(x_0,x_1),$$

$$x_0 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+,$$

$$p_{X_1}(x_1)\sum_{x_0\in\mathcal{X}_0^+} p_{X_0X_1X_2}(x_0,0,x_2)$$

$$-\, p_{X_1}(0)\sum_{x_0\in\mathcal{X}_0^+} p_{X_0X_1X_2}(x_0,x_1,x_2) = 0,$$

$$x_1 \in \mathcal{X}_1^+ \setminus \{0\}, x_2 = \mathcal{X}_0^+ \setminus \{0\},$$

$$\mu_{x_0,x_1,x_2} \geq 0, \quad x_0 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+, x_2 \in \mathcal{X}_0^+,$$

$$\mu_{x_0,x_1,x_2} p_{X_0X_1X_2}(x_0,x_1,x_2) = 0,$$

$$x_0 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+, x_2 \in \mathcal{X}_0^+,$$

where (22) is equivalent to

$$d_H(x_0,x_2) - 1 - \mu_{x_0,x_1,x_2} + \alpha_{x_0,x_1}$$

$$+ \sum_{x_1'\in\mathcal{X}_1^+\setminus\{0\}} \beta_{x_1',x_2} p_{X_1}(x_1') = 0,$$

$$x_0 \in \mathcal{X}_0^+, x_1 = 0, x_2 \in \mathcal{X}_1^+ \setminus \{0\},$$

$$d_H(x_0,x_2) - 1 - \mu_{x_0,x_1,x_2} + \alpha_{x_0,x_1} - \beta_{x_1,x_2} p_{X_1}(0) = 0,$$

$$x_0 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+ \setminus \{0\}, x_2 \in \mathcal{X}_0^+ \setminus \{0\},$$

$$d_H(x_0,x_2) - 1 - \mu_{x_0,x_1,x_2} + \alpha_{x_0,x_1} = 0,$$

$$x_0 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+, x_2 = 0.$$

Let

$$p_{X_0X_1X_2^*}(x_0,x_1,x_2)$$

$$= \begin{cases} p_{X_0X_1}(x_1,x_1) - p_{X_0X_1}(0,x_1), & x_0 = x_1 \neq x_2 = 0 \\ p_{X_0X_1}(x_0,x_1), & x_0 = x_2 \neq x_1 \\ p_{X_0X_1}(0,x_1), & x_0 = x_1 = x_2 \\ 0, & \text{otherwise} \end{cases}$$

for $x_0 \in \mathcal{X}_0^+$, $x_1 \in \mathcal{X}_1^+$, and $x_2 \in \mathcal{X}_0^+$. Let

$$
\begin{aligned}
&\mu^*_{x_0,x_1,x_2} = 0, \quad x_0 = x_2 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+, \\
&\mu^*_{x_0,x_1,x_2} = 0, \quad x_0 = x_1 \in \mathcal{X}_1^+ \setminus \{0\}, x_2 = 0, \\
&\mu^*_{x_0,x_1,x_2} = \alpha^*_{x_0,x_1}, \\
&\qquad\qquad x_0 \neq x_1, x_2 = 0, x_0 \in \mathcal{X}_0^+ \setminus \{0\}, x_1 \in \mathcal{X}_1^+, \\
&\mu^*_{x_0,x_1,x_2} = \alpha^*_{x_0,x_1} + \sum_{x_1' \in \mathcal{X}_1^+ \setminus \{0\}} \beta^*_{x_1',x_2} p_{X_1}(x_1'), \\
&\qquad\qquad x_0 \neq x_2, x_1 = 0, x_0 \in \mathcal{X}_0^+, x_2 \in \mathcal{X}_0^+ \setminus \{0\}, \\
&\mu^*_{x_0,x_1,x_2} = \alpha^*_{x_0,x_1} - \beta^*_{x_1,x_2} p_{X_1}(0), \\
&\qquad x_0 \neq x_2, x_0 \in \mathcal{X}_0^+, x_1 \in \mathcal{X}_1^+ \setminus \{0\}, x_2 \in \mathcal{X}_0^+ \setminus \{0\}, \\
&\alpha^*_{x_0,x_1} = 0, \quad x_0 = x_1 \in \mathcal{X}_1^+ \setminus \{0\}, \\
&\alpha^*_{x_0,x_1} = 1, \quad x_0 = 0, x_1 \in \mathcal{X}_1^+, \\
&\alpha^*_{x_0,x_1} = 1, \quad x_0 \neq x_1, x_0 \in \mathcal{X}_0^+ \setminus \{0\}, x_1 \in \mathcal{X}_1^+ \setminus \{0\}, \\
&\alpha^*_{x_0,x_1} = \frac{1 - \sum_{x_1' \neq x_0, x_1' \in \mathcal{X}_1^+ \setminus \{0\}} p_{X_1}(x_1')}{p_{X_1}(0)}, \\
&\qquad\qquad\qquad\qquad x_0 \in \mathcal{X}_0^+ \setminus \{0\}, x_1 = 0, \\
&\beta^*_{x_1,x_2} = \frac{\alpha^*_{x_2,x_1} - 1}{p_{X_1}(0)}, \quad x_1 \in \mathcal{X}_1^+ \setminus \{0\}, x_2 \in \mathcal{X}_0^+ \setminus \{0\}.
\end{aligned}
$$

It can be verified that the Karush-Kuhn-Tucker conditions are satisfied by the constructed $(p_{X_0X_1X_2^*}, \mu^*, \alpha^*, \beta^*)$. Moreover, note that

$$
\begin{aligned}
&1 - \sum_{x_0 \in \mathcal{X}_0^+} \sum_{x_1 \in \mathcal{X}_1^+} p_{X_0X_1X_2^*}(x_0, x_1, x_0) \\
&= 1 - \sum_{x_0 \in \mathcal{X}_1^+} \sum_{x_1 \in \mathcal{X}_1^+} p_{X_0X_1X_2^*}(x_0, x_1, x_0) \\
&\quad - \sum_{x_0 \in \mathcal{X}_0^+ \setminus \mathcal{X}_1^+} \sum_{x_1 \in \mathcal{X}_1^+} p_{X_0X_1X_2^*}(x_0, x_1, x_0) \\
&= 1 - \sum_{x_0 \in \mathcal{X}_1^+} \Bigg(p_{X_0X_1X_2^*}(x_0, x_0, x_0)
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \sum_{x_1 \neq x_0} p_{X_0X_1X_2^*}(x_0, x_1, x_0)\Bigg) \\
&\quad - \sum_{x_0 \in \mathcal{X}_0^+ \setminus \mathcal{X}_1^+} \sum_{x_1 \in \mathcal{X}_1^+} p_{X_0X_1X_2^*}(x_0, x_1, x_0) \\
&= 1 - \sum_{x_0 \in \mathcal{X}_1^+} \Bigg(p_{X_0X_1}(0, x_0) + \sum_{x_1 \neq x_0} p_{X_0X_1}(x_0, x_1) \Bigg) \\
&\quad - \sum_{x_0 \in \mathcal{X}_0^+ \setminus \mathcal{X}_1^+} \sum_{x_1 \in \mathcal{X}_1^+} p_{X_0X_1}(x_0, x_1) \\
&= 1 - p_{X_0}(0) - \sum_{x_0 \in \mathcal{X}_1^+} \sum_{x_1 \neq x_0} p_{X_0X_1}(x_0, x_1) \\
&\quad - \sum_{x_0 \in \mathcal{X}_0^+ \setminus \mathcal{X}_1^+} \sum_{x_1 \in \mathcal{X}_1^+} p_{X_0X_1}(x_0, x_1) \\
&= 1 - p_{X_0}(0) - \mathbb{P}(X_0 \neq X_1).
\end{aligned}
$$

The proof is complete.

ACKNOWLEDGMENT

The work of Jun Chen was supported by the Early Research Award from the Province of Ontario and by the Natural Science and Engineering Research Council (NSERC) of Canada under a discovery grant. The work of Jia Wang was supported by the NSFC under Grant 60802020 and 973 Program (2010CB731401, 2010CB731406).

REFERENCES

[1] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Signal Process.*, vol. 41, pp. 3445–3462, Dec. 1993.

[2] H. Brunk and N. Farvardin, "Fixed-rate successively refinable scalar quantizers," *Proc. DCC'96*, Snowbird, Utah, Mar. 1996, pp. 250-259.

[3] A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 243-250, Jun. 1996.

[4] H. Brunk and N. Farvardin, "Embedded entropy-constrained trellis coded quantization," *Proceedings of International Symposium on Information Theory*, Cambridge, MA, p. 274, Aug. 1998.

[5] H. Jafarkhani and V. Tarokh, "Design of successively refinable trellis-coded quantizers," *IEEE Trans. on Inf. Theory*, vol. 45, pp. 1490-1497, Jul. 1999.

[6] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.*, vol. 9, pp. 1158-1170, Jul. 2000.

[7] S. Dumitrescu and X. Wu, "Algorithms for optimal multi-resolution quantization," *J. Algorithms*, 50(2004), pp. 1-22.

[8] M. Effros and D. Dugatkin, "Multiresolution vector quantization," *IEEE Trans. Inf. Theory*, vol. 50, pp. 3130-3145, Dec. 2004.

[9] A. El Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, pp. 851-857, Nov. 1982.

[10] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 33, pp. 502-521, Jul. 1987.

[11] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2106-2114, Sep. 2003.

[12] B. Rimoldi, "Successive refinement of information: characterization of the achievable rates," *IEEE Trans. Inf. Theory*, vol. 40, pp. 253-259, Jan. 1994.

[13] J. Wang, J. Chen, L. Zhao, P. Cuff, and H. Permuter, "A random variable substitution lemma with applications to multiple description coding," preprint. Available: http://arxiv.org/abs/0909.3135.

[14] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inf. Theory*, vol. 31, pp. 721-726, Nov. 1985.

[15] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, pp. 269-275, Mar. 1991.

[16] V. Erokhin, "Epsilon-entropy of a discrete random variable," *Theory of Probability and Applications*, vol. 3, pp. 97-100, 1958, English translation.

[17] J. T. Pinkston, "An application of rate-distortion theory to a converse to the coding theorem," *IEEE Trans. Inf. Theory*, vol. IT-15, pp. 66-71, Jan. 1969.

[18] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909-1921, Dec. 1980.

[19] Z. Sun, M. Shao, J. Chen, K. M. Wong, and X. Wu, "Achieving the rate-distortion bound with low-density generator matrix codes," *IEEE Trans. Commun.*, vol. 58, pp. 1643-1653, Jun. 2010.

[20] Y. Zhang, S. Dumitrescu, J. Chen, and Z. Sun, "LDGM-based multiple description coding with Hamming distortion measure," preprint.

[21] M. Wainwright and E. Maneva, "Lossy source encoding via message-passing and decimation over generalized codewords of ldgm codes," *Proceedings of International Symposium on Information Theory*, pp. 1493-1497, Sep. 4-9, 2005.

[22] T. Filler and J. Friedrich, "Binary quantization using belief propagation with decimation over factor graphs of ldgm codes," *Proc. Allerton Conf. Comm., Control, and Computing*, Monticello, IL, Sep. 2007.

[23] S.-Y. Chung, G. D. Forney, T. J. Richardson, and R. Urbanke, "On the design of low-desity parity check codes within 0.0045 dB of the Shannon limit," *IEEE Communications Letters*, vol. 5, pp. 58-60, Feb. 2001.