On the Generalization of Natural Type Selection to Multiple Description Coding

Yuhua Fan, Jia Wang, Jun Sun, *Member, IEEE,* and Jun Chen, *Member, IEEE*

Abstract—Natural type selection was originally proposed by Zamir and Rose for universal single description coding. In this paper, we generalize this principle to universal multiple description coding (MDC). Two schemes based on random codebooks are proposed: one is of fixed distortion and the other is of fixed weight. Their operational sum-rate-distortion functions are derived, which coincide with the EGC (El Gamal-Cover) sum-rate bound if the parameters of the schemes are optimized. It is also shown that in both schemes the joint type of reconstruction codewords can be used to improve the rate-distortion (R-D) performance. Based on our theoretical results, a practical universal scheme is proposed by leveraging the MDC methods based on low-density generator matrix (LDGM) codes. The performance of this scheme is compared experimentally with the EGC bound, which shows its effectiveness.

Index Terms—Data compression, natural type selection, universal multiple description coding.

I. INTRODUCTION

WE study the problem of universal multiple description coding (UMDC), in which no knowledge of the source statistics is assumed, while good operational rate-distortion (R-D) performance is still expected asymptotically. Although significant effort has been devoted to universal lossless/lossy data compression, research on UMDC is still very limited. It is worth noting that the techniques used in designing universal lossy single description codes are not directly applicable to UMDC due to the fact that there are multiple decoders in MDC and each decoder decodes a different set of descriptions.

As pointed out in [3], since a lossy decoder only needs to produce an approximate replica of the source, the aim of its codebook adaptation is not to match itself to the source statistics but to find the "right" set of type classes. Furthermore, a mechanism of natural type selection is proposed in [3] for lossy single description coding. It is shown that, given

a codebook, codewords of certain type will be favored during the coding process, based on which a better codebook can be generated. The theory of natural type selection illuminates the connection among universal source coding, Blahut-Arimoto algorithm, and the rate distortion theory.

In this paper, we generalize the concept of natural type selection to two-channel MDC, in which the source is encoded into two descriptions and each description is either received entirely or lost completely. We propose two schemes based on random codebooks: one is of fixed-distortion and the other is of fixed-weight. For both schemes, we analyze the relationship between sum-rate and distortion, and establish results that are parallel to those given in [3]. It is shown that the R-D performance of the proposed schemes improves progressively as coding proceeds. Inspired by these theoretical results, we then propose a practical universal scheme by leveraging the MDC methods based on low-density generator matrix (LDGM) codes [12]. Experimental results show that the R-D performance of this scheme approaches the El Gamal-Cover (EGC) bound.

Recent work on UMDC includes [1], [2]. In [1], MDC of discrete ergodic sources is studied. The relationship between block MDC and stationary MDC is revealed, based on which a universal coding algorithm is proposed. Given an n-dimensional source vector $\mathbf{x} \in \mathcal{X}^n$, a triple of n-dimensional vectors $(\mathbf{y}, \mathbf{z}, \mathbf{w})$ is first found by solving an optimization problem with simulated annealing, after which $\mathbf{y}$ and $\mathbf{z}$ are compressed separately by the Lempel-Ziv algorithm, and $\mathbf{w}$ is compressed losslessly conditioned on both $\mathbf{y}$ and $\mathbf{z}$. In [2], a universal scheme is presented for multiple description scalar quantization under the assumption that the cells of side quantizers are Voronoi regions. It is also shown that the complexity of the scheme can be greatly reduced by exploiting the structure of quantizers.

The rest of the paper is organized as follows. Section II introduces the notations and conventions used throughout the paper. Section III contains the definition of the EGC region along with a discussion of its properties. Section IV presents two UMDC schemes based on random codebooks, supported by several theoretical results. A practical UMDC scheme is proposed in Section V, and the experimental results are given in Section VI. We conclude the paper in Section VII.

II. NOTATIONS AND CONVENTIONS

We use uppercase letters (e.g., U) to denote random variables, use their lowercase counterparts (e.g., u) to denote their realizations, and use calligraphic letters (e.g., $\mathcal{U}$) to denote the corresponding alphabets. Lowercase boldface letters (e.g.,

u) represent deterministic vectors while uppercase boldface letters (e.g., **U**) represent random vectors or matrices. Given an n-dimensional vector $\mathbf{u} \in \mathcal{U}^n$, its i-th element is denoted by $\mathbf{u}[i]$, and its type is denoted by $t_{\mathbf{u}}(u), u \in \mathcal{U}$. We use $t_{\mathbf{uv}}(\cdot)$ for the joint type of $\mathbf{u}$ and $\mathbf{v}$. Given a matrix $\mathbf{A}$, we use $\mathbf{A}[i][j]$ to denote its (i,j) entry.

We assume that the source is an i.i.d. process generated by a generic random variable X. We denote the source alphabet by $\mathcal{X}$, and the probability mass function (pmf) of X by $p_X(x), x \in \mathcal{X}$. Given two alphabets $\mathcal{U}$ and $\mathcal{V}$, the distortion between $u \in \mathcal{U}$ and $v \in \mathcal{V}$ is represented by $d_{\mathcal{U},\mathcal{V}}(u,v)$. For n-dimensional vectors $\mathbf{u} \in \mathcal{U}^n$ and $\mathbf{v} \in \mathcal{V}^n$, we define $d_{\mathcal{U},\mathcal{V}}(\mathbf{u},\mathbf{v}) \triangleq \frac{1}{n}\sum_{i=1}^{i=n} d_{\mathcal{U},\mathcal{V}}(\mathbf{u}[i],\mathbf{v}[i])$. For simplicity, $d_{\mathcal{U},\mathcal{V}}(\cdot,\cdot)$ is abbreviated to $d(\cdot,\cdot)$ if the underlying alphabets can be inferred from the context. The logarithm function is to base 2 throughout the paper.

III. SOME BASIC RESULTS ON THE EGC REGION

In two-channel MDC, a quintuple $(R_1, R_2, D_0, D_1, D_2)$ is achievable if for every $\epsilon > 0$, there exist two encoding functions $\psi_1 : \mathcal{X}^n \to \{1, 2, \cdots, M_1\}$, $\psi_2 : \mathcal{X}^n \to \{1, 2, \cdots, M_2\}$, and three decoding functions $\phi_1 : \{1, 2, \cdots, M_1\} \to \mathcal{X}_1^n$, $\phi_2 : \{1, 2, \cdots, M_2\} \to \mathcal{X}_2^n$, $\phi_0 : \{1, 2, \cdots, M_1\} \times \{1, 2, \cdots, M_2\} \to \mathcal{X}_0^n$, such that

$$\frac{1}{n}\log(M_i) \le R_i + \epsilon,$$
$$E\left[d\left(\mathbf{X}, \phi_i(\psi_i(\mathbf{X}))\right)\right] \le D_i + \epsilon, \ i = 1, 2,$$
$$E\left[d\left(\mathbf{X}, \phi_0(\psi_1(\mathbf{X}), \psi_2(\mathbf{X}))\right)\right] \le D_0 + \epsilon,$$

where $\mathbf{X}$ is an n-dimensional random source vector.

The EGC theorem [6] states that for two-channel MDC a quintuple $(R_1, R_2, D_0, D_1, D_2)$ is achievable if $(R_1, R_2, D_0, D_1, D_2) \in \text{conv}(\mathscr{A})$, where

$$
\begin{aligned}
\mathscr{A} \triangleq \{&(R_0, R_1, D_0, D_1, D_2) : \exists \text{ random variables } X_i \in \mathcal{X}_i, \\
&i = 0, 1, 2, \text{ such that } R_i \ge I(X; X_i), i = 1, 2, \\
&R_1 + R_2 \ge I(X; X_0, X_1, X_2) + I(X_1; X_2), \\
&E[d(X, X_i)] \le D_i, i = 0, 1, 2\},
\end{aligned}
\tag{1}
$$

and $\text{conv}(\mathscr{A})$ denotes the convex hull of $\mathscr{A}$. Define

$$
\begin{aligned}
\widetilde{\mathscr{A}_{\text{sum}}} \triangleq \{&(R, D_0, D_1, D_2) : \exists (R_1, R_2, D_0, D_1, D_2) \in \\
&\text{conv}(\mathscr{A}) \text{ such that } R \ge R_1 + R_2\}, \\
\mathscr{A}_{\text{sum}} \triangleq \{&(R, D_0, D_1, D_2) : \exists (R_1, R_2, D_0, D_1, D_2) \in \mathscr{A} \\
&\text{such that } R \ge R_1 + R_2\}, \\
R_{\text{sum}}(D_0, D_1, D_2) \triangleq& \\
&\inf\{R : (R, D_0, D_1, D_2) \in \mathscr{A}_{\text{sum}}\}.
\end{aligned}
\tag{2}
$$

The following result follows directly from the above definitions.

Proposition 1. $\text{conv}(\mathscr{A}_{\text{sum}}) = \widetilde{\mathscr{A}_{\text{sum}}}$.

Let $f(x_0, x_1, x_2, x)$, $q(x_0, x_1, x_2)$, $r(x_1)$, and $s(x_2)$ be functions over $\left(\prod_{i=0}^{i=2} \mathcal{X}_i\right) \times \mathcal{X}$, $\prod_{i=0}^{i=2} \mathcal{X}_i$, $\mathcal{X}_1$, and $\mathcal{X}_2$, respectively, where all the alphabets are assumed to be finite.

Define

$$
\begin{aligned}
&\mathscr{G}\left(f(x_0, x_1, x_2, x), q(x_0, x_1, x_2), r(x_1), s(x_2)\right) \\
&\triangleq \sum_{x \in \mathcal{X}} \sum_{x_0 \in \mathcal{X}_0} \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} f(x_0, x_1, x_2, x) \times \\
&\qquad \log \frac{f(x_0, x_1, x_2, x)}{p_X(x)q(x_0, x_1, x_2)r(x_1)s(x_2)},
\end{aligned}
\tag{3}
$$

and

$$
\begin{aligned}
\mathscr{P} \triangleq \{&(\tilde{q}(x_0, x_1, x_2), \tilde{r}(x_1), \tilde{s}(x_2)) : \tilde{q}(x_0, x_1, x_2), \tilde{r}(x_1), \text{and} \\
&\tilde{s}(x_2) \text{ are nonnegative functions defined over} \\
&\mathcal{X}_0 \times \mathcal{X}_1 \times \mathcal{X}_2, \mathcal{X}_1, \text{and } \mathcal{X}_2, \text{respectively;} \\
&\textstyle\sum_{x_1 \in \mathcal{X}_1} \tilde{r}(x_1) = 1; \sum_{x_2 \in \mathcal{X}_2} \tilde{s}(x_2) = 1; \\
&\textstyle\sum_{x_0 \in \mathcal{X}_0} \tilde{q}(x_0, x_1, x_2) = 1, \forall (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2.\}.
\end{aligned}
\tag{4}
$$

Proposition 2. $R_{\text{sum}}(D_0, D_1, D_2)$ is given by (5).

We use $(f^*(x_0, x_1, x_2, x), q^*(x_0, x_1, x_2), r^*(x_1), s^*(x_2))$ to denote the minimizer of the optimization problem given by (5).

Proof of Proposition 2: It can be shown by using the log-sum inequality [5, Theorem 2.7.1] that

$$
\begin{cases}
q^*(x_0, x_1, x_2) = \frac{\sum_x f^*(x_0, x_1, x_2, x)}{\sum_{x, x_0} f^*(x_0, x_1, x_2, x)}, \\
r^*(x_1) = \sum_{x, x_0, x_2} f^*(x_0, x_1, x_2, x), \\
s^*(x_2) = \sum_{x, x_0, x_1} f^*(x_0, x_1, x_2, x),
\end{cases}
\tag{6}
$$

which, together with (2), implies (5). ∎

According to the Karush-Kuhn-Tucker (KKT) conditions [15, Proposition 3.3.7] for the minimization problem in (5), there exist $\lambda_0, \lambda_1, \lambda_2 \ge 0$ such that

$$
\begin{aligned}
&f^*(x_0, x_1, x_2, x) = \\
&\frac{p_X(x)q^*(x_0, x_1, x_2)r^*(x_1)s^*(x_2)e^{-\sum_{i=0}^{i=2}\lambda_i d(x,x_i)}}{\sum_{x_0, x_1, x_2} q^*(x_0, x_1, x_2)r^*(x_1)s^*(x_2)e^{-\sum_{i=0}^{i=2}\lambda_i d(x,x_i)}}.
\end{aligned}
\tag{7}
$$

Define $L(\lambda_0, \lambda_1, \lambda_2)$ by (8). We then have the following result, whose proof is given in Appendix A.

Proposition 3. Let $\left(\widehat{q}(x_0, x_1, x_2), \widehat{r}(x_1), \widehat{s}(x_2)\right)$ be the minimizer of (8) for some $(\lambda_0, \lambda_1, \lambda_2) \ge 0$. Let $\widehat{f}(x_0, x_1, x_2, x)$ be given by (7) with $q^*(\cdot), r^*(\cdot), s^*(\cdot)$ replaced by $\widehat{q}(\cdot), \widehat{r}(\cdot), \widehat{s}(\cdot)$, respectively. Let $\widehat{D}_i = \sum_{x_0, x_1, x_2, x} \widehat{f}(x_0, x_1, x_2, x)d(x, x_i)$. Then, $\left(\widehat{f}(\cdot), \widehat{q}(\cdot), \widehat{r}(\cdot), \widehat{s}(\cdot)\right)$ is the minimizer of (5) with $D_i = \widehat{D}_i, i = 0, 1, 2$. Let $\widehat{R} = \mathscr{G}\left(\widehat{f}(\cdot), \widehat{q}(\cdot), \widehat{r}(\cdot), \widehat{s}(\cdot)\right)$. Then, $\left(\widehat{R}, \widehat{D}_0, \widehat{D}_1, \widehat{D}_2\right) \in \partial(\text{conv}(\mathscr{A}_{\text{sum}}))$.

According to the EGC theorem, the EGC region is formed by convex combinations of points, each of which corresponds to a joint pmf $f(x, x_0, x_1, x_2)$ over $\mathcal{X} \times \mathcal{X}_0 \times \mathcal{X}_1 \times \mathcal{X}_2$. It is shown in [6] that the marginals of $f(x, x_0, x_1, x_2)$, i.e., $f(x_1)$, $f(x_2)$, $f(x_0|x_1, x_2)$, can be used to design the codebooks. However, in universal coding, the pmf $f(x, x_0, x_1, x_2)$ is not known a priori, and the codebooks need to be designed beforehand. To find a universal scheme, it is necessary to reformulate (1) so that the probability distributions for designing the codebooks are decoupled from $f(x, x_0, x_1, x_2)$. With this in mind, we define $\mathscr{G}(\cdot)$ in (3) and establish its relationship

$$R_{\text{sum}}(D_0, D_1, D_2) = \min_{f(\cdot), q(\cdot), r(\cdot), s(\cdot)} \mathscr{G}\left(f(x_0, x_1, x_2, x), q(x_0, x_1, x_2), r(x_1), s(x_2)\right) \tag{5}$$
$$\text{subject to} \sum_{x, x_0, x_1, x_2} f(x_0, x_1, x_2, x) d(x, x_i) \leq D_i, i = 0, 1, 2.$$
$$\sum_{x_0, x_1, x_2} f(x_0, x_1, x_2, x) = p_X(x), \forall x \in \mathcal{X},$$
$$(q(x_0, x_1, x_2), r(x_1), s(x_2)) \in \mathscr{P}, f(x_0, x_1, x_2, x) \geq 0.$$

$$L(\lambda_0, \lambda_1, \lambda_2) \triangleq \min_{q(\cdot), r(\cdot), s(\cdot)} -\sum_x p_X(x) \log \left(\sum_{x_0, x_1, x_2} q(x_0, x_1, x_2) r(x_1) s(x_2) e^{-\sum_{i=0}^{i=2} \lambda_i d(x, x_i)} \right) \tag{8}$$
$$\text{subject to} (q(x_0, x_1, x_2), r(x_1), s(x_2)) \in \mathscr{P}.$$

with $R_{\text{sum}}(D_0, D_1, D_2)$ in (5). The relationship between $R_{\text{sum}}(D_0, D_1, D_2)$ and $\widetilde{\mathscr{A}}_{\text{sum}}$ can be found by combining Propositions 1, 2, and 3.

IV. GENERALIZING NATURAL TYPE SELECTION TO MDC

A. The Basic Coding Process

We first present two 2-channel MDC schemes, which form the basis of the UMDC schemes given in Section IV-C. The schemes presented here are motivated by the EGC scheme [6], though they have several crucial differences. It will be seen that, according to the theoretical results in Section IV-B, these schemes can be viewed as "natural type selectors" for UMDC.

Fig. 1 describes the modules constituting the proposed schemes. The encoder has two alternative forms, depicted in Fig 1(a) and Fig. 1(b), respectively. In Fig. 1(a), for each source vector, the encoder first calls a procedure named SearchWithFixedDistortion. The procedure begins with the codebook generation, using three functions $q(\cdot, \cdot, \cdot)$, $r(\cdot)$, and $s(\cdot)$ which satisfy the following condition:

$$(q(\cdot, \cdot, \cdot), r(\cdot), s(\cdot)) \in \mathscr{P}, \tag{9}$$

where $\mathscr{P}$ is defined in (4). Then, each possible combination of codewords is examined until the distortion constraints are met. The procedure outputs a triple of codewords and a triple of integers (m_0, m_1, m_2) indicating the locations of the codewords in the codebooks. Elias's δ-code [7] is then used to convert m_0, m_1, and m_2 into three binary sequences. The binary sequences corresponding to m_1 and m_2 are transmitted in Descriptions 1 and 2, respectively. The binary sequence of m_0 is split into both descriptions. Fig. 1(b) is similar to Fig. 1(a) except that the encoder calls a procedure named SearchWithFixedWeight, which returns the triple of codewords with the least weighted distortion value in the codebooks.

Fig. 1(c) shows the diagram of a side decoder. The side decoder only decodes a portion of the received bitstream. It first performs integer decoding, then a procedure named SideReconstruction is invoked to reconstruct the codewords. The central decoder is displayed in Fig. 1(d). It performs integer decoding to retrieve (m_0, m_1, m_2), which was previously returned by SearchWithFixedDistortion/Weight at the encoder. Then it calls a procedure named CentralReconstruction. All the procedures mentioned above are given in Fig. 2.

Fig. 1(a), Fig. 1(c), and Fig. 1(d) constitute the scheme of fixed-distortion*. Fig. 1(b), Fig. 1(c), and Fig. 1(d) constitute the scheme of fixed-weight.

In practice, the random sampling performed in the proposed procedures is implemented with pseudo-random number generators (PNGs), which are shared by the encoder and the decoder. Suppose that SearchWithFixedDistortion or SearchWithFixedWeight is invoked for a source vector **x**. Let $q(\cdot, \cdot, \cdot)$, $r(\cdot)$, and $s(\cdot)$ be the functions input to the procedure, and let $(\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2)$ be the triple of codewords returned. To ensure that $\mathbf{x}_i (i = 1, 2)$ is reconstructed by side decoding of the i-th description, and that $\mathbf{x}_0$ is reconstructed by central decoding, it suffices for the decoder to know $q(\cdot, \cdot, \cdot)$, $r(\cdot)$, and $s(\cdot)$. In Section IV-C it will be revealed how the decoder synchronizes its $q(\cdot, \cdot, \cdot)$, $r(\cdot)$, $s(\cdot)$ with those of the encoder. For the time being, we assume that the above condition is satisfied.

B. Theoretical Results

In this section, we derive several results regarding the schemes given in Section IV-A. We begin with some definitions. Let $w(x_0, x_1, x_2)$ be a pmf over $\mathcal{X}_0 \times \mathcal{X}_1 \times \mathcal{X}_2$. Define

$$I_A(p_X(x), w(x_0, x_1, x_2), D_0, D_1, D_2)$$
$$\triangleq \min_{f(\cdot)} \sum_{x \in \mathcal{X}} \sum_{x_0 \in \mathcal{X}_0} \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} f(x_0, x_1, x_2, x) \times$$
$$\log \frac{f(x_0, x_1, x_2, x)}{p_X(x) w(x_0, x_1, x_2)}$$
$$\text{subject to} \sum_{x, x_0, x_1, x_2} f(x_0, x_1, x_2, x) d(x, x_i) \leq D_i,$$
$$i = 0, 1, 2,$$
$$\sum_x f(x_0, x_1, x_2, x) = w(x_0, x_1, x_2),$$
$$\sum_{x_0, x_1, x_2} f(x_0, x_1, x_2, x) = p_X(x), \forall x \in \mathcal{X},$$
$$f(x_0, x_1, x_2, x) \geq 0.$$

Parallel to [3] and the references therein, we refer to $I_A(\cdot)$ as "lower mutual information". If the feasible region of the above optimization problem is empty, we set $I_A(\cdot) = +\infty$.

---

*Lines 14-16 in SearchWithFixedDistortion correspond to the situation of coding failure. When this procedure is used, M (see line 1) should be chosen large enough so that the probability of coding failure is negligible.
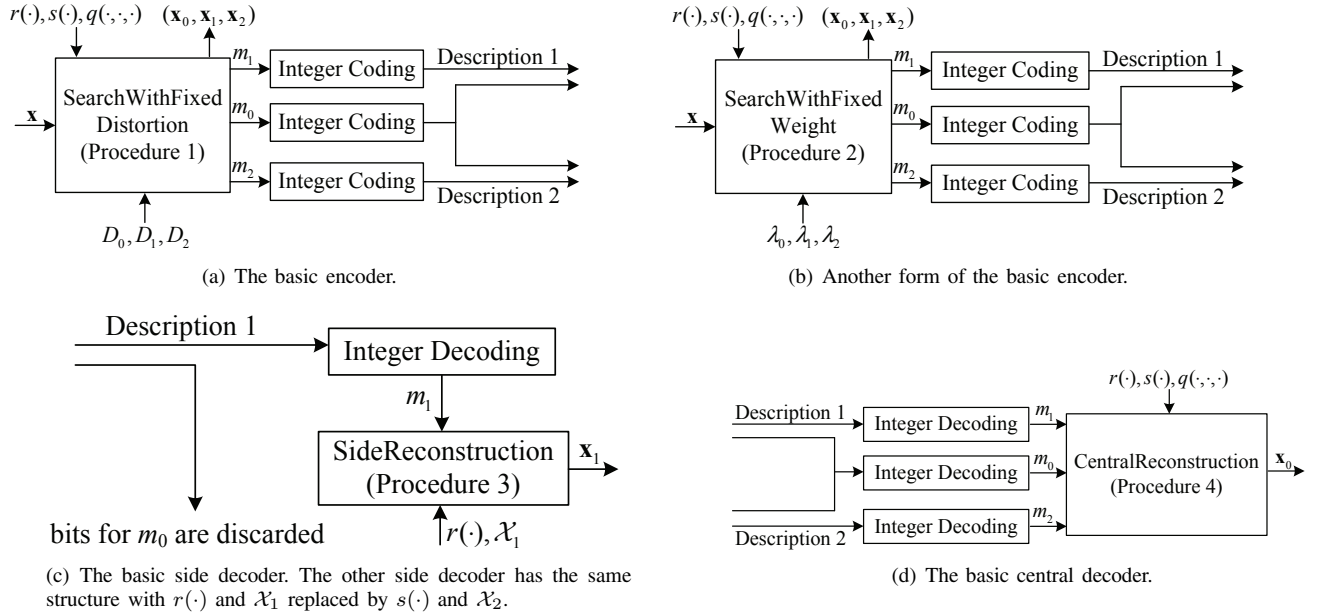
Fig. 1. The basic coding modules. The procedures involved are given in Fig. 2.

For $q(x_0, x_1, x_2)$, $r(x_1)$, and $s(x_2)$ satisfying (9), define

$$\mathscr{D}\left(w(x_0, x_1, x_2) \parallel q(x_0, x_1, x_2)r(x_1)s(x_2)\right)$$
$$\triangleq \sum_{x_0, x_1, x_2} w(x_0, x_1, x_2) \log\left(\frac{w(x_0, x_1, x_2)}{q(x_0, x_1, x_2)r(x_1)s(x_2)}\right)$$

and

$$R_A(p_X, q, r, s, D_0, D_1, D_2)$$
$$\triangleq \min_w \{I_A(p_X, w, D_0, D_1, D_2) + \mathscr{D}(w \parallel qrs)\}, \quad (10)$$

where we abbreviate $p_X(x)$, $w(x_0, x_1, x_2)$, $q(x_0, x_1, x_2)$, $r(x_1)$, $s(x_2)$ to p_X, w, q, r, s, respectively. Note that $I_A(\cdot)$ and $R_A(\cdot)$ will be used to analyze SearchWithFixedDistortion. For SearchWithFixedWeight, let $\lambda_0, \lambda_1, \lambda_2 \geq 0$ and define

$$I_B(p_X, w, \lambda_0, \lambda_1, \lambda_2, \overline{D})$$
$$\triangleq \min_{D_0, D_1, D_2} I_A(p_X, w, D_0, D_1, D_2),$$
$$\text{subject to } \sum_{i=0}^{i=2} \lambda_i D_i \leq \overline{D},$$
$$R_B(p_X, q, r, s, \lambda_0, \lambda_1, \lambda_2, \overline{D})$$
$$\triangleq \min_w \{I_B(p_X, w, \lambda_0, \lambda_1, \lambda_2, \overline{D}) + \mathscr{D}(w \parallel qrs)\}. \quad (11)$$

Using (10), we get

$$R_B(p_X, q, r, s, \lambda_0, \lambda_1, \lambda_2, \overline{D})$$
$$= \min_{D_0, D_1, D_2} R_A(p_X, q, r, s, D_0, D_1, D_2)$$
$$\text{subject to } \sum_{i=0}^{i=2} \lambda_i D_i \leq \overline{D}. \quad (12)$$

The properties of $R_A(\cdot)$ and $R_B(\cdot)$ are summarized as follows.

Proposition 4. If $q(x_0, x_1, x_2)$, $r(x_1)$, and $s(x_2)$ are strictly positive, then
  (i) The minimizers of (10) and (11) are both unique.
  (ii) Both $R_A(p_X, q, r, s, D_0, D_1, D_2)$ and $R_B(p_X, q, r, s, \lambda_0, \lambda_1, \lambda_2, \overline{D})$ are continuous functions of $p_X(\cdot)$ under the norm for $p_X(\cdot)$ defined by $\|p_X(\cdot)\|_1 = \sum_{x \in \mathcal{X}} |p_X(x)|$.

  (iii) $R_A(p_X, q, r, s, D_0, D_1, D_2)$ is continuous in (D_0, D_1, D_2). $R_B(p_X, q, r, s, \lambda_0, \lambda_1, \lambda_2, \overline{D})$ is continuous in $\overline{D}$.

Proof Outline: The first result can be proved following similar steps in [3]. The next two can be verified by using the proof technique in [4, Chapter 2]. ∎

For $R_A(\cdot)$, we also have the following result, which follows from (5).

Proposition 5. $R_{\text{sum}}(D_0, D_1, D_2) = \min_{q, r, s} R_A(p_X, q, r, s, D_0, D_1, D_2)$.

The following proposition can be proved using the definition of $R_B(\cdot)$.

Proposition 6. Given $p_X, q, r, s, \lambda_0, \lambda_1, \lambda_2$, there exists a $\overline{D}_{\max} \geq 0$ such that
  (i) $R_B(p_X, q, r, s, \lambda_0, \lambda_1, \lambda_2, \overline{D}) > 0, \forall \overline{D} < \overline{D}_{\max}$, and is zero otherwise,
  (ii) $R_B(p_X, q, r, s, \lambda_0, \lambda_1, \lambda_2, \overline{D})$ is a convex and strictly decreasing function of $\overline{D}$ when $0 < \overline{D} < \overline{D}_{\max}$.

Next, we introduce two lemmas, which will be used to prove our main results.

Lemma 1. Let $\mathscr{M}_k = \{(m_0, m_1, m_2) : m_0, m_1, m_2 \text{ are positive integers}, m_0 m_1 m_2 = k.\}$. Then,

$$\lim_{m \to \infty} \frac{\log(\sum_{k=1}^{k=m} |\mathscr{M}_k|)}{\log m} = 1. \quad (13)$$

Lemma 2. Suppose that $q(x_0, x_1, x_2)$, $r(x_1)$, and $s(x_2)$ satisfy the constraint in (9). Let $\mathbf{x}$ be some n-dimensional vector output by the source. Let $\mathbf{X}_0$, $\mathbf{X}_1$, and $\mathbf{X}_2$ be three n-dimensional random vectors. The elements of $\mathbf{X}_1$ and $\mathbf{X}_2$ are drawn i.i.d. from the pmfs $r(x_1)$ and $s(x_2)$, respectively. The i-th element of $\mathbf{X}_0$, denoted by $\mathbf{X}_0[i]$, is generated by random

---

Procedure 1 SearchWithFixedDistortion($\mathbf{x}, D_0, D_1, D_2, q(\cdot,\cdot,\cdot), r(\cdot), s(\cdot)$)

Require: $q(\cdot,\cdot,\cdot)$, $r(\cdot)$, and $s(\cdot)$ are functions satisfying the condition given by (9). $\triangleright$ $\mathbf{x}$ is a source vector.

1: $n \leftarrow$ the dimension of $\mathbf{x}$. $M \leftarrow$ a positive integer; define three n-dimensional vectors: $\mathbf{x}_0 \in \mathcal{X}_0^n$, $\mathbf{x}_1 \in \mathcal{X}_1^n$, and $\mathbf{x}_2 \in \mathcal{X}_2^n$; define two $M \times n$ matrices: $\mathbf{A}_1 \in \mathcal{X}_1^{M \times n}$ and $\mathbf{A}_2 \in \mathcal{X}_2^{M \times n}$; define an $M \times M \times M \times n$ matrix $\mathbf{A}_0 \in \mathcal{X}_0^{M \times M \times M \times n}$

2: fill in $\mathbf{A}_1$ and $\mathbf{A}_2$ by iid sampling from the pmfs $r(x_1), x_1 \in \mathcal{X}_1$ and $s(x_2), x_2 \in \mathcal{X}_2$, respectively

3: **for** each $(m_0, m_1, m_2) \in \{1, 2, \cdots, M\}^3$ **do**

4:     **for** $i = 1$ to n **do**

5:         generate $\mathbf{A}_0[m_0][m_1][m_2][i]$ by iid sampling from $q(x_0, \mathbf{A}_1[m_1][i], \mathbf{A}_2[m_2][i])$, $x_0 \in \mathcal{X}_0$

6: **for** $k = 1$ to M **do**

7:     $\mathcal{M}_k \leftarrow \{(m_0, m_1, m_2) : m_0, m_1, m_2 \text{ are positive integers}, m_0 m_1 m_2 = k.\}$

8:     **for** each $(m_0, m_1, m_2) \in \mathcal{M}_k$ **do**

9:         **for** $i = 1$ to n **do**

10:             $\mathbf{x}_1[i] \leftarrow \mathbf{A}_1[m_1][i]$; $\mathbf{x}_2[i] \leftarrow \mathbf{A}_2[m_2][i]$; $\mathbf{x}_0[i] \leftarrow \mathbf{A}_0[m_0][m_1][m_2][i]$

11:         compute $d(\mathbf{x}, \mathbf{x}_i), i = 0, 1, 2$

12:         **if** $d(\mathbf{x}, \mathbf{x}_i) \leq D_i, i = 0, 1, 2$ **then**

13:             **return** $(m_0, m_1, m_2), (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2)$

14: **for** $i = 1$ to n **do**

15:     $\mathbf{x}_1[i] \leftarrow \mathbf{A}_1[1][i]$; $\mathbf{x}_2[i] \leftarrow \mathbf{A}_2[1][i]$; $\mathbf{x}_0[i] \leftarrow \mathbf{A}_0[1][1][1][i]$

16: **return** $(1, 1, 1), (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2)$

Procedure 2 SearchWithFixedWeight($\mathbf{x}, \lambda_0, \lambda_1, \lambda_2, q(\cdot,\cdot,\cdot), r(\cdot), s(\cdot)$)

1: execute line 1 to 5 in Procedure 1

2: $(m_0, m_1, m_2) \leftarrow \arg\min_{(k_0, k_1, k_2)} \{\lambda_0 d(\mathbf{x}, \mathbf{A}_0[k_0][k_1][k_2]) + \sum_{i=1}^{i=2} \lambda_i d(\mathbf{x}, \mathbf{A}_i[k_i]), \text{subject to } k_0 k_1 k_2 \leq M\}$

3: **for** $i = 1$ to n **do**

4:     $\mathbf{x}_1[i] \leftarrow \mathbf{A}_1[m_1][i]$; $\mathbf{x}_2[i] \leftarrow \mathbf{A}_2[m_2][i]$; $\mathbf{x}_0[i] \leftarrow \mathbf{A}_0[m_0][m_1][m_2][i]$

5: **return** $(m_0, m_1, m_2), (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2)$

Procedure 3 SideReconstruction($m, g(\cdot), \mathcal{Y}$)

Require: m is a positive integer. $\mathcal{Y}$ is a finite alphabet. $g(\cdot)$ is a pmf over $\mathcal{Y}$.

1: $n, M \leftarrow$ same as those in SearchWithFixedDistortion/Weight; define an n-dimensional vector $\mathbf{y} \in \mathcal{Y}^n$; define an $M \times n$ matrix $\mathbf{A} \in \mathcal{Y}^{M \times n}$

2: fill in $\mathbf{A}$ by iid sampling from the pmf $g(y), y \in \mathcal{Y}$

3: **for** $i = 1$ to n **do**

4:     $\mathbf{y}[i] \leftarrow \mathbf{A}[m][i]$

5: **return** $\mathbf{y}$

Procedure 4 CentralReconstruction($m_0, m_1, m_2, q(\cdot,\cdot,\cdot), r(\cdot), s(\cdot)$)

Require: m_0, m_1, m_2 are positive integers. $q(\cdot,\cdot,\cdot)$, $r(\cdot)$, $s(\cdot)$ are functions satisfying the condition given by (9).

1: $n, M \leftarrow$ same as those in SearchWithFixedDistortion/Weight; define an n-dimensional vector $\mathbf{x}_0 \in \mathcal{X}_0^n$; define two $M \times n$ matrices: $\mathbf{A}_1 \in \mathcal{X}_1^{M \times n}$ and $\mathbf{A}_2 \in \mathcal{X}_2^{M \times n}$; define an $M \times M \times M \times n$ matrix $\mathbf{A}_0 \in \mathcal{X}_0^{M \times M \times M \times n}$

2: fill in $\mathbf{A}_1$ and $\mathbf{A}_2$ by iid sampling from the pmfs $r(x_1), x_1 \in \mathcal{X}_1$ and $s(x_2), x_2 \in \mathcal{X}_2$, respectively

3: **for** each $(k_0, k_1, k_2) \in \{1, 2, \cdots, M\}^3$ **do**

4:     **for** $i = 1$ to n **do**

5:         generate $\mathbf{A}_0[k_0][k_1][k_2][i]$ by iid sampling from $q(x_0, \mathbf{A}_1[k_1][i], \mathbf{A}_2[k_2][i])$, $x_0 \in \mathcal{X}_0$

6: **for** $i = 1$ to n **do**

7:     $\mathbf{x}_0[i] \leftarrow \mathbf{A}_0[m_0][m_1][m_2][i]$

8: **return** $\mathbf{x}_0$

Fig. 2. The procedures which are called by the coding modules given in Fig. 1. $\mathcal{X}_0$, $\mathcal{X}_1$, and $\mathcal{X}_2$ are finite alphabets. D_0, D_1, and D_2 are distortions. On line 11 of Procedure 1, $d(\mathbf{x}, \mathbf{x}_i) = \frac{1}{n} \sum_{j=1}^{j=n} d_{\mathcal{X}, \mathcal{X}_i}(\mathbf{x}[j], \mathbf{x}_i[j])$.

sampling from the pmf $q(x_0, \mathbf{X}_1[i], \mathbf{X}_2[i])$, $x_0 \in \mathcal{X}_0$. Then,

$$-\frac{1}{n}\log(Q_A(\mathbf{x}, D_0, D_1, D_2))$$
$$\triangleq -\frac{1}{n}\log(\Pr\{d_{\mathcal{X}, \mathcal{X}_i}(\mathbf{x}, \mathbf{X}_i) \leq D_i, i = 0, 1, 2\})$$
$$= R_A(t_{\mathbf{x}}, q, r, s, D_0, D_1, D_2) + o(1), \quad (14)$$
$$-\frac{1}{n}\log(Q_B(\mathbf{x}, \lambda_0, \lambda_1, \lambda_2, \overline{D}))$$
$$\triangleq -\frac{1}{n}\log(\Pr\{\sum_{i=0}^{i=2} \lambda_i d(\mathbf{x}, \mathbf{X}_i) \leq \overline{D}\})$$
$$= R_B(t_{\mathbf{x}}, q, r, s, \lambda_0, \lambda_1, \lambda_2, \overline{D}) + o(1), \quad (15)$$

where $\lim_{n \to \infty} o(1) = 0$.

The proof of Lemma 1 is given in Appendix B. Lemma 2 follows in similar lines to the derivations in [9] and [3]. Equipped with the above lemmas as well as the properties of $R_A(\cdot)$ and $R_B(\cdot)$, we are ready to analyze the procedures shown in Fig. 2. Our next result reveals how $R_A(p_X, q, r, s, D_0, D_1, D_2)$ is related to SearchWithFixedDistortion. It also shows that this procedure plays the role of a "natural type selector".

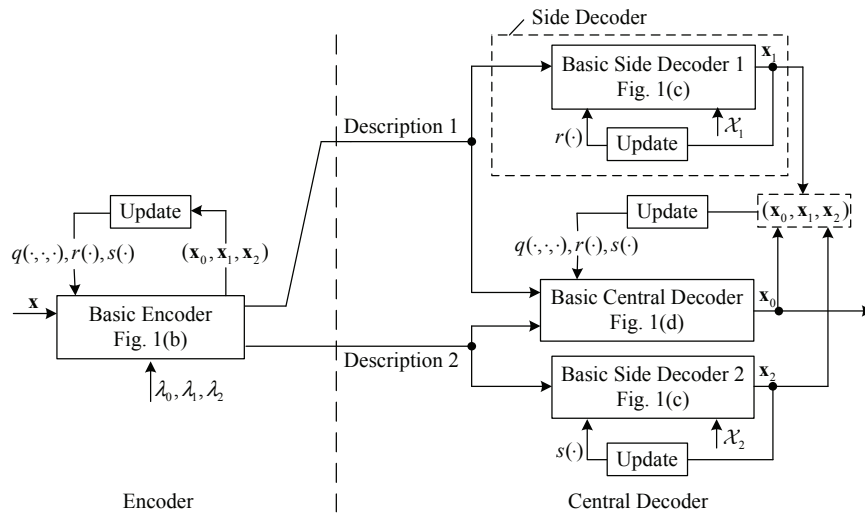*Proposition* 7. Let $\mathbf{X}$ be an n-dimensional random source vector. Let
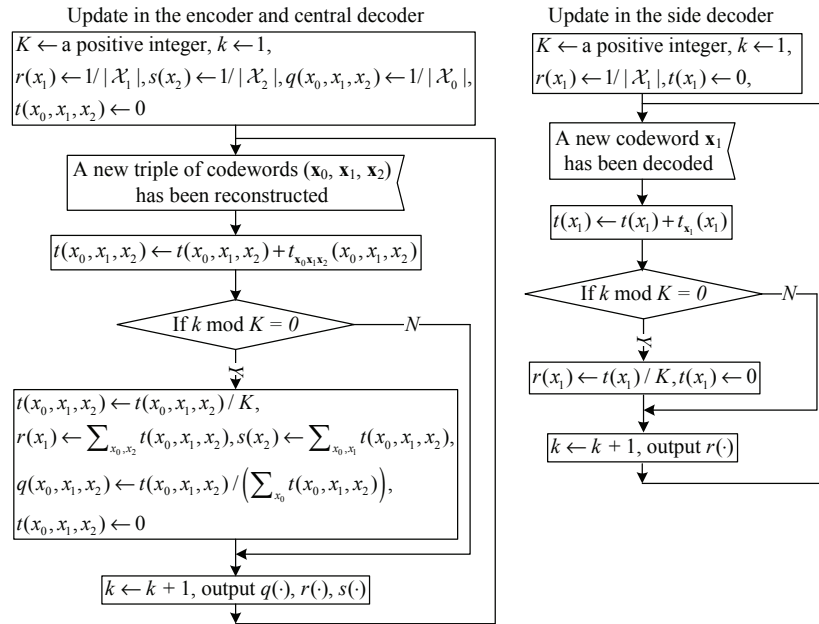
Fig. 3.   The complete framework of UMDC.

Fig. 4.   The update modules in Fig. 3: updating the functions used to generate the codebooks.

$$((M_0, M_1, M_2), (\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2)) =$$
$$\text{SearchWithFixedDistortion}(\mathbf{X}, D_0, D_1, D_2, q(\cdot, \cdot, \cdot), r(\cdot), s(\cdot)),$$

where (M_0, M_1, M_2) is a triple of positive integers; $(\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2)$ is a triple of n-dimensional vectors; $q(\cdot, \cdot, \cdot)$, $r(\cdot)$, and $s(\cdot)$ are positive-valued functions satisfying the constraint in (9). Suppose in SearchWithFixedDistortion, $M \gg 2^{nR_A(p_X, q, r, s, D_0, D_1, D_2)}$. Then $\forall \epsilon > 0, \exists n_0 > 0, \forall n > n_0$,

$$\Pr\left\{\left|\frac{\log(M_0 M_1 M_2)}{n} - R_A(p_X, q, r, s, D_0, D_1, D_2)\right| < \epsilon\right\}$$
$$> 1 - \epsilon, \qquad (16)$$

$$\Pr\left\{\|t_{\mathbf{x}_0 \mathbf{x}_1 \mathbf{x}_2} - w_A\|_1 < \epsilon\right\} > 1 - \epsilon, \qquad (17)$$

where w_A is the minimizer of (10).

Remark 1. Let $L(\cdot)$ be the length function of Elias's δ-code [7]. It is known [7] that $L(m) \sim \log(m)$. Since Elias's δ-scheme is used to encode each $M_i, i = 0, 1, 2$, into a binary sequence, asymptotically the rate of M_i is $\frac{1}{n} \log(M_i)$ and the sum-rate equals $\frac{1}{n} \log(M_0 M_1 M_2)$. Combining this with (16) and Proposition 5, we note that if $q(\cdot, \cdot, \cdot)$, $r(\cdot)$, and $s(\cdot)$ are chosen appropriately, then asymptotically the coding system depicted in Fig. 1 with Fig. 1(a) as the encoder achieves the EGC sum-rate bound.

Remark 2. By Proposition 4(i), the minimizer of (10) is unique. Thus, (17) implies that the codewords returned by SearchWithFixedDistortion have a joint type approximately equal to w_A. As a consequence, this procedure can be interpreted as a "type selector".

Remark 3. The conditions given in Proposition 7 for $q(\cdot, \cdot, \cdot)$, $r(\cdot)$, $s(\cdot)$ are slightly stronger than those given by (9). We require $q(\cdot, \cdot, \cdot)$, $r(\cdot)$, $s(\cdot)$ to be strictly positive in order
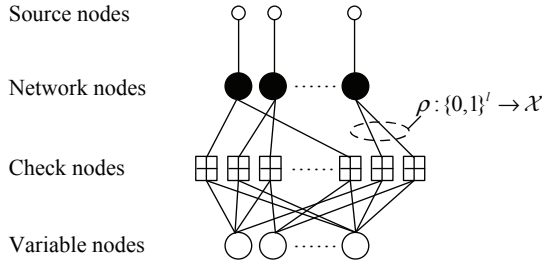
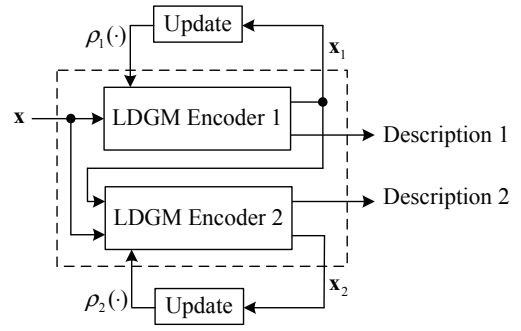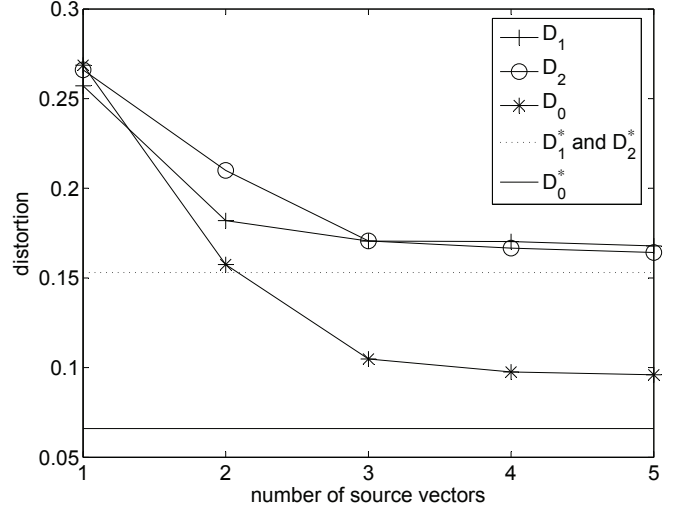Fig. 5. The multilevel LDGM code proposed in [10].

Fig. 6. Updating the codebooks of the MDC scheme proposed in [12] shown inside the dashed box.

Fig. 7. Convergence behavior of the UMDC scheme on a nonuniform binary source with Hamming distortion. $p_X(0) = 0.75$. $R_1 = R_2 = 0.231$ bit/symbol. $(R_1, R_2, D_0^*, D_1^*, D_2^*)$ is located on the EGC bound. The length of each source vector is 10000.

to prove (16) and (17). But note that to ensure Search-WithFixedDistortion to be executable, the conditions given by (9) suffice. It is unclear whether (16) and (17) hold if $q(x_0, x_1, x_2)r(x_1)s(x_2) = 0$ for some (x_0, x_1, x_2).

The proof of Proposition 7 is given in Appendix C. For SearchWithFixedWeight, we have the following result, whose proof is given in Appendix D.

Proposition 8. Let $\mathbf{X}$ be an n-dimensional random source vector. Let

$$((M_0, M_1, M_2), (\mathbf{X}_0, \mathbf{X}_1, \mathbf{X}_2)) =$$
$$\text{SearchWithFixedWeight}(\mathbf{X}, \lambda_0, \lambda_1, \lambda_2, q(\cdot, \cdot, \cdot), r(\cdot), s(\cdot)),$$

where $(\lambda_0, \lambda_1, \lambda_2) \geq 0$ and $q(\cdot, \cdot, \cdot)$, $r(\cdot)$, $s(\cdot)$ are positive-valued functions satisfying the constraint in (9). Suppose in SearchWithFixedWeight, $M = \left\lfloor 2^{nR_B(p_X, q, r, s, \lambda_0, \lambda_1, \lambda_2, \overline{D})} \right\rfloor$ for some $0 < \overline{D} < \overline{D}_{\max}$. Then, $\forall \epsilon > 0$, $\exists n_0 > 0$, $\forall n > n_0$,

$$\Pr\left\{ \left| \overline{D} - \sum_{i=0}^{i=2} \lambda_i d(\mathbf{X}, \mathbf{X}_i) \right| < \epsilon \right\} > 1 - \epsilon, \quad (18)$$

$$\Pr\left\{ \|t_{\mathbf{X}_0 \mathbf{X}_1 \mathbf{X}_2} - w_B\|_1 < \epsilon \right\} > 1 - \epsilon, \quad (19)$$

where w_B is the minimizer of (11).

Remark 4. In view of (19), SearchWithFixedWeight can also be interpreted as a "type selector".

Remark 5. By [7] the sum-rate is asymptotically $\frac{1}{n}\log(M_0 M_1 M_2)$. Since $M_0 M_1 M_2 \leq M$, the sum-rate of SearchWithFixedWeight is asymptotically upper-bound by $\frac{1}{n}\log(M)$. While SearchWithFixedWeight performs a full search and returns the triple with the minimum weighted distortion. Thus, for SearchWithFixedWeight, we conjecture that $\lim_{n\to\infty} \frac{1}{n} E[\log(M_0 M_1 M_2) - \log(M)] = 0$.

Remark 6. For reasons similar to those given in Remark 3, we require $q(\cdot, \cdot, \cdot)$, $r(\cdot)$, $s(\cdot)$ to be positive-valued in Proposition 8.

C. UMDC with Random Codebooks

Fig. 3 shows the fixed-weight UMDC scheme, whose encoder is built upon Fig. 1(b). The scheme can be divided into two parts: the encoder and the central decoder. The side decoder is a subset of the central decoder and is shown inside a dashed box. Besides the modules given in Fig. 1, the scheme includes several new modules named "Update". Fig. 4 shows how these modules update $q(\cdot), r(\cdot), s(\cdot)$, which are used for codebook generation. With the additional modules,

the decoder is able to synchronize its codebooks with those of the encoder.

Let $\mathbf{X}^{(k)}$ be the k-th n-dimensional random vector output by the source, and let $\mathbf{X}_0^{(k)}, \mathbf{X}_1^{(k)}, \mathbf{X}_2^{(k)}$ be the reconstruction codewords of $\mathbf{X}^{(k)}$. Moreover, let $\left(q^{(k)}, r^{(k)}, s^{(k)}\right)$ denote the k-th triple of functions output by the "Update" module at the encoder. For each k, the encoder invokes SearchWithFixedWeight with $\left(q^{(k-1)}, r^{(k-1)}, s^{(k-1)}\right)$ and produces a triple of integers $\left(M_0^{(k)}, M_1^{(k)}, M_2^{(k)}\right)$ indicating the locations of the reconstruction codewords in the codebooks. $\left\{M_1^{(k)} : k \geq 1\right\}$ and $\left\{M_2^{(k)} : k \geq 1\right\}$ are transmitted in Descriptions 1 and 2, respectively, and $\left\{M_0^{(k)} : k \geq 1\right\}$ is split into both descriptions. If there is no transmission error, that is, $\left\{\left(M_0^{(k)}, M_1^{(k)}, M_2^{(k)}\right) : k \geq 1\right\}$ is received correctly, then for each k, the central decoder could reproduce $\left(q^{(k)}, r^{(k)}, s^{(k)}\right)$ by using its "Update" module given in Fig. 4 with the previously reconstructed codewords. If only one description is received correctly, say the description containing $\left\{M_1^{(k)} : k \geq 1\right\}$, then according to Fig. 4, for each k the side decoder could reproduce $r^{(k)}(\cdot)$ with the previously reconstructed codewords $\mathscr{X}_1^{(k)} = \left\{\mathbf{X}_1^{(k')} : \lfloor k/K \rfloor K - K + 1 \leq k' \leq \lfloor k/K \rfloor K\right\}$. Note that the

function $r^{(k)}(\cdot)$ generated at the encoder depends only on $\mathscr{X}_1^{(k)}$. Therefore, the functions required by each decoder for codebook generation can always be obtained from the types of previously reconstructed codewords, and the encoder needs neither to know which description is received, nor to transmit $\left(q^{(k)}, r^{(k)}, s^{(k)}\right)$.

To analyze the R-D performance of the UMDC scheme, we need the following result for $R_B(\cdot)$, which can be proved by using the log-sum inequality [5, Theorem 2.7.1].

Proposition 9. Let

$$\widehat{q}(x_0, x_1, x_2) = \frac{w_B(x_0, x_1, x_2)}{\sum_{x_0} w_B(x_0, x_1, x_2)},$$

$$\widehat{r}(x_1) = \sum_{x_0, x_2} w_B(x_0, x_1, x_2),$$

$$\widehat{s}(x_2) = \sum_{x_0, x_1} w_B(x_0, x_1, x_2),$$

where $w_B(x_0, x_1, x_2)$ is the minimizer of (11). Then,

$$R_B(p_X, \widehat{q}, \widehat{r}, \widehat{s}, \lambda_0, \lambda_1, \lambda_2, \overline{D})$$
$$\leq R_B(p_X, q, r, s, \lambda_0, \lambda_1, \lambda_2, \overline{D}). \qquad (20)$$

The following proposition shows that the weighted distortion of the codewords reconstructed by the proposed scheme does not increase with time when n tends to infinity.

Proposition 10. Suppose in SearchWithFixedWeight, $M = \left\lfloor 2^{nC} \right\rfloor$, where $0 < C < 2H(p_X)$. Then, $\forall \delta > 0, \forall k, \forall \Delta k$,

$$\lim_{n \to \infty} \Pr \left\{ \sum_{i=0}^{i=2} \lambda_i d\left(\mathbf{X}^{(k+\Delta k)}, \mathbf{X}_i^{(k+\Delta k)}\right) \right.$$
$$\left. - \sum_{i=0}^{i=2} \lambda_i d\left(\mathbf{X}^{(k)}, \mathbf{X}_i^{(k)}\right) > \delta \right\} = 0, \qquad (21)$$

where $(\lambda_0, \lambda_1, \lambda_2) > 0$ [†].

Proof Outline: For brevity, we assume $K = 1$ for the modules given in Fig. 4. The proof of (21) for $K > 1$ is similar and is omitted. When $K = 1$,

$$\begin{cases} q^{(k)}(x_0, x_1, x_2) = \frac{t^{(k)}(x_0, x_1, x_2)}{\sum_{x_0} t^{(k)}(x_0, x_1, x_2)}, \\ r^{(k)}(x_1) = \sum_{x_0, x_2} t^{(k)}(x_0, x_1, x_2), \\ s^{(k)}(x_2) = \sum_{x_0, x_1} t^{(k)}(x_0, x_1, x_2), \end{cases} \qquad (22)$$

where $t^{(k)}(x_0, x_1, x_2)$ denotes the joint type of $\left(\mathbf{X}_0^{(k)}, \mathbf{X}_1^{(k)}, \mathbf{X}_2^{(k)}\right)$. Note that if $0 < C < 2H(p_X)$, then for any $(q, r, s) \in \mathscr{P}, q, r, s > 0$, and $(\lambda_0, \lambda_1, \lambda_2) > 0$, there exists a $\overline{D} > 0$ such that

$$R_B(p_X, q, r, s, \lambda_0, \lambda_1, \lambda_2, \overline{D}) = C.$$

Besides, according to (11), for positive-valued $q(\cdot, \cdot, \cdot)$, $r(\cdot), s(\cdot)$, the minimizer of (11) is also positive-valued. Let $w_B^{(k)}(x_0, x_1, x_2)$ be the minimizer of (11) with $q = q^{(k-1)}, r = r^{(k-1)}, s = s^{(k-1)}$. In view of (19), we have

$$\lim_{n \to \infty} \Pr \left\{ t^{(k)} = w_B^{(k)} \right\} = 1. \qquad (23)$$

[†] Here we assume $(\lambda_0, \lambda_1, \lambda_2) > 0$ instead of $(\lambda_0, \lambda_1, \lambda_2) \geq 0$ because in the case $\lambda_0 \lambda_1 \lambda_2 = 0$ the problem degenerates to successive refinement or single description coding.

Combining (23), (22), (20), (18), Proposition 6(ii), and the condition $M = \left\lfloor 2^{nC} \right\rfloor$ with $0 < C < 2H(p_X)$, we find that $\forall \delta > 0, \forall k$,

$$\lim_{n \to \infty} \Pr \left\{ \sum_{i=0}^{i=2} \lambda_i d\left(\mathbf{X}^{(k)}, \mathbf{X}_i^{(k)}\right) \right.$$
$$\left. - \sum_{i=0}^{i=2} \lambda_i d\left(\mathbf{X}^{(k-1)}, \mathbf{X}_i^{(k-1)}\right) > \delta \right\} = 0,$$

from which (21) follows. ∎

One can get the fixed-distortion UMDC scheme by replacing the basic encoder in Fig. 3 with that in Fig. 1(a). It can be shown that as n tends to infinity the sum-rate of this UMDC scheme does not increase with time. It is still an open problem whether the R-D performance of the proposed scheme converges to the EGC bound.

V. APPLICATION TO PRACTICAL MDC

We first review the multilevel LDGM code proposed in [10], which is designed for single description coding. Fig. 5 shows the factor graph of a multilevel LDGM code. The code consists of four levels of nodes. Suppose there are n source nodes, n network nodes, and $l \cdot n$ check nodes where l is some integer. Each check node takes binary values. Each network node is connected to one source node and l check nodes, and its value is determined by the values of its adjacent check nodes through a mapping $\rho : \{0, 1\}^l \to \hat{\mathcal{X}}$. The reconstruction codeword $\hat{\mathbf{x}} \in \hat{\mathcal{X}}^n$ is formed by putting the values of all the network nodes into a vector.

Suppose $\hat{\mathcal{X}} = \mathcal{X} = \{1, 2, \cdots, |\mathcal{X}|\}$. Note that according to [5]

$$R(D) = \min_{p_{\hat{\mathcal{X}}|\mathcal{X}}(\hat{x}|x)} \{I(X; \hat{X}) : E[d(X, \hat{X})] \leq D\}$$
$$= \mathscr{D}\left(p_{\hat{\mathcal{X}}|\mathcal{X}}^*(\hat{x}|x) p_X(x) \| p_{\hat{\mathcal{X}}}^*(\hat{x}) p_X(x)\right), \qquad (24)$$

where $R(D)$ is the R-D function of $p_X(x)$. It is shown in [10] that the following mapping is an asymptotically optimal choice for $\rho(\cdot)$:

$$\rho(c_1, c_2, \cdots, c_l)$$
$$= \begin{cases} \hat{x} \text{ if } \sum_{\hat{y}=1}^{\hat{y}=\hat{x}-1} \left\lfloor 2^l p_{\hat{\mathcal{X}}}^*(\hat{y}) \right\rfloor \leq \sum_{k=1}^{k=l} 2^{k-1} c_k < \sum_{\hat{y}=1}^{\hat{y}=\hat{x}} \left\lfloor 2^l p_{\hat{\mathcal{X}}}^*(\hat{y}) \right\rfloor, \\ 1 \text{ if } \sum_{k=1}^{k=l} 2^{k-1} c_k \geq \sum_{\hat{y}=1}^{\hat{y}=|\mathcal{X}|} \left\lfloor 2^l p_{\hat{\mathcal{X}}}^*(\hat{y}) \right\rfloor. \end{cases}$$
$$(25)$$

Denote by $\mathbf{G}$ the generator matrix of the LDGM code, and by $\mathbf{c}$ the vector composed of the values of the check nodes. Note that $\mathbf{c}$ is of dimension nl and $\mathbf{c} = \mathbf{v}\mathbf{G}$ for some $\mathbf{v}$. Suppose that the entries of $\mathbf{G}$ are generated randomly and independently of each other with a Bernoulli distribution. Then, $t_{\mathbf{c}}$ approaches the uniform distribution as $n \to \infty$ for almost all choices of $\mathbf{v}$ (see [10]). With the $\rho(\cdot)$ given by (25), $t_{\hat{\mathbf{x}}}$, the type of the reconstruction codeword, approaches $p_{\hat{\mathcal{X}}}^*(\hat{x})$ as $l, n \to \infty$, which, according to (24), is necessary for the asymptotic optimality of the whole LDGM code.

In [12], a two-channel MDC scheme using the above LDGM code is proposed for binary sources with Hamming

distortion. The scheme employs two LDGM codebooks, one for each description, and the descriptions are generated sequentially. Given an n-dimensional source vector $\mathbf{x}$, the first description is produced with the penalties assigned according to the source symbols. The penalties for the second description is given based on the first description as well as the source vector. Denote by $\mathbf{x}_j (j = 1, 2)$ the side reconstruction codeword generated with the j-th description. The central reconstruction codeword $\mathbf{x}_0$ is given by the following equation [12]:

$$\mathbf{x}_0[i] = \begin{cases} 0 \text{ if } \mathbf{x}_1[i] = \mathbf{x}_2[i] = 0, \\ 1 \text{ otherwise,} \end{cases} \tag{26}$$

where $1 \leq i \leq n$.

Denote by $\rho_j : \{0,1\}^{l_j} \to \mathcal{X}_j, j = 1, 2$, the nonlinear mapping used in the LDGM codebook of the j-th description. Based on the MDC scheme proposed in [12], a practical universal scheme is given in Fig. 6. Each time a source vector is encoded into two descriptions, and $\rho_j(\cdot)$ is updated according to the following equation:

$$\rho_j(c_1, c_2, \cdots, c_{l_j}) = \begin{cases} 0 \text{ if } \sum_{k=1}^{k=l_j} 2^{k-1} c_k < \left\lfloor 2^{l_j} t_{\mathbf{x}_j}(0) + 0.5 \right\rfloor, \\ 1 \text{ otherwise.} \end{cases}$$

The practical universal scheme given above only contains two codebooks and does not generate refinement information, which is a simplified version of the scheme given in Fig. 3.

VI. EXPERIMENTAL RESULTS

We implement the UMDC scheme given in Fig. 6. Fig. 7 shows the convergence behavior of the scheme when applied to a nonuniform binary source with Hamming distortion. We set the length of each source vector to 10000, and set $l_1 = l_2 = 4$. The generator matrices of the LDGM codes are designed using the algorithm given in [13] for BIAWGN (Binary Input Additive White Gaussian Noise) channels. For the message passing and decimation over the LDGM codes, the corresponding parameters are chosen according to [10] and [12]. The central decoder combines both descriptions with (26). In the "Update" modules depicted in Fig. 4, we set $K = 1$. The curves in Fig. 7 are averaged over 5 independent trials. In Fig. 7, it is observed that the UMDC scheme converges quickly. Initially, the central distortion is slightly larger than the side distortions, because the central decoder always combines both descriptions with (26) and the codebooks at that time are far from being optimal. However, as more source vectors are coded, the central distortion decreases faster, and is eventually less than the side distortions. Tab. I compares the performance of the universal scheme (when it converges) with the non-universal scheme proposed in [11] and the EGC bound in the no-excess sum rate case [14]. On average, the central and side distortions of the proposed scheme are comparable with those of the non-universal scheme, and are about 0.022, 0.010, 0.007 larger than those given by the EGC theorem, respectively.

VII. CONCLUSIONS

We consider universal coding in the context of multiple descriptions. We propose two UMDC schemes using random codebooks. The principle of natural type selection is generalized to MDC and the supporting theoretical results are proved. It is shown that in each scheme the types of reconstruction codewords can help to improve the R-D performance. We then give a practical universal scheme based on the LDGM-based MDC methods. The practical scheme derives its principles from our theoretical framework. Each time a source vector is encoded, the nonlinear mappings inside the LDGM codes are updated according to the types of reconstruction codewords. Experimental results show that the R-D performance of the schemes gradually approaches the EGC bound.

APPENDIX A

Proof of Proposition 3: Given $\lambda_0, \lambda_1, \lambda_2 \geq 0$, define

$$\mathscr{L}(f(x_0, x_1, x_2, x), q(x_0, x_1, x_2), r(x_1), s(x_2), \lambda_0, \lambda_1, \lambda_2)$$
$$\triangleq \mathscr{G}(f(x_0, x_1, x_2, x), q(x_0, x_1, x_2), r(x_1), s(x_2))$$
$$+ \sum_{i=0}^{i=2} \sum_{x_0, x_1, x_2, x} \lambda_i f(x_0, x_1, x_2, x) d(x, x_i).$$

For functions $f(x_0, x_1, x_2, x), q(x_0, x_1, x_2), r(x_1), s(x_2)$ satisfying the constraints in (5) with $D_i = \widehat{D}_i$, we have

$$\mathscr{L}\left(\widehat{f}, \widehat{q}, \widehat{r}, \widehat{s}, \lambda_0, \lambda_1, \lambda_2\right)$$
$$= L(\lambda_0, \lambda_1, \lambda_2) \leq \mathscr{L}(f, q, r, s, \lambda_0, \lambda_1, \lambda_2). \tag{27}$$

Since $\sum f(x_0, x_1, x_2, x) d(x, x_i) \leq \widehat{D}_i, i = 0, 1, 2$, it follows that $\mathscr{G}\left(\widehat{f}, \widehat{q}, \widehat{r}, \widehat{s}\right) \leq \mathscr{G}(f, q, r, s)$, which means that $\left(\widehat{f}, \widehat{q}, \widehat{r}, \widehat{s}\right)$ is the minimizer of (5) with $D_i = \widehat{D}_i$. Notice that the inequality in (27) holds irrespective of the values of $\widehat{D}_i, i = 0, 1, 2$. Thus, $\left(\widehat{R}, \widehat{D}_0, \widehat{D}_1, \widehat{D}_2\right)$ is located on a supporting hyperplane of $\mathscr{A}_{\text{sum}}$, which implies that it belongs to $\partial(\text{conv}(\mathscr{A}_{\text{sum}}))$. ∎

APPENDIX B

Proof of Lemma 1: For each positive integer k, define $\alpha(k) \triangleq |\{l : l|k, l > 0\}|$, where "$l|k$" stands for "$l$ divides k". Then

$$\sum_{k=1}^{k=m} |\mathscr{M}_k| = \sum_{k=1}^{k=m} \sum_{l|k} \alpha(l) = \sum_{k=1}^{k=m} \left\lfloor \frac{m}{k} \right\rfloor \alpha(k).$$

Thus,

$$m \sum_{k=1}^{k=m} \frac{\alpha(k)}{k} - \sum_{k=1}^{k=m} \alpha(k) \leq \sum_{k=1}^{k=m} |\mathscr{M}_k| \leq m \sum_{k=1}^{k=m} \frac{\alpha(k)}{k}.$$

TABLE I
COMPARISON OF THE R-D PERFORMANCE OF THE PROPOSED UMDC SCHEME (3RD COLUMN), THE NON-UNIVERSAL SCHEME [11] (4TH COLUMN), AND THE EGC BOUND IN THE NO-EXCESS-RATE CASE [14] (5TH COLUMN) FOR BINARY SOURCES WITH HAMMING DISTORTION. THE UNIT OF RATE IS BIT/SYMBOL.

$p_X(0)$	(R_1, R_2)	(D_0, D_1, D_2)(UMDC)	(D_0, D_1, D_2)[11]	(D_0^*, D_1^*, D_2^*)
0.5	(0.320, 0.320)	(0.091, 0.236, 0.275)	(0.088, 0.235, 0.284)	(0.068, 0.230, 0.264)
0.5	(0.231, 0.231)	(0.157, 0.286, 0.293)	(0.149, 0.275, 0.318)	(0.124, 0.273, 0.287)
0.5	(0.395, 0.395)	(0.052, 0.205, 0.254)	(0.050, 0.206, 0.270)	(0.033, 0.203, 0.251)
0.5	(0.167, 0.628)	(0.047, 0.345, 0.135)	(0.046, 0.333, 0.150)	(0.032, 0.332, 0.134)
0.75	(0.320, 0.320)	(0.048, 0.137, 0.145)	(0.050, 0.139, 0.148)	(0.026, 0.125, 0.136)
0.75	(0.231, 0.231)	(0.096, 0.168, 0.164)	(0.088, 0.162, 0.166)	(0.066, 0.153, 0.153)
0.75	(0.395, 0.395)	(0.021, 0.119, 0.134)	(0.023, 0.121, 0.137)	(0.002, 0.110, 0.124)
0.75	(0.167, 0.628)	(0.016, 0.185, 0.068)	(0.018, 0.188, 0.069)	(0.002, 0.178, 0.060)

According to [8, p.57, p.70],

$$\sum_{k=1}^{k=m} \frac{\alpha(k)}{k} = \frac{1}{2}\log^2(m) + 2C_0 \log(m) + O(1),$$

$$\sum_{k=1}^{k=m} \alpha(k) = m\log m + (2C_0 - 1)m + O(\sqrt{m}),$$

where C_0 is the Euler's constant. Then, we have

$$C_1 m \log^2 m \leq \sum_{k=1}^{k=m} |\mathcal{M}_k| \leq C_2 m \log^2 m,$$

where C_1 and C_2 are constants. Hence, (13) is proved. ∎

APPENDIX C

Proof of Proposition 7: For brevity, we use $R_A(p_X, \delta)$ as the shorthand for $R_A(p_X, q, r, s, D_0 - \delta, D_1 - \delta, D_2 - \delta)$. To establish (16), we need to prove that $\forall \epsilon, 0 < \epsilon < R_A(p_X, 0)$, $\exists n_1 > 0, \forall n > n_1$,

$$\Pr\left\{\frac{1}{n}\log(M_0 M_1 M_2) < R_A(p_X, 0) - \epsilon\right\} < \epsilon/2, \quad (28)$$

and that $\forall \epsilon > 0, \exists n_2, \forall n > n_2$,

$$\Pr\left\{\frac{1}{n}\log(M_0 M_1 M_2) > R_A(p_X, 0) + \epsilon\right\} < \epsilon/2. \quad (29)$$

We prove (28) by characterizing the size of the codebook used by SearchWithFixedDistortion, and prove (29) by relating it to the set of jointly typical sequences.

We first prove (28). Define $N_n \triangleq \lfloor 2^{n(R_A(p_X,0)-\epsilon)} \rfloor$. In view of Lemma 1, there exists $n_1' > 0$, for all $n > n_1'$,

$$\left|\bigcup_{k=1}^{k=N_n} \mathcal{M}_k\right| \leq 2^{n(R_A(p_X,0)-3\epsilon/4)}. \quad (30)$$

By Proposition 4(ii), $\exists \xi > 0, \forall$pmf $t(\cdot)$ over $\mathcal{X}$, if $\|t(\cdot) - p_X(\cdot)\|_1 \triangleq \sum_{x \in \mathcal{X}} |t(x) - p_X(x)| < \xi$,

$$|R_A(t, q, r, s, D_0, D_1, D_2) - R_A(p_X, 0)| < \epsilon/4. \quad (31)$$

For any finite alphabet $\mathcal{Y}$ and any pmf over it, say $g(y), y \in \mathcal{Y}$, we use $T_{n,\varepsilon}(g)$ to denote the strongly ε-typical set of $g(\cdot)$ in $\mathcal{Y}^n$ [5], *i.e.*,

$$T_{n,\varepsilon}(g) \triangleq \{\mathbf{y} \in \mathcal{Y}^n : |t_{\mathbf{y}}(y) - g(y)| < \varepsilon/|\mathcal{Y}|, \forall y \in \mathcal{Y}\}.$$

Let $\mathbf{X}$ be an n-dimensional random source vector. Then, $\exists n_1'' > 0$ such that $\forall n > n_1''$,

$$\Pr\{\mathbf{X} \in T_{n,\xi}(p_X)\} > 1 - \epsilon/4.$$

Note that $T_{n,\xi}(p_X) \subset \{\mathbf{x} : \|t_{\mathbf{x}} - p_X\|_1 < \xi\}$. Combining this fact with (14) and (31) implies that $\exists n_1''' > 0, \forall n > n_1'''$, $\forall \mathbf{x} \in T_{n,\xi}(p_X)$,

$$Q_A(\mathbf{x}, D_0, D_1, D_2) \leq 2^{-n(R_A(p_X,0)-\epsilon/2)}. \quad (32)$$

Then, $\forall n > n_1 \triangleq \max\{n_1', n_1'', n_1'''\}$,

$$\Pr\left\{M_0 M_1 M_2 < 2^{n(R_A(p_X,0)-\epsilon)}\right\}$$

$$\leq \frac{\epsilon}{4} + \sum_{\mathbf{x} \in T_{n,\xi}(p_X)} \left|\bigcup_{k=1}^{k=N_n} \mathcal{M}_k\right| Q_A(\mathbf{x}, D_0, D_1, D_2) \times$$

$$\prod_{i=1}^{i=n} p_X(\mathbf{x}[i])$$

$$\overset{(a)}{\leq} 2^{n(R_A(p_X,0)-3\epsilon/4)} 2^{-n(R_A(p_X,0)-\epsilon/2)} + \epsilon/4$$

$$= 2^{-n\epsilon/4} + \epsilon/4,$$

where (a) is due to (30) and (32). This proves (28).

Next we prove (29). By Proposition 4(iii), $\forall \epsilon > 0, \exists \delta > 0$, such that $|R_A(p_X, \delta) - R_A(p_X, 0)| < \epsilon/4$. In view of (10), we have

$$R_A(p_X, \delta)$$
$$= \min_{f(\cdot)} \mathscr{G}(f, q, r, s)$$
$$\text{subject to } \sum_{x_0, x_1, x_2} f(x_0, x_1, x_2, x) = p_X(x),$$
$$f(x_0, x_1, x_2, x) \geq 0,$$
$$\sum_{x_0, x_1, x_2, x} f(x_0, x_1, x_2, x)d(x, x_i) \leq D_i - \delta,$$
$$i = 0, 1, 2,$$

where $\mathscr{G}(f, q, r, s)$ is defined in (3). Let $f_\delta^*(x_0, x_1, x_2, x)$ be the minimizer of the above problem. Then,

$$\sum_{x_0, x_1, x_2, x} f_\delta^*(x_0, x_1, x_2, x)d(x, x_i) \leq D_i - \delta, i = 0, 1, 2,$$

$$(33)$$

which implies that $\exists \eta_0 > 0, \forall \eta, 0 < \eta < \eta_0, \forall (\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}) \in T_{n,\eta}(f_\delta^*)$,

$$\sum_{x,x_0,x_1,x_2} t_{\mathbf{x}_0,\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}}(x_0,x_1,x_2,x)d(x,x_i) \leq D_i - \delta/2,$$

$$i = 0, 1, 2. \quad (34)$$

Define the following random events:

$$\mathcal{E}_1 \triangleq \{\exists(m_0,m_1,m_2), m_0 m_1 m_2 \leq 2^{n(R_A(p_X,0)+\epsilon)},$$
$$d(\mathbf{X}, \mathbf{A}_i[m_i]) \leq D_i, i = 1, 2,$$
$$d(\mathbf{X}, \mathbf{A}_0[m_0][m_1][m_2]) \leq D_0\},$$
$$\mathcal{E}_2 \triangleq \{\exists(m_0,m_1,m_2), m_0 m_1 m_2 \leq 2^{n(R_A(p_X,\delta)+\epsilon/2)},$$
$$(\mathbf{A}_0[m_0][m_1][m_2], \mathbf{A}_1[m_1], \mathbf{A}_2[m_2], \mathbf{X}) \in T_{n,\eta}(f_\delta^*)\},$$

where $\mathbf{A}_i, i = 0, 1, 2$, are the matrices defined in SearchWith-FixedDistortion, and

$$\mathbf{A}_i[m_i] \triangleq (\mathbf{A}_i[m_i][1], \mathbf{A}_i[m_i][2], \cdots, \mathbf{A}_i[m_i][n]), i = 1, 2,$$
$$\mathbf{A}_0[m_0][m_1][m_2]$$
$$\triangleq (\mathbf{A}_0[m_0][m_1][m_2][1], \mathbf{A}_0[m_0][m_1][m_2][2], \cdots,$$
$$\mathbf{A}_0[m_0][m_1][m_2][n]).$$

Due to (34) and the fact that $R_A(p_X, \delta) < R_A(p_X, 0) + \epsilon/4$, we have $\Pr\{\mathcal{E}_2\} \leq \Pr\{\mathcal{E}_1\}$. Thus,

$$\Pr\left\{\tfrac{1}{n} \log(M_0 M_1 M_2) > R_A(p_X, 0) + \epsilon\right\} \leq \Pr\{\overline{\mathcal{E}_1}\} \leq \Pr\{\overline{\mathcal{E}_2}\}.$$

It can be shown by leveraging the proof technique in [6] that $\Pr\{\overline{\mathcal{E}_2}\} \to 0$ as $n \to \infty$, which proves (29).

Finally, we prove (17). Suppose there is another procedure SearchWithFixedDistortion$'$ same as SearchWithFixedDistortion except that it does not return until an integer triple (M_0', M_1', M_2') is found such that

(i) $\|t_{\mathbf{A}_0[M_0'][M_1'][M_2']\mathbf{A}_1[M_1']\mathbf{A}_2[M_2']} - w_A\| \geq \epsilon$,
(ii) $d(\mathbf{A}_1[M_1'], \mathbf{X}) \leq D_1, \quad d(\mathbf{A}_2[M_2'], \mathbf{X}) \leq D_2,$
   $d(\mathbf{A}_0[M_0'][M_1'][M_2'], \mathbf{X}) \leq D_0.$

Suppose that the codebooks and source inputs of both procedures are the same. Then,

$$\Pr\{\|t_{\mathbf{X}_0 \mathbf{X}_1 \mathbf{X}_2} - w_A\| < \epsilon\}$$
$$\geq \Pr\{M_0 M_1 M_2 < M_0' M_1' M_2'\}$$
$$= 1 - \Pr\{M_0 M_1 M_2 = M_0' M_1' M_2'\}.$$

Define

$$R_{A,\epsilon}(p_X, \delta)$$
$$\triangleq \min_{f(\cdot)} \mathscr{G}(f, q, r, s)$$
$$\text{subject to } \left\|\sum_x f(x_0, x_1, x_2, x) - w_A(x_0, x_1, x_2)\right\| \geq \epsilon,$$
$$\sum_{x_0,x_1,x_2} f(x_0, x_1, x_2, x) = p_X(x),$$
$$f(x_0, x_1, x_2, x) \geq 0,$$
$$\sum_{x_0,x_1,x_2,x} f(x_0, x_1, x_2, x)d(x, x_i) \leq D_i - \delta,$$
$$i = 0, 1, 2;$$

if the feasible region is empty, we define $R_{A,\epsilon}(p_X, \delta) \triangleq +\infty$. We have $R_{A,\epsilon}(p_X, 0) > R_A(p_X, 0)$, which is due to the

uniqueness of w_A. Let $\epsilon' = \min\left\{\tfrac{1}{2}\epsilon, \tfrac{1}{2}(R_{A,\epsilon}(p_X, 0) - R_A(p_X, 0))\right\}$. By (16),

$$\lim_{n\to\infty} \Pr\{M_0 M_1 M_2 < 2^{n(R_A(p_X,0)+\epsilon')}\} > 1 - \tfrac{\epsilon}{2}.$$

Note that

$$\Pr\{\|t_{\mathbf{X}_0 \mathbf{X}_1 \mathbf{X}_2} - w_A\| < \epsilon\}$$
$$\geq \Pr\{M_0 M_1 M_2 < M_0' M_1' M_2'\}$$
$$\geq \Pr\Big\{M_0 M_1 M_2 < 2^{n(R_A(p_X,0)+\epsilon')} \text{ and }$$
$$M_0' M_1' M_2' \geq 2^{n(R_{A,\epsilon}(p_X,0)-\epsilon')}\Big\}$$
$$> 1 - \Pr\Big\{M_0 M_1 M_2 \geq 2^{n(R_A(p_X,0)+\epsilon')}\Big\} -$$
$$\Pr\Big\{M_0' M_1' M_2' < 2^{n(R_{A,\epsilon}(p_X,0)-\epsilon')}\Big\}.$$

Thus, to establish (17), it suffices to prove

$$\lim_{n\to\infty} \Pr\Big\{M_0' M_1' M_2' < 2^{n(R_{A,\epsilon}(p_X,0)-\epsilon')}\Big\} \leq \tfrac{\epsilon}{2}. \quad (35)$$

Let $(\mathbf{Y}_0, \mathbf{Y}_1, \mathbf{Y}_2)$ be the triple of random vectors such that

(i) $(\mathbf{Y}_0, \mathbf{Y}_1, \mathbf{Y}_2)$ is independent of $\mathbf{X}$.
(ii) $(\mathbf{Y}_0[i], \mathbf{Y}_1[i], \mathbf{Y}_2[i])$ is independent of $(\mathbf{Y}_0[j], \mathbf{Y}_1[j], \mathbf{Y}_2[j]), \forall i \neq j$.
(iii) $\Pr\{(\mathbf{Y}_0[i], \mathbf{Y}_1[i], \mathbf{Y}_2[i]) = (x_0, x_1, x_2)\} = r(x_1)s(x_2)q(x_0, x_1, x_2)$.

Since

$$\Pr\{\|t_{\mathbf{Y}_0 \mathbf{Y}_1 \mathbf{Y}_2} - w_A\| \geq \epsilon, d(\mathbf{Y}_i, \mathbf{x}) \leq D_i, i = 0, 1, 2\}$$
$$= \sum_{t:\|t-w_A\|>\epsilon} 2^{-n(I_A(t_{\mathbf{x}},t,D_0,D_1,D_2)+o(1))} \Pr\{t_{\mathbf{Y}_0 \mathbf{Y}_1 \mathbf{Y}_2} = t\}$$
$$= 2^{-n(\min_{w:\|w-w_A\|>\epsilon}\{I_A(t_{\mathbf{x}},w,D_0,D_1,D_2)+\mathscr{D}(w\|qrs)\}+o(1))}$$
$$= 2^{-n(R_{A,\epsilon}(t_{\mathbf{x}},0)+o(1))},$$

where the first equality is derived based on the results in [9], and $\underline{\lim}_{p\to p_X} R_{A,\epsilon}(p, 0) \geq R_{A,\epsilon}(p_X, 0)$, there exists $\xi > 0$ such that for n sufficiently large,

$$\Pr\{M_0' M_1' M_2' < 2^{n(R_{A,\epsilon}(p_X,0)-\epsilon')}\}$$
$$\leq \frac{\epsilon}{2} + \sum_{\mathbf{x}\in T_{n,\xi}(p_X)} \left(\prod_{i=1}^{i=n} p_X(\mathbf{x}[i])\right) \left|\bigcup_{k=1}^{k=2^{n(R_{A,\epsilon}(p_X,0)-\epsilon')}} \mathscr{M}_k\right| \times$$
$$\Pr\{\|t_{\mathbf{Y}_0 \mathbf{Y}_1 \mathbf{Y}_2} - w_A\| \geq \epsilon,$$
$$d(\mathbf{Y}_i, \mathbf{x}) \leq D_i, i = 0, 1, 2\}$$
$$\overset{(a)}{\leq} \frac{\epsilon}{2} + 2^{n(R_{A,\epsilon}(p_X,0)-\epsilon'+o(1))} 2^{-n(R_{A,\epsilon}(p_X,0)-\frac{\epsilon'}{2}+o(1))}$$
$$\to \frac{\epsilon}{2},$$

where (a) follows from Lemma 1. This completes the proof of (35)[‡]. ∎

APPENDIX D

Proof of Proposition 8: Let D_0, D_1, D_2 be values such that $R_B(p_X, q, r, s, \lambda_0, \lambda_1, \lambda_2, \overline{D}) = R_A(p_X, q, r, s, D_0,$

[‡]If $\underline{\lim}_{p\to p_X} R_{A,\epsilon}(p, 0) = R_{A,\epsilon}(p_X, 0) = +\infty$, then $\exists \eta$, if $\|t_{\mathbf{x}} - p_X\| \leq \eta$, $\Pr\{M_0 M_1 M_2 = M_0' M_1' M_2' | \mathbf{X} = \mathbf{x}\} = 0$.

$D_1, D_2)$ and $\sum_{i=0}^{i=2} \lambda_i D_i \leq \overline{D}$ [§]. To prove (18), we only need to prove

$$\Pr\left\{\sum_{i=0}^{i=2} \lambda_i d(\mathbf{X}, \mathbf{X}_i) < \overline{D} - \epsilon\right\} < \epsilon/2, \qquad (36)$$

and

$$\Pr\left\{\sum_{i=0}^{i=2} \lambda_i d(\mathbf{X}, \mathbf{X}_i) \leq \overline{D} + \epsilon\right\} > 1 - \epsilon/2, \qquad (37)$$

as $n \to \infty$. The proof of (36) is similar to that of (28) and is omitted. For (37), note that

$$\Pr\left\{\sum_{i=0}^{i=2} \lambda_i d(\mathbf{X}, \mathbf{X}_i) \leq \overline{D} + \epsilon\right\}$$

$$\geq \Pr\left\{\exists(m_0, m_1, m_2), m_0 m_1 m_2 \leq 2^{n R_A(p_X, q, r, s, D_0, D_1, D_2)},\right.$$

$$d(\mathbf{X}, \mathbf{A}_i[m_i]) \leq D_i + \frac{\epsilon}{\sum_{i=0}^{i=2} \lambda_i}, i = 1, 2,$$

$$\left. d(\mathbf{X}, \mathbf{A}_0[m_0][m_1][m_2]) \leq D_0 + \frac{\epsilon}{\sum_{i=0}^{i=2} \lambda_i}\right\},$$

which tends to 1 as $n \to \infty$ according to the proof of (29). Thus, (18) is proved.

We proceed to prove (19). Due to (18), $\forall \delta(0 < \delta < \epsilon)$, if n is sufficiently large,

$$\Pr\{\|t_{\mathbf{X}_0 \mathbf{X}_1 \mathbf{X}_2} - w_B\| < \epsilon\}$$
$$\geq (1 - \delta) \times$$
$$\Pr\left\{\|t_{\mathbf{X}_0 \mathbf{X}_1 \mathbf{X}_2} - w_B\| < \epsilon \,\middle|\, \left|\overline{D} - \sum_{i=0}^{i=2} \lambda_i d(\mathbf{X}, \mathbf{X}_i)\right| \leq \delta\right\}. \qquad (38)$$

Define

$$R_{B,\epsilon}(p_X, \delta)$$
$$\triangleq \min_{f(\cdot)} \mathscr{G}(f, q, r, s)$$
$$\text{subject to } \left\|\sum_x f(x_0, x_1, x_2, x) - w_B(x_0, x_1, x_2)\right\| \geq \epsilon,$$
$$\sum_{i=0}^{i=2} \sum_{x_0, x_1, x_2, x} f(x_0, x_1, x_2, x) \lambda_i d(x, x_i) \leq \overline{D} + \delta,$$
$$\sum_{x_0, x_1, x_2} f(x_0, x_1, x_2, x) = p_X(x),$$
$$f(x_0, x_1, x_2, x) \geq 0,$$

and $R_{B,\epsilon}(p_X, \delta) = +\infty$ if the feasible region is empty. If $\exists \delta_1(0 < \delta_1 < \epsilon)$ such that $R_{B,\epsilon}(p_X, \delta_1) = +\infty$, then according to (38), eventually, $\Pr\{\|t_{\mathbf{X}_0 \mathbf{X}_1 \mathbf{X}_2} - w_B\| < \epsilon\} \geq 1 - \delta_1 \geq 1 - \epsilon$, and (19) is proved.

Next we consider the case in which $R_{B,\epsilon}(p_X, \delta) < +\infty$, $0 < \delta < \epsilon$. In this case, we actually have $R_{B,\epsilon}(p_X, \delta) < +\infty$, $\forall \delta > 0$, since $R_{B,\epsilon}(p_X, \delta)$ is at least non-increasing in δ. Furthermore, by using upper and lower limits, we find that

$$\lim_{\delta \to 0^+} R_{B,\epsilon}(p_X, \delta) = R_{B,\epsilon}(p_X, 0). \qquad (39)$$

By Proposition 4(i), $R_{B,\epsilon}(p_X, \delta) > R_{B,0}(p_X, \delta) = R_B(p_X, q, r, s, \lambda_0, \lambda_1, \lambda_2, \overline{D} + \delta)$. Thus, (39) implies that $\exists \delta_2$

[§]Actually, the equality holds due to Proposition 6.

$(0 < \delta_2 < \epsilon)$ such that $R_{B,\epsilon}(p_X, 0) \geq R_{B,\epsilon}(p_X, \delta_2) > R_{B,0}(p_X, 0)$.

Suppose there is another procedure SearchWithFixed-Weight' same as SearchWithFixedWeight except that it returns the integer triple (M'_0, M'_1, M'_2) which satisfies the following equation:

$$(M'_0, M'_1, M'_2)$$
$$= \text{argmin}_{(k_0, k_1, k_2)}\Big\{\lambda_0 d(\mathbf{X}, \mathbf{A}_0[k_0][k_1][k_2]) +$$
$$\sum_{i=1}^{i=2} \lambda_i d(\mathbf{X}, \mathbf{A}_i[k_i]),$$
$$\text{subject to } k_0 k_1 k_2 \leq M,$$
$$\|t_{\mathbf{A}_0[k_0][k_1][k_2] \mathbf{A}_1[k_1] \mathbf{A}_2[k_2]} - w_B\| \geq \epsilon\Big\}.$$

Then,

$$\Pr\{\|t_{\mathbf{X}_0 \mathbf{X}_1 \mathbf{X}_2} - w_B\| < \epsilon\}$$
$$\geq \Pr\Big\{\sum_{i=0}^{i=2} \lambda_i d(\mathbf{X}, \mathbf{X}_i) < \lambda_0 d(\mathbf{X}, \mathbf{A}_0[M'_0][M'_1][M'_2]) +$$
$$\sum_{i=1}^{i=2} \lambda_i d(\mathbf{X}, \mathbf{A}_i[M'_i])\Big\}$$
$$\geq 1 - \Pr\left\{\sum_{i=0}^{i=2} \lambda_i d(\mathbf{X}, \mathbf{X}_i) \geq \overline{D} + \frac{\delta_2}{2}\right\}$$
$$- \Pr\Big\{\lambda_0 d(\mathbf{X}, \mathbf{A}_0[M'_0][M'_1][M'_2]) + \sum_{i=1}^{i=2} \lambda_i d(\mathbf{X}, \mathbf{A}_i[M'_i])$$
$$\leq \overline{D} + \frac{\delta_2}{2}\Big\}.$$

In view of (37), to prove (19) it suffices to show that

$$\lim_{n \to \infty} \Pr\Big\{\lambda_0 d(\mathbf{X}, \mathbf{A}_0[M'_0][M'_1][M'_2]) + \sum_{i=1}^{i=2} \lambda_i d(\mathbf{X}, \mathbf{A}_i[M'_i])$$
$$\leq \overline{D} + \frac{\delta_2}{2}\Big\} < \epsilon, \qquad (40)$$

which can be proved in a way similar to that for (35). Thus, (19) is proved. ∎

REFERENCES

[1] S. Jalali and T. Weissman, "Multiple description coding of discrete ergodic sources," in *Proc. 2009 Allerton Conf. Commun. Contr. Comput.*, pp. 1256–1261.

[2] A. György, T. Linder, and G. Lugosi, "Tracking the best quantizer," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1604–1625, Apr. 2008.

[3] R. Zamir and K. Rose, "Natural type selection in adaptive lossy compression," *IEEE Trans. Inf. Theory*, vol. 47, no. 1, pp. 99–111, Jan. 2001.

[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Akadémiai Kiadó, 1986.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[6] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 851–857, Nov. 1982.

[7] P. Elias, "Universal codeword sets and representations of the integers," *IEEE Trans. Inf. Theory*, vol. 21, pp. 194–203, 1975.

[8] T. M. Apostol, *Introduction to Analytic Number Theory*. Springer-Verlag, 1976.

[9] Z. Zhang, E. H. Yang, and V. Wei, "The redundancy of source coding with a fidelity criterion–part one: known statistics," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, pp. 71–91, Jan. 1997.

[10] 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, no. 6, pp. 1643–1653, June 2010.

[11] S. Dumitrescu, J. Chen, and Y. Zhang, "Multiple description coding for binary sources," in *Proc. 2009 Int. Conf. Wireless Commun. Signal Process.*

[12] Y. Zhang, S. Dumitrescu, J. Chen, and Z. Sun, "Multiple description coding based on LDGM codes," in *Proc. 2009 IEEE Inf. Theory Workshop*, pp. 85–89.

[13] S. -Y. Chung, "On the construction of some capacity-approaching coding schemes," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[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] D. P. Bertsekas, *Nonlinear Programming*, 2nd edition. Athena Scientific, 1999.

Yuhua Fan received the B.Sc. and M.S. degrees from Shanghai Jiao Tong University, China, in 2005 and 2008, respectively, both in electronic engineering. He is currently pursuing the Ph.D. degree at the same university. His research interests include image and video coding and processing.

Jia Wang received the B.Sc. degree in electronic engineering, the M.S. degree in pattern recognition and intelligence control, and the Ph.D. degree in electronic engineering from Shanghai Jiao Tong University, China, in 1997, 1999, and 2002, respectively. He is currently an Associate Professor of the Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, and a member of the Shanghai Key Laboratory of Digital Media Processing and Transmission. His research interests include multiuser information theory and its application in video coding.

Jun Sun (M'02) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, in 1989 and 1992, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, China, in 1995. As a Professor, he is now with the Institute of Image Communication and Information Processing of Shanghai Jiao Tong University. His research interests include image communication, HDTV, and mobile communication.

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 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.

He has received several awards for his research, including the Josef Raviv Memorial Postdoctoral Fellowship in 2006, the Early Researcher Award from the Province of Ontario in 2010, and the IBM Faculty Award in 2010.