An Introduction to Multiple Description Coding

Sheng Jing

October 16, 2006

In this report, we formulate the Multiple Description (MD) coding problem, which was created couple of decades ago. Significant research effort has been devoted to completely characterizing the rate-distortion region of the MD coding problem. Various Information-Theoretical approaches to this problem have been taken, which generate beautiful results. We summarize part of the research effort in this report.

1 Introduction

The MD coding problem was not created as a pure information-theoretical puzzle. As the author stated in [1], “Multiple Description coding has come full circle from explicit practical motivation to theoretical novelty and back to engineering application”. MD coding was invented at Bell Laboratories during the 1970s in connection with communicating speech over the telephone network. At that time, though the telephone network enjoys good reliability, outages of transmission is inevitable, mainly due to device failures, routine maintenance or upgrades. Rather than diverting calls to standby transmission links in case of transmission outage, it may be clever to split the information from a single call onto two separate links or paths. Some early attempts by channel splitting are summarized in [1].

The channel splitting idea inspires the following question: “If an information source is described with two separate descriptions, what are the concurrent limitations on qualities of these descriptions taken separately and jointly?” [1]. This question eventually came to be known as the MD coding problem.

Before presenting the problem formulation of MD coding, we shall first introduce the basic definitions of rate-distortion theory and state Shannon’s rate-distortion theorem.

1.1 Shannon’s Rate-Distortion Theory

When information is transferred over a channel at a rate above the channel’s capacity, distortion in the recovery of the information is inevitable. The branch of information theory devoted to characterize the relationship between achievable distortion and required rate is called the rate-distortion theory. The most important result in the rate-distortion theory is perhaps Shannon's rate-distortion theorem [2], which is restated in this section.

We assume that $X_i, i = 1, 2, \cdots$ is a sequence of i.i.d. discrete random variables drawn according to a common probability mass function $p(x), x \in \mathcal{X}$. We are given a reconstruction space $\hat{\mathcal{X}}$ together with an associated distortion measure

$$d: \qquad \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}. \tag{1}$$

A description of $\underline{x} \in \hat{\mathcal{X}}^n = \hat{\mathcal{X}} \times \cdots \times \hat{\mathcal{X}}$ is a map i: $\hat{\mathcal{X}}^n \to \{1, \cdots, 2^{nR}\}$, where R is the rate of description in bits per source symbol of $\underline{x}$. A reconstruction of $\underline{x}$ is a map $\hat{x}$: $\{1, \cdots, 2^{nR}\} \to \hat{\mathcal{X}}^n$. The distortion incurred through this pair of description and reconstruction is defined by

$$d^n \quad = \quad \mathbf{E}\left[\frac{1}{n}\sum_{k=1}^{n} d\left(X_k, \hat{x}_k(i(\underline{X}))\right)\right]. \tag{2}$$

The distortion d is said to be achievable with rate R for the source sequence $\{X_i\}_{i=1}^n$ if for $n = 1, 2, \cdots$, there exists a sequence of rate R descriptions i: $\hat{\mathcal{X}}^n \to \{1, \cdots, 2^{nR}\}$ and reconstructions $\hat{x}$: $\{1, \cdots, 2^{nR}\} \to \hat{\mathcal{X}}^n$ such that $d^n \leq d$, for all n sufficiently large.

Rate-Distortion Function The rate-distortion function $R(d)$ is the infimum of all rates R achieving distortion d on a given stochastic process $\{X_i\}_{i=1}^{\infty}$.

Theorem 1.1 (Shannon's Rate-Distortion Theorem [2])**.** *If $\{X_i\}_{i=1}^{\infty}$ are i.i.d. discrete finite alphabet random variables with probability mass function $p(x)$, then*

$$R(d) \quad = \quad \inf_{\mathcal{P}(d)} I(X; \hat{X}), \tag{3}$$

where

$$\mathcal{P}(d) \quad = \quad \left\{ p(\hat{x}|x) : \sum_{x,\hat{x}} p(x)\, p(\hat{x}|x)\, d(x, \hat{x}) \leq d \right\}. \tag{4}$$

We can calculate the rate-distortion function for several special sources and distortion measures.

Corollary 1.2 (Bernoulli Source with Hamming Distortion)**.** *The rate-distortion function for a Bernoulli(α) Source with Hamming Distortion is*

$$R(d) \quad = \quad \begin{cases} H(\alpha) - H(d), & 0 \leq d \leq \min\{\alpha, 1-\alpha\}, \\ 0, & d > \min\{\alpha, 1-\alpha\}, \end{cases} \tag{5}$$

where $H(\cdot)$ is the entropy function of a binary random variable.

Corollary 1.3 (Gaussian Source with Squared Error Distortion)**.** *The rate-distortion function for a $\mathcal{N}(0, \sigma^2)$ source with squared error distortion is*

$$R(d) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{d}, & 0 \le d \le \sigma^2, \\ 0, & d > \sigma^2. \end{cases} \tag{6}$$

2 MD Coding: Problem Formulation

The 2-channel 3-receiver MD coding problem is represented in Figure 1.

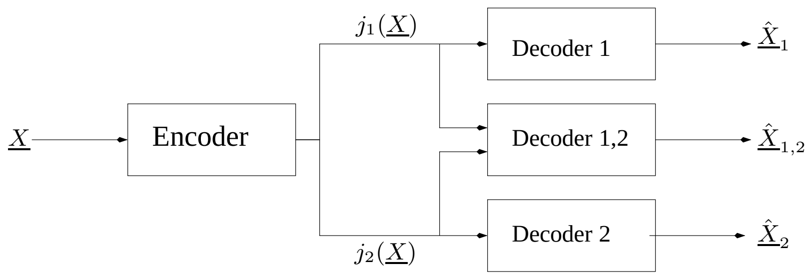

Figure 1: 2-Channel 3-Receiver MD Coding Problem

The encoder is presented with a sequence of i.i.d. source symbols $\{X_i\}_{i=1}^{\infty}$. Each source symbol is distributed according to a probability mass function $p(x), x \in \mathcal{X}$. We are given three reconstruction spaces $\hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2, \hat{\mathcal{X}}_{1,2}$, together with the associated distortion measures

$$d_t: \quad \mathcal{X} \times \hat{\mathcal{X}}_t \to \mathbb{R}, \qquad t = 1, 2, \{1,2\}. \tag{7}$$

The distortion measure on n-sequences is defined by the average per-symbol distortion

$$d_t^n(\underline{x}, \underline{\hat{x}}_t) = 1/n \sum_{k=1}^{n} d_t(x_k, \hat{x}_{tk}), \quad t = 1, 2, \{1,2\}, \tag{8}$$

where $\underline{x} = \{x_1, \cdots, x_n\} \in \mathcal{X}^n$ and $\underline{\hat{x}}_t = \{\hat{x}_{t1}, \cdots, \hat{x}_{tn}\} \in \hat{\mathcal{X}}_t^n$. The encoding and decoding functions are defined by

$$f_t: \quad \mathcal{X}^n \to \{1, \cdots, M_t\}, \quad t = 1, 2 \tag{9}$$

$$g_t: \quad \{1, \cdots, M_t\} \to \hat{\mathcal{X}}_t^N, \quad t = 1, 2 \tag{10}$$

$$g_{1,2}: \quad \{1, \cdots, M_1\} \times \{1, \cdots, M_2\} \to \hat{\mathcal{X}}_{1,2}. \tag{11}$$

Denote $\underline{X} = (X_1, \cdots, X_n) \in \mathcal{X}^n$. Define

$$\begin{aligned} \underline{\hat{X}}_t &= g_t(f_t(\underline{(}X))), \quad t = 1, 2 & (12) \\ \underline{\hat{X}}_{1,2} &= g_{1,2}(f_1(\underline{(}X)), f_2(\underline{(}X))), & (13) \end{aligned}$$

and

$$D_t = \mathbf{E}\left[d_t^n(\underline{X}, \underline{\hat{X}}_t)\right], \quad t = 1, 2, \{1, 2\}. \tag{14}$$

The quintuple $(f_1, f_2, g_1, g_2, g_{1,2})$ is called a code with parameter $(n, M_1, M_2, D_0, D_1, D_{1,2})$.

Achievable Rate-Distortion Vector We shall say (R_1, R_2) is achievable for distortion $\underline{d} = (d_1, d_2, d_{1,2})$ if, for all $\epsilon > 0$, there exists for n sufficiently large a code with parameters $(n, M_1, M_2, D_0, D_1, D_{1,2})$, where

$$\begin{aligned} M_t &< 2^{(R_t+\epsilon)n}, \quad t = 1, 2 & (15) \\ D_t &< d_t + \epsilon, \quad t = 1, 2, \{1, 2\}. & (16) \end{aligned}$$

Rate-Distortion Region The rate-distortion region $\mathcal{R}(\underline{d})$ for distortion $\underline{d} = (d_1, d_2, d_{1,2})$ is the closure of the set of achievable rate vectors (R_1, R_2) inducing distortions $\leq \underline{d}$.

Achievable Rate-Distortion Region Any subset of the rate-distortion region is called an achievable rate-distortion region. Another common name for achievable rate-distortion region is inner bound to the rate-distortion region.

3 Results on Achievable Rate-Distortion Region

The following two sets of sufficient conditions for $(R_1, R_2, d_1, d_2, d_{1,2})$ to be achievable was deduced by El Gammal and Cover in [4] which will be later referred to as the EGC* resp. EGC region.

Theorem 3.1 (EGC* Achievable RD Region)**.** *Let $X_1, X_2, \cdots$ be a sequence of i.i.d. finite alphabet random variables drawn according to a probability mass function $p(x)$. Let $d_m(\cdot, \cdot)$ be bounded. An achievable rate distortion region for distortion $\underline{d} = (d_1, d_2, d_{1,2})$ is given by the convex hull of all (R_1, R_2) such that*

$$\begin{aligned} R_1 &\geq I(X; U), & (17) \\ R_2 &\geq I(X; V), & (18) \\ R_1 + R_2 &\geq I(X; U, V) + I(U; V), & (19) \end{aligned}$$

for some random variables U and V jointly distributed with a generic source random variable X such that there *exist random variables of the forms,*

$$\hat{X}_1 = g_1(U), \tag{20}$$

$$\hat{X}_2 = g_2(V), \tag{21}$$

$$\hat{X}_{1,2} = g_{1,2}(U, V), \tag{22}$$

such that $\mathbf{E}[d_t(X, \hat{X}_t)] \leq d_t, t = 1, 2, \{1, 2\}$.

Theorem 3.2 (EGC Achievable RD Region)**.** *The quintuple* $(R_1, R_2, d_1, d_2, d_{1,2})$ *is achievable if there exist random variables* $\hat{X}_1, \hat{X}_2, \hat{X}_{1,2}$ *jointly distributed with a generic source random variable* X *such that*

$$R_t \geq I(X; \hat{X}_t), \quad t = 1, 2, \tag{23}$$

$$R_1 + R_2 \geq I(X; \hat{X}_1, \hat{X}_2, \hat{X}_{1,2}) + I(\hat{X}_1; \hat{X}_2), \tag{24}$$

$$d_t \geq \mathbf{E}[d_t(X, \hat{X}_t)], \quad t = 1, 2, \{1, 2\}. \tag{25}$$

Let $\mathcal{R}_{EGC^\star}$ resp. $\mathcal{R}_{EGC}$ denote the $EGC^\star$ resp. EGC achievable rate-distortion region. Actually, $\mathcal{R}_{EGC^\star}$ and $\mathcal{R}_{EGC}$ are closely related, as stated in the following theorem.

Theorem 3.3.

$$\mathcal{R}_{EGC^\star} \subset \mathcal{R}_{EGC}$$

The EGC* region is also included here since it (and also the EGC region) turns out to be optimal in the special case of Gaussian source and squared error distortion measure, which will be summarized in the following.

3.1 Special Cases: Gaussian Source with Squared Error Distortion

For the special case of Gaussian source with squared error distortion, the MD coding rate-distortion region was preliminarily deduced in [3] and [4]. The authors of [5] fixed some remaining inconsistencies and characterized the entire rate-distortion region.

Theorem 3.4 (MD Coding RD Region: Gaussian Source with Squared Error Distortion [3] [4] [5])**.** *For i.i.d. Gaussian source sequence,* $X_i \sim \mathcal{N}(0, \sigma^2)$*, with squared error distortion measure, the MD coding rate-distortion region for Figure 1 is the set of quintuples* $(R_1, R_2, d_1, d_2, d_{1,2})$ *satisfying the following conditions.*

(1). Given that $0 \le d_{1,2} \le d_1 + d_2 - \sigma^2$*, then the rate pair* (R_1, R_2) *is achievable if*

$$R_1 \geq \frac{1}{2}\log\frac{1}{d_1} \tag{26}$$

$$R_2 \geq \frac{1}{2}\log\frac{1}{d_2} \tag{27}$$

$$R_1 + R_2 \geq \frac{1}{2}\log\frac{1}{d_{1,2}} \tag{28}$$

(2). Given that $d_1 + d_2 - \sigma^2 \le d_{1,2} \le \left(\frac{1}{d_1} + \frac{1}{d_1} - \frac{1}{\sigma^2}\right)^{-1}$*, then the rate pair* (R_1, R_2) *is achievable if*

$$R_1 \geq \frac{1}{2}\log\frac{1}{d_1} \tag{29}$$

$$R_2 \geq \frac{1}{2}\log\frac{1}{d_2} \tag{30}$$

$$R_1 + R_2 \geq \frac{1}{2}\log\frac{1}{d_{1,2}} + \frac{1}{2}\log\frac{(\sigma^2 - d_{1,2})^2}{(\sigma^2 - d_{1,2})^2 - \left(\sqrt{(\sigma^2 - d_1)(\sigma^2 - d_2)} - \sqrt{(d_1 - d_{1,2})(d_2 - d_{1,2})}\right)^2} \tag{31}$$

(3). Given that $\left(\frac{1}{d_1} + \frac{1}{d_1} - \frac{1}{\sigma^2}\right)^{-1} \le d_{1,2} \le +\infty$*, then the rate pair* (R_1, R_2) *is achievable if*

$$R_1 \geq \frac{1}{2}\log\frac{1}{d_1} \tag{32}$$

$$R_2 \geq \frac{1}{2}\log\frac{1}{d_2} \tag{33}$$

Remark: For Gaussian sources and squared error distortion measure, Theorem 3.4 not only characterizes an achievable rate-distortion region but also states that this region is the best achievable rate-distortion region. The rate-distortion region in Theorem 3.4 is obtained by evaluating and optimizing the EGC* rate-distortion region in Theorem 3.1. The converse part (showing the optimality of the region) was presented in [3]. The main technicality of the converse part shows a lower bound to the mutual information between the two side receiver's reconstructions (Figure 1), $I(\hat{X}_1^n; \hat{X}_2^n)$, for a given rate pair (R_1, R_2) and side receiver's distortions (d_1, d_2). Denote the reconstructions at the two side receivers as $\hat{X}_1^n$ and $\hat{X}_2^n$. The lower bound is,

$$I(\hat{X}_1^n; \hat{X}_2^n) \geq \frac{n}{2}\log\frac{\sigma^2}{\sigma^2 - \left(\sqrt{(\sigma^2 - d_1)(\sigma^2 - d_2)} - \sqrt{(d_1 - d_{1,2})(d_2 - d_{1,2})}\right)^2} \tag{34}$$

The above lower bound is the cause for the tradeoff between the central receiver and side receivers. On the one hand, if the central decoder needs to approach Shannon's rate-distortion bound (Theorem 1.1), the two descriptions need to be approximately independent, which renders the two side reconstructions also approximately independent. Therefore, the approximate independence condition $I(\hat{X}_1^n; \hat{X}_2^n) \approx 0$ necessitates that $\sigma^2 + d_{1,2} \approx d_1 + d_2$, which means at least one of d_1 and d_2 is close to the source's variance (i.e. performs poorly). On the other hand, if both side decoders need to approach Shannon's rate-distortion bound, then the above lower bound on $I(\hat{X}_1^n; \hat{X}_2^n)$ is far away from zero, which means that the two descriptions sent over two channels are highly correlated. Therefore, the central decoder's distortion performance is bounded away from the rate-distortion bound since receiving two highly correlated descriptions is not much better than receiving only one description.

4 More Results on Achievable Rate-Distortion Region

It would be nice if the optimality of the EGC achievable rate-distortion region is also true in more general cases other than the Gaussian source with squared error distortion case. Unfortunately, the EGC region is not always the exact rate-distortion region. Towards this end, Zhang and Berger proposed the following achievable rate-distortion region in [6]. In [6], a concrete example was also constructed to show that the EGC region is strictly non-optimal.

Theorem 4.1 (Berger's Achievable Rate-Distortion Region)**.** *A rate pair (R_1, R_2) is achievable for the distortion vector $\underline{d} = (d_1, d_2, d_{1,2})$ if there exist random variables $\hat{X}_0, \hat{X}_1, \hat{X}_2$ jointly distributed with a generic source random variable X such that*

$$R_1 \geq I(X; \hat{X}_0, \hat{X}_1), \tag{35}$$

$$R_2 \geq I(X; \hat{X}_0, \hat{X}_2), \tag{36}$$

$$R_1 + R_2 \geq 2I(X; \hat{X}_0) + I(\hat{X}_1; \hat{X}_2 | \hat{X}_0) + I(X; \hat{X}_1, \hat{X}_2 | \hat{X}_0), \tag{37}$$

and there exist $\phi_1, \phi_2, \phi_{1,2}$ which satisfy

$$\mathbf{E}\left[d(X, \phi_t(\hat{X}_0, \hat{X}_t))\right] \leq d_t, \quad t = 1, 2, \tag{38}$$

$$\mathbf{E}\left[d(X, \phi_{1,2}(\hat{X}_0, \hat{X}_t, \hat{X}_2)))\right] \leq d_{1,2}. \tag{39}$$

Remark The example in [6], which shows the non-optimality of the EGC region, deals with the binary symmetric Hamming problem. Define the the following two functions,

$$\bar{R}_{EGC}(D) = \inf_{(R_1, R_2, d_1, d_2, d_{1,2}) \in \bar{\mathcal{R}}_{EGC}} \left\{ R_1 + R_2 : d_{1,2} = 0, d_1 + d_2 \leq 2D \right\}, \tag{40}$$

$$R_{ZB}(D) = \inf_{(R_1, R_2, d_1, d_2, d_{1,2}) \in \mathcal{R}_{ZB}} \left\{ R_1 + R_2 : d_{1,2} = 0, d_1 + d_2 \leq 2D \right\}, \tag{41}$$

where $\bar{\mathcal{R}}_{EGC}$ is the (closure) EGC region and $\mathcal{R}_{ZB}$ is Berger's region. In [6], $\bar{R}_{EGC}(D)$ is explicitly calculated to take the following form,

$$\bar{R}_{EGC}(D) \;=\; 2H(1/2 + D) - (1/2)H(4D) - 2D. \tag{42}$$

In [6], an upper bound on $R_{ZB}(D)$ is also derived. Comparing $\bar{R}_{EGC}(D)$ and the upper bound on $R_{ZB}(D)$ verifies the non-optimality of the EGC region.

Though Berger's region exceeds the EGC region in the constructed example, no concrete evidence shows that the former region subsumes the latter region. More recently, the authors in [7] presented a new achievable rate-distortion region, which is claimed to subsume both the EGC region and Berger's region. The main theorem in [7] is for the L-channel $(2^L - 1)$-receiver MD coding problem. We present in the following its simplified version for the 2-channel 3-receiver MD coding problem.

Theorem 4.2 (Goyal's Achievable Rate-Distortion Region)**.** *The rate vector* (R_1, R_2) *is achievable for the distortion requirement* $\underline{d} = (d_1, d_2, d_{1,2})$ *if there exist random variables* $\hat{X}_0, \hat{X}_1, \hat{X}_2, \hat{X}_{1,2}$ *(*$\hat{X}_0$ *takes value in some finite alphabet* $\mathcal{X}_0$ *and each* $\hat{X}_t$ *takes value in the reconstruction alphabet* $\mathcal{X}_t$*,* $t = 1, 2, \{1, 2\}$*) such that*

$$R_1 \;\geq\; I(X; \hat{X}_0, \hat{X}_1), \tag{43}$$

$$R_2 \;\geq\; I(X; \hat{X}_0, \hat{X}_2), \tag{44}$$

$$R_1 + R_2 \;\geq\; 2I(X; I(X; \hat{X}_0) + I(\hat{X}_1; \hat{X}_2|\hat{X}_0) + I(X; \hat{X}_1, \hat{X}_2|\hat{X}_0), \tag{45}$$

$$d_t \;\geq\; \mathbf{E}\left[d_t(X, \hat{X}_t)\right], \quad t = 1, 2, \{1, 2\}. \tag{46}$$

Remark By letting X_0 in Theorem 4.2 be a constant, the Goyal's achievable rate-distortion region reduces to the EGC region. Therefore, Goyal's region subsumes the EGC region. Furthermore, simple argument shows that Goyal's region also contains all points in Berger's region. However, it is not known yet if Goyal's region strictly improves Berger's region or not.

5 L-Channel MD Coding

Up to now, we have focused on the 2-channel 3-receiver MD coding problem. The generalization to L-channel $(2^L - 1)$-receiver MD coding problem (Figure 2) is a natural next step. In the L-channel $(2^L - 1)$-receiver MD coding problem, there are L encoding functions $f_l(\cdot) : \mathcal{X} \to \{1, \cdots, M_l\}, l \in \mathcal{L} = \{1, \cdots, L\}$. The L descriptions $j_l = f_l(\underline{X}), l \in \mathcal{L}$ are sent over L channels. There are $2^L - 1$ receivers, denoted as $g_{\mathcal{K}}(\cdot)$, for each

$\mathcal{K} \in \mathcal{L}$ and $\mathcal{K} \neq \emptyset$. The reconstruction at each receiver is represented by

$$\underline{\hat{X}}_{\mathcal{K}} \quad = \quad g_{\mathcal{K}}(j_k : k \in \mathcal{K}). \tag{47}$$

The distortions $d_{\mathcal{K}}^n$ are also defined as the average per-symbol distortion. The rate-distortion region of the L-channel $(2^L - 1)$-receiver MD coding problem is composed of all achievable length-$(L + (2^L - 1))$ vectors $(R_l : l \in \mathcal{L}; \quad d_{\mathcal{K}} : \mathcal{K} \subset \mathcal{L}, \mathcal{K} \neq \emptyset)$.

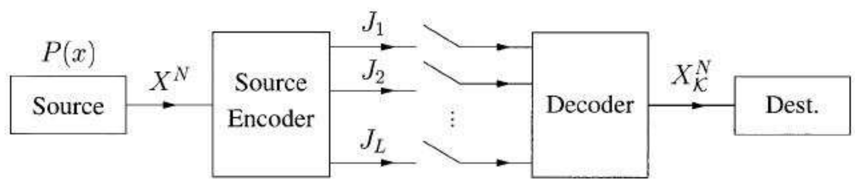

Figure 2: L-Channel $(2^L - 1)$-Receiver MD Coding Problem

In [7], the authors presented an achievable rate-distortion region for the L-channel $(2^L - 1)$-receiver MD coding problem.

Theorem 5.1 (L-Channel MD Coding: Achievable Rate-Distortion Region [7])**.** *Let $X_{(2^{\mathcal{L}})}$ be any set of 2^L random variables jointly distributed with X, where X_0 takes value in some finite alphabet $\hat{\mathcal{X}}_0$ and each $X_{\mathcal{K}}$ takes value in the reproduction alphabet $\hat{\mathcal{X}}_{\mathcal{K}}, \mathcal{K} \neq \emptyset$. Then the rate-distortion region contains the rates and distortions satisfying*

$$d_{\mathcal{K}} \quad \geq \quad \mathbf{E}\left[d_{\mathcal{K}}(X, X_{\mathcal{K}})\right], \tag{48}$$

$$R_{\mathcal{K}} \quad \geq \quad (|\mathcal{K}| - 1)I(X; X_0) - H(X_{(2^{\mathcal{K}})}|X) + \sum_{\mathcal{M} \subset \mathcal{K}} H(X_{\mathcal{M}}|X_{(2^{\mathcal{M}} - \{\mathcal{M}\})}). \tag{49}$$

6 Conclusions

The Multiple Description Coding problem is recapped in this report. Historical results are summarized and restated. Most of the references cited in this report are on the theoretical side of the MD coding problem. Though beautiful results exist, the rate-distortion region of the MD coding problem still remains an open problem at the current stage.

References

[1] Vivek K. Goyal, *Multiple Description Coding: Compression Meets the Network*, IEEE Signal Processing Magazine, pages 74-93, September 2001;

[2] C. E. Shannon, *Coding Theorems for a Discrete Source with a Fidelity Criterion*, IRE Nat. Conv. Rec. Mar. 1959; reprinted in *Key Papers in the Development of Information Theory*, New York; IEEE;

[3] L. Ozarow, *On a Source-Coding Problem with Two Channels and Three Receivers*, The Bell System Technical Journal, Vol. 59, No. 10, December 1980;

[4] Abbas A. El Gammal, Thomas M. Cover, *Achievable Rates for Multiple Descriptions*, IEEE Transaction on Information Theory, Vol. 28, No. 6, November 1982;

[5] Hanying Feng, Michelle Effros, *On the achievable region for multiple description source codes on Gaussian sources*, Proceedings. IEEE International Symposium on Information Theory, 29 June-4 July 2003;

[6] Zhen Zhang, Toby Berger, *New Results in Binary Multiple Descriptions*, IEEE Transaction on Information Theory, Vol. 33, No. 4, July 1987;

[7] Raman Venkataramani, Gerhard Kramer, Vivek K. Goyal, *Multiple Description Coding with Many Channels*, IEEE Transaction on Information Theory, Vol. 49, No. 9, September 2003;

[8] J. K. Wolf, A. D. Wyner and J. Ziv, *Source Coding for Multiple Descriptions*, The Bell System Technical Journal, Vol. 59, No. 8, October 1980;