Multiple Description Coding:

Typical Arguments and Random Binning

Charles Swannack

Abstract

We reexamine the problem of Multiple Description coding considered last week. We briefly review the import concepts and results. An overview of the basic methods of proof are provided. We further formulate this problem in the context of distributed source coding and from this framework examine a practical coding scheme for the symmetric problem provided by Pradhan et. al.

Introduction

The problem of multiple description (MD) coding has roots in both practical and theoretical communities. This problem while theoretically challenging also provides a valuable framework for reliable transmission over a packet network.

Today we will not only overview the methods that are used to prove the achievability of the best know MD rate-distortion regions, but also overview a constructive approach to the problem. The MD problem may be roughly described as follows [1]:

> *Suppose that we wish to describe a stochastic process through a communication network. Therefore, we send L descriptions in hopes that a few of them will get through. However, due to link failures, only a random subset of the descriptions get through. What is the region of achievable distortions for every possible set of link failures?*

A depiction of the general problem may be seen in Figure 1. In general we will consider a discrete source that produces a sequence symbols $\{X_i\}_{i=i}^{\infty}$ that are independent and identically-distributed (*i.i.d.*) with respect to the probability mass function $p(x)$, for $x \in \mathcal{X}$ for some given alphabet $\mathcal{X}$. We will denote the set of received description indices as $\mathcal{K}$ and denote the corresponding decoder and reconstruction space as $g_{\mathcal{K}}$ and $\hat{\mathcal{X}}_{\mathcal{K}}$. Note that the decoder *and* the reconstruction space depend on the set of received description indices $\mathcal{K}$ and that the set $\mathcal{K}$ is unknown to the encoder.

As noted last week it is not possible in general to achieve the rate-distortion function for every possible subset of indices. Indeed, we saw in Ozarow's Theorem [2] (the three receiver case with a Gaussian source and mean squared distortion) if the distortion on the side decoders is equal to the distortion-rate function $D(R) = \exp -2R$ then the joint decoder can not achieve a distortion that is better than $D(R)/2$. This is far from $D^2(R) = D(R + R)$ if $D(R) \ll 1$. Intuitively, this is due to the fact that if the side decoders achieve the optimal distortion-rate then they must be highly correlated and thus the extra description provides very little information. However, a precise quantitative statement for the general L-channel problem for either a Gaussian or a general source is still not known.

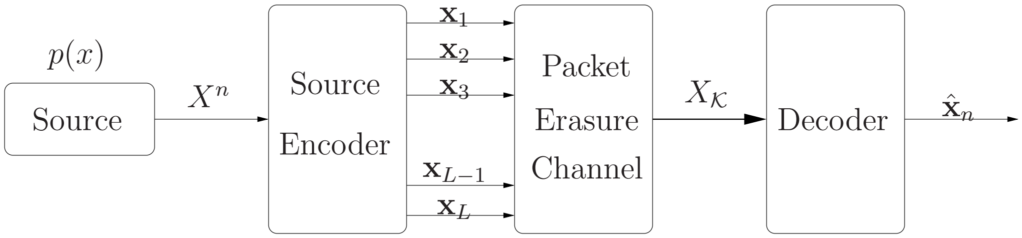

Figure 1: The MD Problem

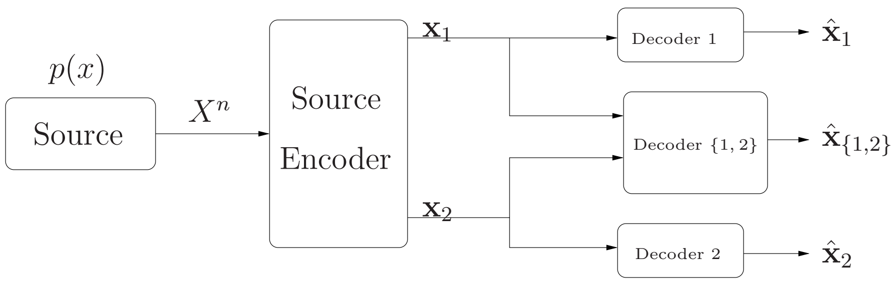

Figure 2: The 2-User Multiple Description Problem

We will reexamine the general rate-distortion regions that were discussed last time. Today, we will more closely examine the proofs of the theorem's of El Gamal and Cover [1], Zhang and Berger [3] as well as the region of Venkataramani, Kramer, and Goyal [4]. This time we will focus on the perspective of decoding with side information. Additionally, we will consider the more practically motivated question of finding "good" achievable methods of encoding to achieve an arbitrary distortion specification. We will first consider the common methods for proving the achievability of a given rate-distortion vector. Then we will consider a new constructive method that uses channel erasure codes as well as distributed source coding to encode for links with a symmetric rate.

2-Description MD Regions

The most basic problem in multiple description coding is the 2-description with three receiver problem. As discussed last time the exact region of achievable rates is known in the case of a Gaussian source. Indeed, this is the result of Ozarow [2]. The exact problem setup can be seen in Figure 2. We now recall one of the earliest characterizations of the 2-description coding problems.

Theorem 1. (The EGC 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:*

$$R_i \geq I(X; \hat{X}_i) \qquad \text{for } i \in \{1,2\} \tag{1}$$

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

$$D_{\mathcal{K}} \geq \mathbb{E}\left\{ d\left(X, \hat{X}_{\mathcal{K}}\right)\right\} \qquad \text{for } \mathcal{K} \in \{1, 2, \{1,2\}\} \tag{3}$$

The proof of this theorem follows for what is now a quite standard argument using random codes and typical sets. We refer the reader to [5] for a complete discussion on typical sets and their properties. We now provide a brief sketch of the proof of Theorem 1.

Sketch of Proof. We sketch the proof with the following steps:

0) **Given:**
   a) $\mathcal{X}_1$ and $\mathcal{X}_2$ such that $\mathcal{X}_1$ covers $\mathcal{X}^n$ with distortion D_1 and $\mathcal{X}_2$ covers $\mathcal{X}^n$ with distortion D_2.
   b) A collection of code books that $\mathcal{X}_0$ conditionally cover $\mathcal{X}^n$ with distortion d_0.
1) **Random Code Generation**
   a) Draw $2^{nR'_1}$ vectors uniformly from $T_\epsilon(\mathcal{X}_1)$.
   b) Draw $2^{nR'_2}$ vectors uniformly from $T_\epsilon(\mathcal{X}_2)$.
   c) For each jointly typical $(\mathbf{x}_1(i), \mathbf{x}_2(j))$ construct at codebook by drawing $2^{n\Delta}$ vectors from $T_\epsilon(\mathbf{x}_0|\mathbf{x}_1(i), \mathbf{x}_2(j))$.
2) **Encoding:** Given an $\mathbf{x} \in \mathcal{X}^n$ find an (i, j, k) such that $(\mathbf{x}, \mathbf{x}_0(k), \mathbf{x}_1(i), \mathbf{x}_2(j))$ are in the set of all jointly typical sequences if possible. Otherwise, set $(i, j, k) = (0, 0, 0)$. Split, k in to two so that $k = (k_1, k_2)$. Description 1 is then (i, k_1), description 2 is (j, k_2).
3) **Decoding:**
   a) Decoder 1: Receives (i, k_1) and announces $\mathbf{x}_1(i)$
   b) Decoder 2: Receives (j, k_2) and announces $\mathbf{x}_2(j)$
   c) Decoder $\{1, 2\}$: Receives (i, j, k) and announces $\mathbf{x}_0(i, j, k)$
3) **Distortion:** Can show,

$$\mathbb{E}\{\mathbf{d}\} = (1 - P_{\mathrm{e}})(\mathbf{D} + \epsilon) + P_{\mathrm{e}} d_{\max}$$

   where under the conditions of the theorem $P_{\mathrm{e}} \to 0$.

□

This region is the optimal region for a Gaussian source with square error distortion and was conjectured to be the optimal region in general. However, note that in the decoding stage of the above proof the descriptions k_1 and k_2 are thrown away when single description is received. While, this does not matter in the case of a Gaussian source with square error distortion, a slight modification to the proof above led Zhang and Berger to show that the EGC region is not in general optimal by examining the case of a binary source with a Hamming distortion. The region of Zhang and Berger (ZB) is described in the following.

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

$$\begin{aligned} R_i &\geq I(X; \hat{X}_i, \hat{X}_0) \qquad \textit{for } i \in \{1, 2\} & (4)\\ R_1 + R_2 &\geq 2I(X; \hat{X}_0) + I(\hat{X}_1; \hat{X}_2|\hat{X}_0) + I(X; \hat{X}_1, \hat{X}_2|\hat{X}_0) + I(\hat{X}_1; \hat{X}_2) & (5)\\ d_{\mathcal{K}} &\geq \mathbb{E}\left\{d\left(X, \hat{X}_{\mathcal{K}}\right)\right\} \qquad \textit{for } \mathcal{K} \in \{1, 2, \{1, 2\}\} & (6) \end{aligned}$$

where $X_1 = g_1(\hat{X}_0, \hat{X}_1)$*,* $X_2 = g_1(\hat{X}_0, \hat{X}_2)$*, and* $X_{\{1,2\}} = g_{\{1,2\}}(\hat{X}_0, \hat{X}_1, \hat{X}_2)$ *for some functions* $g_1, g_2, g_{\{1,2\}}$.

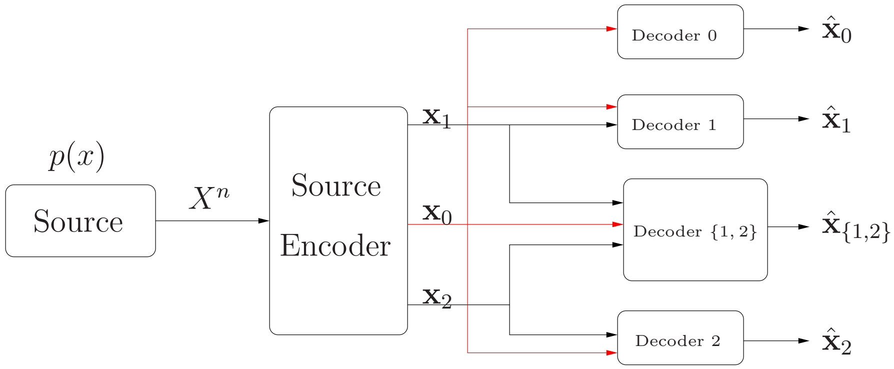

Figure 3: The Problem Zhang and Berger

The proof of this region follows much like that of the proof of Theorem 1. However, while the technical arguments rely on many of the same properties of typical sets the proof of Theorem 2 has two major conceptual differences. The first is that which addresses the possible penalty by neglecting the partial descriptions k_1 and k_2. The second is a common coarse description that is added to the description for each channel. In fact, Zhang and Berger arrive at Theorem 2 by first considering the three description and 4 receiver problem depicted in Figure 3. Note that this figure only differs from that of Figure 2 by the addition of the common refinement X_0 and the extra decoder corresponding to X_0. It is clear that any vector of achievable rate-distortion vales for this problem can be simply related to the 2 description problem. We now sketch the proof of Theorem 2 in order to outline the differences.

Sketch of Proof. We sketch the proof with the following steps:

0) **Given:** Random variables $\hat{X}_0$,$\hat{X}_1$ and $\hat{X}_2$ that satisfy the conditions of Theorem 2.

1) **Random Code Generation**
   a) Draw $2^{nR'_0}$ vectors uniformly from $T_\epsilon(\mathcal{X}_0)$.
   b) Draw $2^{nR'_1}$ vectors uniformly from $T_\epsilon(\mathcal{X}_1|\mathbf{x}_0(i))$.
   c) Draw $2^{nR'_2}$ vectors uniformly from $T_\epsilon(\mathcal{X}_2|\mathbf{x}_0(i))$.

2) **Encoding:** Given an $\mathbf{x} \in \mathcal{X}^n$ find an (i,j,k) such that $(\mathbf{x}, \mathbf{x}_0(k), \mathbf{x}_1(i), \mathbf{x}_2(j))$ are in the set of all jointly typical sequences if possible. Otherwise, set $(i,j,k) = (0,0,0)$. Thus, the first description is (i,k) and the second is (j,k).

3) **Decoding:**
   a) Decoder 1: Receives (i,k) and announces $g_1(\mathbf{x}_0(k), \mathbf{x}_1(i))$.
   b) Decoder 2: Receives (j,k) and announces $g_2(\mathbf{x}_0(k), \mathbf{x}_2(j))$.
   c) Decoder $\{1,2\}$: Receives (i,j,k) and announces $g_{\{1,2\}}(\mathbf{x}_0(k), \mathbf{x}_1(i), \mathbf{x}_2(j))$.

3) **Distortion:** Can show,

$$\mathbb{E}\{\mathbf{d}\} = (1-P_\mathrm{e})(\mathbf{D}+\epsilon) + P_\mathrm{e} d_{\max}$$

   where under the conditions of the theorem $P_\mathrm{e} \to 0$.

□

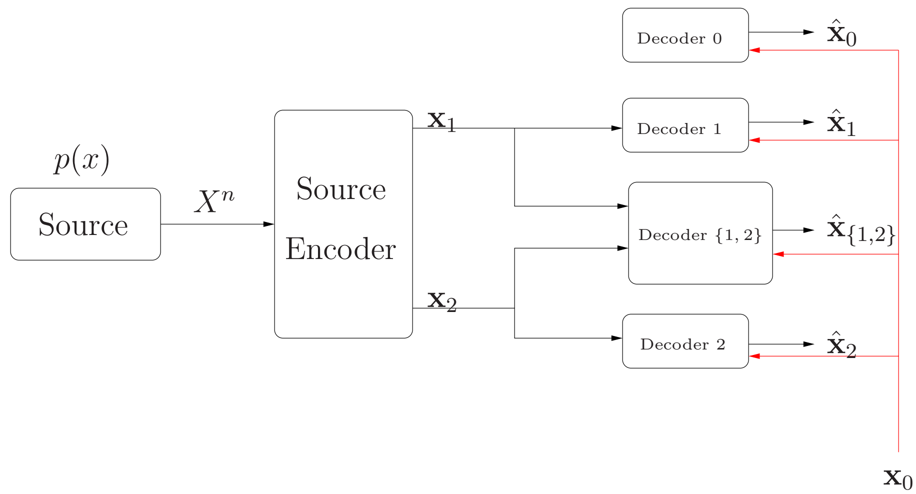

Figure 4: An alternate view of the problem Zhang and Berger

Using this framework Venkataramani, Kramer and Goyal [4] extended this result to the general problem with L-channels.

Theorem 3. *Let $X_1, X_2, \ldots, X_{\mathcal{K}}, \ldots, X_{2^{\mathcal{L}}}$ be a collection of 2^L random variables indexed by the subsets of the power set $2^{\mathcal{L}}$ of the description indices $\mathcal{L} = \{1, 2, \ldots, L\}$ that are jointly distributed with the source random variable X. Let $X_\emptyset$ take on values in an arbitrary finite alphabet $\mathcal{X}_\emptyset$ and each $X_{\mathcal{K}}$ takes on values in a reconstruction alphabet $\mathcal{X}_{\mathcal{K}}$ for $\mathcal{K} \neq \emptyset$. Then, the rate-distortion region contains the rates and distortions satisfying*

$$\begin{aligned} D_{\mathcal{K}} &\geq \mathbb{E}\{d_{\mathcal{K}}(X, X_{\mathcal{K}})\} && (7)\\ R_{\mathcal{K}} &\geq (|\mathcal{K}| - 1)\, I(X; X_\emptyset) - I(X, X_{(\mathcal{K})}) \\ &\quad + \sum_{\mathcal{M} \subseteq \mathcal{K}} H(X_{\mathcal{M}} | X_{(2^{\mathcal{M}} - \mathcal{M})}) && (8) \end{aligned}$$

for every $\mathcal{K} \in 2^{\mathcal{K}} - \emptyset$ and where $X_{(\mathcal{K})} \triangleq \{X_{\mathcal{M}} : \mathcal{M} \in \mathcal{K}\}$.

It is important to note that in this framework the random variable $X_\emptyset$ plays the same role as X_0 in Theorem 2. That is, $X_\emptyset$ can be interpreted as side information that is available at each decoder. That is, by neglecting the decoder corresponding to X_0 in Figure 3 and viewing X_0 as side information at the decoder we may be tempted to think of the Theorem 2 (and for that matter Theorem 3) as that depicted in Figure 4. Note however, since this side information is not available at the decoder we must transmit it over each channel and thus suffer the leading term in Theorem 3. This point of view leads us to a quite practical achievable scheme studied by Pradhan et. al. [6, 7]. We begin by first reviewing a fundamental result in the distributed encoding of correlated sources.

The Slepian–Wolf Problem

We now briefly recall the problem studied by Slepian and Wolf that we require in the sequel. We refer the reader to [5, pg. 407–416] for a more complete discussion. Consider a pair of jointly distributed random variables (X, Y) with joint distribution $p(x, y)$ and consider jointly encoding both X and Y for transmission to a common receiver. It should be clear that a rate at least $H(X) + H(Y)$ is sufficient by separately encoding X and Y. However, in a fundamental paper by Slepian and Wolf it is shown that it is sufficient to use a total rate of $H(X, Y)$ even if the encoding is done separately. This can be shown by *randomly binning* the codewords then decoding based on knowledge of the joint statistics of X and Y. That is, by uniformly assigning every codeword of $\mathcal{X}^n$ to a random bin and uniformly assigning every codeword of $\mathcal{Y}^n$ to a random bin and having each encoder transmit the bin indices. The two indices are then jointly decoded based on joint typicality.

It should be clear that in this framework we could reprove Theorem 2 using Figure 4. It is this same approach that the authors of [6, 7] take. We now briefly review these results.

The region of Pradhan et. al.

We now turn our focus to the more practical question of the symmetric multiple description problem, whereby all descriptions are encoded at the same rate. Such a scenario occurs in current data transfer protocols for example. We require equal protection for all the data so that the distortion is a function of the number of descriptions lost only. That is, if n_{lost} many packets are lost and there are L total descriptions then receiving either description indexed by $\{1, 2, \ldots, n - n_{\text{lost}}\}$ or $\{n_{\text{lost}} + 1, n_{\text{lost}} + 2, \ldots, n\}$ incurs the same distortion. Thus, the achievable rate distortion region is described by the $L + 1$ vector $(R, D_1, D_2, \ldots, D_L)$ where D_k is the distortion incurred if k descriptions are received.

We note that in the case of two descriptions our intuition is well grounded. For example, as we noted earlier in the 2 description problem we know that the two descriptions must be nearly independent if we wish to approach the rate distortion function for the combined rate, i.e. if $d_{\{1,2\}} \approx D(R_1 + R_2)$ then X_1 and X_2 must be nearly independent. Further, it is rather straight forward to encode for this case since we know exactly which two descriptions are received. However, in the L channel case there is inherent uncertainty at the encoder about which descriptions are received.

Suppose for now that we know that at least k descriptions are received. Then one may as a first approach employ a rate kR rate-distortion code that achieves the optimal distortion and encode it for the multiple channels using a (L, k) erasure code. Thus, if any k descriptions are received we may recover the data and obtain the optimal distortion. However, if more than k descriptions are received the decoder does not gain any new information. Alternatively, we could generate L independent codebooks in order to gain new information for each received description. It is now not clear, however, if we may still achieve the optimal distortion for a rate kR code. It is a key result of [6,7] that in the case of a Gaussian source one may use independent codebooks in order to continually decrease the distortion as more descriptions are received while simultaneously achieving the optimal distortion for a rate kR code. To be more precise, we let $D^{(k)}(R) = D(kR)$ be the distortion achieved by receiving k or more descriptions using a (L, k) erasure code. Then if the decoder receives r extra descriptions the result of [6, 7] states that a distortion of

$$D_r^{(k)}(R) = \frac{k}{\frac{k+r}{D^{(k)}(R)} - r}$$

is achievable. Note that if $r = 0$, i.e. exactly k descriptions are received, then this scheme achieves the optimal distortion. Further, since $D^{(k)}(R) < 1$, the distortion decreases as we receive extra information (which was not the case for the erasure code approach). The authors of [6, 7] provide the following more general theorem for which the preceding results are a consequence.

Theorem 4. *Let $X_1, X_2, \ldots, X_n$ be a collection of random variables that are jointly and symmetrically distributed with an arbitrary source random variable X. Then, if at least k descriptions are received and*

$$\begin{aligned} D_{\mathcal{K}} &\geq \mathbb{E}\left\{d(X, g_{\mathcal{K}}(X_{\mathcal{K}}))\right\} \qquad \textit{for } J \in \mathcal{L}_k & (9) \\ R &\geq \frac{1}{k}H(X_1, X_2, \ldots, X_k) - \frac{1}{n}H(X_1, X_2, \ldots, X_n|X) & (10) \end{aligned}$$

for some decoding functions g_J, where $\mathcal{L}_k = \{\mathcal{M} \subset \{1, 2, \ldots, L\} : |\mathcal{M}| \geq k\}$, the rate-distortion vector $(R, D_k, D_{k+1}, \ldots, D_L)$ is achievable.

Sketch of Proof. We sketch the proof with the following steps:

0) **Given:** A joint distribution $p(x_1, x_2, \ldots, x_L|x)$ such that the conditions of the theorem are satisfied.

1) **Random Code Generation:** Construct a random codebook $\mathcal{X}_i$ independently for $i \in \{1, 2, \ldots, L\}$ by selecting $2^{nR'}$ codewords uniformly from $T_\epsilon(X_i)$.

2) **Random Binning:** To each codebook associate 2^{nR} bins each containing approximately $2^{n(R'-R)}$ codewords which are assigned by drawing uniformly from $\mathcal{X}_i$ with replacement.

3) **Encoding:** Given an $\mathbf{x} \in \mathcal{X}^n$ find indices $i_1, i_2, \ldots, i_L$ such that the corresponding codewords and source, $(\mathbf{x}, \mathbf{x}_1(i_1), \mathbf{x}_2(i_2), \ldots, \mathbf{x}_L(i_L))$, are jointly typical. Let $i_l = 0$ if no such indices exist. Transmit the smallest bin index that contains $\mathbf{x}_1(i_l)$ over channel l if such an index exists. Otherwise, transmit 0.

4) **Decoding:** For a set of received bin indices indexed by $\mathcal{K}$ the decoder finds for $k_i \in \mathcal{K}$ the $(\mathbf{x}, \mathbf{x}_{k_1}, \mathbf{x}_{k_2}, \ldots, \mathbf{x}_{k_L})$ that are jointly typical and each are contained in the bin corresponding to the received index. If more than one such set exists declare an error. Otherwise, decode to $g_{\mathcal{K}}(\mathbf{x}_{\mathcal{K}})$.

5) **Distortion:** Can show,

$$\mathbb{E}\left\{\mathbf{d}\right\} = (1 - P_{\mathrm{e}})(\mathbf{D} + \epsilon) + P_{\mathrm{e}} d_{\max}$$

where under the conditions of the theorem $P_{\mathrm{e}} \to 0$.

□

Conclusion

We have presented the results of [3, 4] in terms of distributed source coding and have presented the achievable scheme of [6]. While it is still unclear whether the many user regions of [4, 6] are the optimal characterization of the achievable rate-distortion pairs, they provide valuable insights in to practical encoding schemes.

References

[1] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 851–857, November 1982.

[2] L. Ozarow, "On a source–coding problem with two channels and three receivers," *BSTJ*, pp. 1909–1921, May 1980.

[3] Z. Zhang and T. Berger, "New results in multiple descriptions," *IEEE Transactions on Information Theory*, vol. 33, no. 4, pp. 502–521, July 1987.

[4] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Transactions on Information Theory*, vol. 49, no. 9, pp. 2106–2114, September 2003.

[5] T. M. Cover and J. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons Inc., 1997.

[6] S. S. Pradhan, R. Puri, and K. Ramchandran, "n-channel symmetric multiple descriptions – part I: (n, k) source-channel erasure codes," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 47–61, January 2004.

[7] ——, "n-channel symmetric multiple descriptions – part II: An achievable rate-distortion region," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1377–1392, April 2006.