Multiple Descriptions with Feed-forward: A Single-letter Achievable Rate Region.

Ramji Venkataramanan and S. Sandeep Pradhan
Department of EECS, University of Michigan, Ann Arbor, MI 48105
rvenkata@umich.edu, pradhanv@eecs.umich.edu

Abstract— We study the two-channel multiple descriptions problem for an i.i.d source, with feed-forward to one or both side-decoders. We derive a single-letter achievable rate-region that strictly includes the best known rate-region for multiple descriptions without feed-forward. In point-to-point source coding, feed-forward does not improve the rate-distortion function of a discrete memoryless source. In contrast, we provide an example to show that our region can be strictly smaller than the optimal region without feed-forward. The proof of the result uses a block-Markov superposition source coding strategy.

I. INTRODUCTION

Consider a communication network in which we wish to compress a streaming source of data into packets at one node and transmit them to another node. Assume there is a chance that a packet might be lost and never reaches its destination. So we compress each block of data simultaneously into two different packets and send them through different routes. We get a good reconstruction on reception of either packet, but we would like a better reconstruction if both packets are received; in other words, the packets need to refine one another. How should we compress the source into two different descriptions? This, in essence, is the multiple descriptions problem, first posed by Gersho, Ozarow, Witsenhausen and others.

The multiple descriptions set-up is shown in Figure 1. In the standard problem, both S_1 and S_2 are open. $\mathbf{X} = \{X_n\}_{n=1}^{\infty}$ is a source with known distribution. The encoder encodes each block of source samples in two different ways: decoder 1 receives R_1 bits/sample and produces reconstruction $\hat{X}_1$. Similarly, decoder 2 receives R_2 bits/sample and produces $\hat{X}_2$. Decoder 0 receives the full $R_1 + R_2$ bits and produces reconstruction $\hat{X}_0$. Assume suitable distortion measures have been defined for all decoders; let D_1, D_2, D_0 denote the average distortions with which decoders $1, 2$ and 0 are able to reconstruct the source. The problem is to determine the set of all quintuples $(R_1, R_2, D_1, D_2, D_0)$ that are achievable in the usual Shannon sense. This problem has been studied in several notable papers, including [1]–[10]. In this paper, we study multiple descriptions source coding with feed-forward.

To explain the notion of feed-forward in simple terms, let us first consider the point-to-point case. In the usual fixed-rate lossy source coding problem, there is a source X that has to be reconstructed with some distortion D. The encoder takes
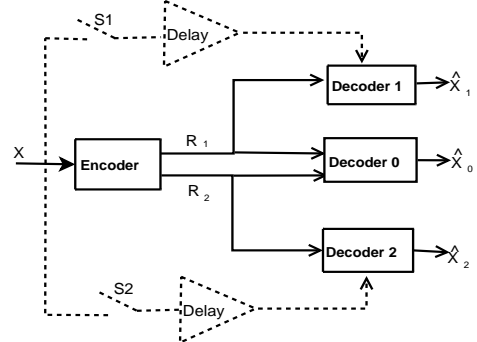
Fig. 1. The multiple descriptions problem

a block of, say, N source samples and maps it to an index in a codebook. The decoder uses this index to reconstruct the N source samples. In source coding with feed-forward, the encoder works in a similar fashion and sends an index to the decoder. The decoder generates the reconstructions sequentially: in order to reconstruct each source sample, the decoder has access to the index *and* some past source samples. Let $X_n, \hat{X}_n$ denote the source and reconstruction samples at time n, respectively. If the source samples are available with a delay k after the index is sent, the decoder has knowledge of the index plus the source samples until time $n - k$ to produce $\hat{X}_n$. We call this set-up feed-forward with delay k.

The notion of feed-forward is applicable to multi-terminal problems as well. Figure 1 shows a multiple descriptions system with feed-forward. Assume that switch S_1 is closed and the source samples are sequentially available with a delay k after the indices are sent. To generate $\hat{X}_{1n}$, decoder 1 has knowledge of the index in a codebook (of rate R_1) plus the source samples until time $n - k$. A similar situation occurs for decoder 2 when S_2 is closed. In this paper, we study the rate vs. distortion trade-offs- the achievable quintuples $(R_1, R_2, D_1, D_2, D_0)$ - when one or both of S_1 and S_2 are closed.

Source coding with feed-forward is relevant in many different settings. The problem was motivated and studied from a communications perspective in [11]–[13], as a variant of source coding with side information. For example, consider the source to be a field that needs to compressed and communicated from one node to another in a network. This field (e.g. a seismic or acoustic field) could propagate through the

medium at a slow rate and become available at one or more decoding nodes as side-information with some delay.

Source coding with feed-forward is also related closely to prediction. In fact, it was first considered in the context of competitive prediction in [14]. The problem of choosing the best predictor of a random process from an exponentially large class of predictors is equivalent to the source coding problem with feed-forward. The following problem is another example that motivates our study of multiple descriptions with feed-forward. There are four agents Alice, Bob, Carol and Dave. Alice has access to an equiprobable binary source; Bob, Carol and Dave are interested in reconstructing the source sequence. Dave needs perfect reconstruction, while Bob and Carol each want to reconstruct with the fraction of their errors being at most d. Accordingly, Bob and Carol agree to buy some information from Alice separately, and Dave agrees to buy the information available to both Bob and Carol. Further assume that after Bob, Carol and Dave reconstruct each source sample, Alice reveals to Carol (but not to Bob and Dave) whether she made an error or not. The minimum rates of information that Alice would have to supply to Bob and Carol under this scenario is the multiple description rate-distortion region with feed-forward to Carol only. Using our results, we show in Section III that the partial feed-forward in the above example (i.e., to Carol only) allows information rates smaller than the best possible rates without feed-forward.

In [11], a simple multiple-description coding scheme was presented for i.i.d. Gaussian sources with feed-forward to *all* decoders (0,1 and 2) with delay $k = 1$. The coding scheme was shown to achieve the optimal rate-distortion region for the i.i.d Gaussian source with feed-forward. In this paper, we present an achievable rate-region for any discrete memoryless source with arbitrary feed-forward delay k, when one or both of S_1 and S_2 are closed. In point-to-point source coding, feed-forward does not improve the rate-distortion function of a discrete memoryless source. Our results show that for multiple descriptions, the rate-distortion region can be strictly smaller with feed-forward.

II. PROBLEM STATEMENT AND MAIN RESULTS

Consider a discrete memoryless source X with finite alphabet $\mathcal{X}$. We assume that the source samples X_n, $n = 1, 2, \ldots$ are independent and identically distributed (i.i.d) according to a probability mass function $P_X(x)$. Let $\widehat{\mathcal{X}}_0, \widehat{\mathcal{X}}_1, \widehat{\mathcal{X}}_2$ denote the finite reconstruction spaces of decoder 0,1 and 2, respectively. Each reconstruction has an associated single letter distortion measure:

$$d_m : \mathcal{X} \times \widehat{\mathcal{X}}_m \to \mathbb{R}, \qquad m = 0, 1, 2.$$

The distortion on $N-$length sequences is the average of the per-letter distortion. For all $x^N \in \mathcal{X}^N, \hat{x}^N \in \widehat{\mathcal{X}}^N$,

$$d_m(x^N, \hat{x}_m^N) \triangleq \frac{1}{N} \sum_{n=1}^{N} d_m(x_n, \hat{x}_{mn}), \quad m = 0, 1, 2.$$

A. Feed-forward to only one decoder

Without loss of generality assume S_1 is open and S_2 is closed in Figure 1.

Definition 1: An $(N, 2^{NR_1}, 2^{NR_2})$ multiple description code with delay k feed-forward of block length N and rate R consists of:

1) Encoder mappings

$$e_m : \mathcal{X}^N \to \{1, \ldots, 2^{NR_m}\}, \qquad m = 1, 2 \quad (1)$$

2) Mappings for decoders 0 and 1

$$g_0 : \{1, \ldots, 2^{NR_1}\} \times \{1, \ldots, 2^{NR_2}\} \to \widehat{\mathcal{X}}_0^N$$
$$g_1 : \{1, \ldots, 2^{NR_1}\} \to \widehat{\mathcal{X}}_1^N \qquad (2)$$

3) A sequence of mappings for decoder 2 [1]

$$g_{2n} : \{1, \ldots, 2^{NR_2}\} \times \mathcal{X}^{n-k} \to \widehat{\mathcal{X}}_2, \quad n = 1, \ldots, N. \quad (3)$$

The encoder maps each N-length source sequence to a pair of indices in in $\{1, \ldots, 2^{NR_1}\} \times \{1, \ldots, 2^{NR_2}\}$. The decoders receive their respective indices. In addition, to reconstruct the nth sample, decoder 2 has access to the source samples until time $(n-k)$. Achievable rates are defined in the usual Shannon sense.

Definition 2: (R_1, R_2) is an achievable rate pair for distortion (D_0, D_1, D_2) if there exists a sequence, indexed by N, of $(N, 2^{NR_1}, 2^{NR_2})$ multiple description codes with feed-forward delay k, such that for sufficiently large N,

$$Ed_m(X^N, \hat{X}_m^N) \le D_m, \quad m = 0, 1, 2. \quad (4)$$

The rate distortion region $R(D_0, D_1, D_2)$ is the closure of the set of achievable rate pairs for distortion (D_0, D_1, D_2).

Our main result is the following theorem.

Theorem 1: A quintuple $(R_1, R_2, D_0, D_1, D_2)$ is achievable - with delay k feed-forward to decoder 2 only- if there exist random variables $U, \hat{X}_1, \hat{X}_2, \hat{X}_0$ jointly distributed with the source X such that

$R_1 > I(X; \hat{X}_1 U)$

$R_2 > I(X; \hat{X}_2|U) + \max\{0, R_1 - I(X\hat{X}_2; \hat{X}_1|U)\}$

$R_1 + R_2 > I(X; \hat{X}_1 U) + I(X; \hat{X}_2|U) + I(X; \hat{X}_0|\hat{X}_1 \hat{X}_2 U)$
$\qquad + I(\hat{X}_1; \hat{X}_2|XU) + \max\{0, R_1 - I(X\hat{X}_2; \hat{X}_1|U)\}$

$Ed_m(X; \hat{X}_m) \le D_m, \quad m = 0, 1, 2$

The proof of the theorem is given in Section IV. Notice that the rate-region specified by the theorem does not depend on the feed-forward delay k, i.e., the region is achievable for any finite delay k. We can compare the rate region with the achievable rates for multiple descriptions without feed-forward. The multiple descriptions rate-distortion region (without feed-forward) is known only for certain special cases (see [1], [3], [4], [7]). The best known achievable region for a general i.i.d source is due to Zhang and Berger [5] which we reproduce below in a slightly modified, but equivalent, form.

---

[1] It is understood that for $n \le k$, $\mathcal{X}^{n-k}$ is the empty set.

Theorem 2 ([5]): A quintuple $(R_1, R_2, D_0, D_1, D_2)$ is achievable (without feed-forward) if there exist random variables $U, \hat{X}_1, \hat{X}_2, \hat{X}_0$ jointly distributed with the source X such that

$R_1 > I(X; \hat{X}_1 U)$

$R_2 > I(X; \hat{X}_2 U)$

$R_1 + R_2 > I(X; \hat{X}_1 U) + I(X; \hat{X}_2 U) + I(X; \hat{X}_0 | \hat{X}_1 \hat{X}_2 U)$
$$+ I(\hat{X}_1; \hat{X}_2 | XU)$$

$Ed_m(X; \hat{X}_m) \leq D_m, \quad m = 0, 1, 2$

To see that the rate region of Theorem 1 represents an improvement the no-feed-forward case, consider any set of random variables $U, \hat{X}_1, \hat{X}_2, \hat{X}_0$ jointly distributed with X. Set $R_1 = I(X; \hat{X}_1 U) + \epsilon$ for some small $\epsilon > 0$. We can have one of two situations:

1) $I(X; \hat{X}_1 U) + \epsilon \leq I(X \hat{X}_2; \hat{X}_1 | U)$.
   In this case, Theorem 1 tells us that

   $$R_2 = I(X \hat{X}_1; \hat{X}_2 | U) + I(X; \hat{X}_0 | \hat{X}_1 \hat{X}_2 U) + \epsilon$$

   is achievable. This represents a savings of $I(U; X)$ bits/sample over the minimum R_2 without feed-forward (specified by Theorem 2).
2) $I(X; \hat{X}_1 U) + \epsilon > I(X \hat{X}_2; \hat{X}_1 | U)$.
   From Theorem 1,

   $$R_2 = I(X \hat{X}_1; \hat{X}_2 | U) + I(X; \hat{X}_0 | \hat{X}_1 \hat{X}_2 U)$$
   $$+ [I(X; U) - I(\hat{X}_2; \hat{X}_1 | XU)] + \epsilon$$

   is achievable, a savings of $I(\hat{X}_2; \hat{X}_1 | XU)$ bits/sample over the no-feed-forward case.

Of course, the potential savings in rate may be greater since we have only presented an achievable rate region. The coding strategy in [5] first quantizes the block of source samples X^N to U^N, which is sent to all the decoders. Decoders $0, 1, 2$ then produce their reconstructions conditioned on U^N. The coding strategy in Theorem 1 uses the feed-forward to decoders 1 and 2 to convey U^N 'cheaply' to all decoders.

B. Feed-forward to both decoders 1 and 2

Switches S_1 and S_2 in Figure 1 are *both* closed. An $(N, 2^{NR_1}, 2^{NR_2})$ multiple description code with delay k feed-forward is defined in the same way as the previous subsection, except that the decoder mappings are modified. The mapping defining decoder 0 is given by

$$g_0 : \{1, \ldots, 2^{NR_1}\} \times \{1, \ldots, 2^{NR_2}\} \to \widehat{\mathcal{X}}_0^N \qquad (5)$$

We have a sequence of mappings for decoders 1 and 2

$$g_{mn} : \{1, \ldots, 2^{NR_2}\} \times \mathcal{X}^{n-k} \to \widehat{\mathcal{X}}_m, \; n = 1, \ldots, N, \quad m = 1, 2. \qquad (6)$$

In addition to the index, both decoders 1 and 2 have access to the source samples until time $(n - k)$.

Achievable rates are defined in the usual Shannon sense as before. Clearly, the region of Theorem 1 is achievable. The rate region obtained by switching the roles of R_1 and R_2 in Theorem 1 is also achievable. Thus, taking the convex hull

of the union of these two rate-regions yields a possibly larger rate-region.

We now present an example to show that the rate-region with feed-forward to only one decoder can be strictly greater than the optimal no feed-forward rate-distortion region.

III. EXAMPLE

Consider an i.i.d binary source X with pmf $P_X(0) = P_X(1) = 1/2$. The reconstruction spaces are all binary and the distortion measures are Hamming, i.e., $d(x, \hat{x}_m) = \delta_{x = \hat{x}_m}, \quad m = 0, 1, 2$. Suppose decoders 1 and 2 want to reconstruct X with distortion d, while decoder 0, needs error-free reconstruction. We want to characterize the function

$$r_{sum}(d) \triangleq \inf\{R_1 + R_2 : (R_1, R_2, 0, d, d) \in \mathbf{R}\}. \quad (7)$$

A lower bound to $r_{sum}(d)$ without feed-forward was obtained in [5, Theorem 3, Section VIII] [2]:

$$r_{sum}(d)_{no-ff} \geq 2 - h\left(\frac{4d + 1 - \sqrt{12d^2 - 4d + 1}}{2}\right). \quad (8)$$

Assume only decoder 2 gets feed-forward with delay k. In our problem, this means that k time units after decoder 2 reconstructs a sample, it is told whether that reconstruction was correct or not. Let U be a binary-valued random variable and fix the conditional distribution $P_{U, \hat{X}_1, \hat{X}_2, \hat{X}_0 | X}$ as

$$P_{U|X} \cdot P_{\hat{X}_1, \hat{X}_2 | XU} \cdot P_{\hat{X}_0 | XU \hat{X}_1 \hat{X}_2}$$

where

- $P_{U|X}$: Fix a parameter D_0 and define

  $$P_{U|X}(0|0) = P_{U|X}(1|1) = 1 - D_0$$
  $$P_{U|X}(0|1) = P_{U|X}(1|0) = D_0. \qquad (9)$$

- $P_{\hat{X}_1, \hat{X}_2 | XU}$: When U and X agree, both deocders 1 and 2 produce correct reconstructions.

  $$P_{\hat{X}_1, \hat{X}_2 | XU}(00|00) = P_{\hat{X}_1, \hat{X}_2 | XU}(00|00) = 1 \quad (10)$$

  When U and X disagree, at least one of them produces the correct reconstruction, according to:

  $$P_{\hat{X}_1, \hat{X}_2 | XU}(01|01) = P_{\hat{X}_1, \hat{X}_2 | XU}(10|01) = d/D_0$$
  $$P_{\hat{X}_1, \hat{X}_2 | XU}(00|01) = 1 - 2d/D_0$$
  $$P_{\hat{X}_1, \hat{X}_2 | XU}(01|10) = P_{\hat{X}_1, \hat{X}_2 | XU}(10|10) = d/D_0 \quad (11)$$
  $$P_{\hat{X}_1, \hat{X}_2 | XU}(11|10) = 1 - 2d/D_0$$

- $P_{\hat{X}_0 | XU \hat{X}_1 \hat{X}_2}$: $\hat{X}_0$ is a function of $(U, \hat{X}_1, \hat{X}_2)$.

  $$\hat{X}_0 = \begin{cases} 0 & \text{if } (\hat{X}_1 = \hat{X}_2 = 0) \\ 1 & \text{if } (\hat{X}_1 = \hat{X}_2 = 1) \\ 1 - U & \text{if } (\hat{X}_1 \neq \hat{X}_2) \end{cases} \quad (12)$$

It is easy to check that this joint distribution achieves the distortion triple $(D_1 = d, D_2 = d, D_0 = 0)$. Using this joint distribution in Theorem 1, we can obtain an achievable

---

[2]There is a typo in the statement of Theorem 3 in [5]. The correct version (given here) can be obtained from the proof of that theorem.
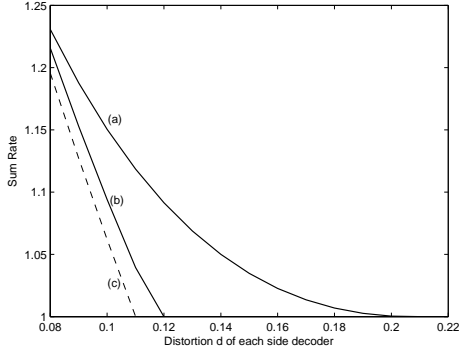
Fig. 2. (a) Zhang-Berger lower bound on $r_{sum}(d)$ without FF, (b) Achievable sum-rate with feed-forward to one decoder , (c) Rate-distortion lower bound on $r_{sum}(d)$ with FF.

rate-region when only decoder 2 receives feed-forward. The relevant information quantities are calculated below, with $h(.)$ used to denote the binary entropy function.

$$I(X;U) = H(U) - H(U|X) = 1 - h(D_0).$$
$$I(\hat{X}_2; \hat{X}_1|XU) = H(\hat{X}_2|XU) - H(\hat{X}_1|\hat{X}_2XU)$$
$$= D_0 h(\frac{d}{D_0}) - D_0\left(1 - \frac{d}{D_0}\right)h(\frac{d}{D_0 - d}).$$
$$I(X;\hat{X}_2|U) = I(X;\hat{X}_1|U) = H(\hat{X}_1|U) - H(\hat{X}_1|UX)$$
$$= h(D_0 - d) - D_0 h(\frac{d}{D_0}).$$
$$I(X;\hat{X}_0|\hat{X}_1\hat{X}_2U) = 0.$$
$$(13)$$

(13) contains all the expressions required to compute the rate-region of Theorem 1. For each d, we can fix the value D_0 to yield the best rate-constraint. Hence we obtain an achievable upper bound to $r_{sum}(d)$ in (7) with feed-forward to only one decoder. This is plotted in graph (b) in Figure 2 for distortions $d \geq 0.08$. Graph (a) is the Zhang-Berger lower bound (8) to $r_{sum}(d)$ without feed-forward. We see that for all the distortions considered, we do better than the optimal no feed-forward rate with feed-forward to one decoder . Since decoders 1 and 2 produce reconstructions with distortion d, R_1 and R_2 have to each be greater than the Shannon rate-distortion function $R(d) = 1 - h(d)$. This is true both with and without feed-forward. Thus a simple lower bound to $r_{sum}(d)$ with feed-forward is

$$r_{sum}(d) > 2(1 - h(d)),$$

which is plotted in graph (c) of Figure 2.

Of particular interest is the case when the sum rate $R_1 + R_2 = 1$. This is the case of *no excess rate* to the central decoder [4]. Setting $D_0 = 0.25945$, Theorem 1 tells us that $(R_1 = 0.5, R_2 = 0.5)$ is achievable for $d = 0.12$ with feed-forward to one decoder. In comparison, it was shown in [15] that with rates of (0.5,0.5) and no feed-forward, the minimum achievable distortion at each side-decoder is $(\sqrt{2} - 1)/2 = 0.207$.
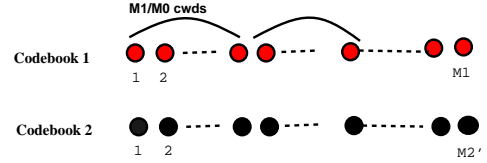

Fig. 3. Codebooks for decoders 1 and 2

IV. PROOF OF THEOREM 1

Assume delay k feed-forward, i.e. each source sample is available at the decoder $N + k$ time units after it is available to the encoder (N will be a measure of block-length in the block-Markov coding scheme). First fix the joint distribution

$$P_X(x) \cdot P_{U,\hat{X}_1,\hat{X}_2,\hat{X}_0|X}(u, \hat{x}_1, \hat{x}_2, \hat{x}_0|x).$$

For the sake of brevity, we will often drop the subscripts in the distributions and use only the arguments. Upper-case letters will be used for random variables and lower-case letters for their realizations. Vectors will be denoted in bold letters.

To prove the theorem, we shall use the properties of strongly ϵ-typical sequences [16]. Length-$N/2$ vectors $x^{N/2}, \hat{x}_1^{N/2}, \hat{x}_2^{N/2}$ are said to be jointly typical if their joint type (composition) is approximately $P_{X,\hat{X}_1,\hat{X}_2}$. The set of all jointly ϵ-typical tuples $(X^{N/2}, \hat{X}_1^{N/2}, \hat{X}_2^{N/2})$ is denoted $T_\epsilon(\mathbf{X}, \hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2)$. Similar definitions of typicality hold for other joint and conditional distributions.

We divide the source sequence in a large number of blocks, say B, with each block containing $\frac{N}{2}$ source symbols. To exploit the feed-forward, we shall use a block-Markov superposition strategy [17], [18] spanning adjacent blocks. The ideas of non-random binning and restricted encoding, introduced in [18], will be used in the proof. The block-Markov coding scheme is described in detail below.

Random Coding: Let $M_0 = 2^{\frac{N}{2}R_0}$, $M_1 = 2^{\frac{N}{2}R_1}$ and $M_2' = 2^{\frac{N}{2}R_2'}$. Choose $\mathbf{U}(1), \ldots, \mathbf{U}(M_0)$ independently according to a uniform distribution over the set $T_\epsilon(\mathbf{U})$ of all the ϵ-typical $N/2$-vectors $\mathbf{U}$. For each $\mathbf{U}(i)$, choose a codebook $\hat{\mathbf{X}}_1^i(1), \ldots, \hat{\mathbf{X}}_1^i(M_1)$ of $N/2$-vectors, independently according to a uniform distribution over the set $T_\epsilon(\hat{\mathbf{X}}_1|\mathbf{U}(i))$. Similarly choose $\hat{\mathbf{X}}_2^i(1), \ldots, \hat{\mathbf{X}}_2^i(M_2')$ from $T_\epsilon(\hat{\mathbf{X}}_2|\mathbf{U}(i))$.

Partition each $\hat{\mathbf{X}}_1^i$ codebook into M_0 disjoint cells, so that each cell has M_1/M_0 elements (we assume for simplicity that M_1/M_0 is an integer). One simple way to partition the codebook is shown in Figure 3.

Encoding: We consider encoding of a source sequence $\mathbf{x}$ spanning B blocks, each block containing $N/2$ source symbols. We denote the bth block by $\mathbf{x}_b$, $b = 1, \ldots B$. Thus

$$\mathbf{x} = [x_1, x_2, \ldots, x_{BN/2}] = [\mathbf{x}_1\mathbf{x}_2 \ldots \mathbf{x}_B].$$

Broadly speaking, there is a 'cloud center' $\mathbf{u}_b$ for each block b, conditioned on which reconstructions $\hat{\mathbf{x}}_{1b}$ and $\hat{\mathbf{x}}_{2b}$ are produced at decoders 1 and 2.

- *Step* 0: Set $\mathbf{u}_1 = \mathbf{U}(1)$.

Time instant	1	...	$N/2$	...	$2N/2$	...	$N(b+1)/2$	·	$BN/2$
Source	X_1	...	$X_{N/2}$	...	$X_{2N/2}$	...	$X_{N(b+1)/2}$	·	$X_{BN/2}$
Encoder					(w_{11}, w_{21})		(w_{1b}, w_{2b})		(w_{1B-1}, w_{2B-1})
FF at decoder	-	-	-	-	-	$X_1 \ldots$	...	$X_{bN/2} \ldots$	
Reconstruction					$\hat{\mathbf{x}}_{11}, \hat{\mathbf{x}}_{21}$		$\hat{\mathbf{x}}_{1b}, \hat{\mathbf{x}}_{2b}$		$\hat{\mathbf{x}}_{1B-1}, \hat{\mathbf{x}}_{2B-1}$

- *Step* b ($b = 1, \ldots, B - 1$): Assume $\mathbf{u}_b$ is known to be equal to $\mathbf{U}(i)$. Encode $\mathbf{x}_b$ as follows. Observe the next length-$N/2$ block $\mathbf{x}_{b+1}$ and find a $j \in \{1, \ldots, M_0\}$ such that $(\mathbf{x}_{b+1}, \mathbf{U}(j)) \in T_\epsilon(\mathbf{X}, \mathbf{U})$. Set

$$\mathbf{u}_{b+1} = \mathbf{U}(j).$$

  If no such j is found, set $\mathbf{u}_{b+1} = \mathbf{U}(1)$.
  So we now have $\mathbf{u}_b = \mathbf{U}(i), \mathbf{u}_{b+1} = \mathbf{U}(j)$. Pick $(w_{1b}, w'_{2b}) \in \{1, \ldots, M_1\} \times \{1, \ldots, M'_2\}$ such that $(\mathbf{x}_b, \mathbf{u}_b, \hat{\mathbf{X}}_1^i(w_{1b}), \hat{\mathbf{X}}_2^i(w'_{2b})) \in T_\epsilon(\mathbf{X}, \mathbf{U}, \hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2)$ *and* $\hat{\mathbf{X}}_1^i(w'_{2b})$ belongs to the jth cell of the $\hat{\mathbf{X}}_1^i$ codebook. If no such (w_{1b}, w'_{2b}) is found, set w_{1b} to a random index in the jth cell of the $\hat{\mathbf{X}}_1^i$ codebook, and similarly set w'_{2b} to a random index in the $\hat{\mathbf{X}}_2^i$ codebook.
  Note that we perform *restricted encoding*, i.e., we do not search for reconstructions in the entire $\hat{\mathbf{X}}_1^i$ codebook-instead, we restrict ourselves to one cell within the codebook. As explained in the sequel, restricted encoding enables decoder 2 to take advantage of the feed-forward. Decoders 1 and 2 will receive w_{1b} and w'_{2b}, respectively and produce reconstructions $\hat{\mathbf{x}}_{1b}$ and $\hat{\mathbf{x}}_{2b}$. Later, decoder 2 learns $\mathbf{x}_b$ precisely through feed-forward and tries to decode $\hat{\mathbf{x}}_{1b}$ using the knowledge of $(\mathbf{x}_b, \hat{\mathbf{x}}_{2b})$. To facilitate this, the encoder might need to send some extra bits to decoder 2 in addition to w'_{2b}. These extra bits sent to decoder 2 are represented as an additional index w''_{2b} from an appropriately chosen codebook of rate $2^{R''_2 N/2}$. The total rate R_2 sent to decoder 2 is thus $R'_2 + R''_2$. In summary, the encoder sends w_{1b} to decoder 1, (w'_{2b}, w''_{2b}) to decoder 2 and $(w_{1b}, w'_{2b}, w''_{2b})$ to decoder 0.
- Step B: Assume $\mathbf{u}_B = \mathbf{U}(i)$. Set $j = 1$ and do encoding as in previous steps: pick $(w_{1b}, w'_{2b}) \in \{1, \ldots, M_1\} \times \{1, \ldots, M'_2\}$ such that $(\mathbf{x}_b, \mathbf{u}_b, \hat{\mathbf{X}}_1^i(w_{1b}), \hat{\mathbf{X}}_2^i(w'_{2b})) \in T_\epsilon(\mathbf{X}, \mathbf{U}, \hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2)$ *and* $\hat{\mathbf{X}}_1^i(w_{1b})$ is in the first cell of the $\hat{\mathbf{X}}_1^i$ codebook.

Decoding: Since there is a growing amount of information available at the decoder (due to feed-forward), the time-line of observations at the encoder and decoder is important. Recall that a source sample is available to the decoder $N + k$ time units after it is produced. The time-line of various events at the encoder and decoder with $k = 1$ is shown in Table I.

- *Step* b ($b = 1, 2, \ldots, B$): At time instant $(b + 1)N/2$, $w_{1b}, (w'_{2b}, w''_{2b}), (w_{1b}, w'_{2b}, w''_{2b})$ are received by the decoders $1, 2, 0$, respectively. As described in the sequel, $\mathbf{u}_b = \mathbf{U}(i)$ has been decoded in the previous step just before time $(b+1)N/2$. Therefore, the appropriate codebooks $\hat{\mathbf{X}}_1^i, \hat{\mathbf{X}}_2^i$ are used and reconstructions $\hat{\mathbf{x}}_{1b}, \hat{\mathbf{x}}_{2b}, \hat{\mathbf{x}}_{0b}$

are produced using w_{1b}, w'_{2b} and (w_{1b}, w'_{2b}), respectively. By time instant $(b + 2)N/2$, decoder 2 has received the first b $N/2$-blocks of source samples $\mathbf{x}_1, \ldots, \mathbf{x}_b$ through feed-forward [3]. Decoder 2 now decodes the reconstruction of decoder 1- it tries to find $\hat{\mathbf{x}}_{1b}$ from the $\hat{\mathbf{X}}_1^i$ codebook such that $(\mathbf{x}_b, \mathbf{u}_b, \hat{\mathbf{x}}_{1b}, \hat{\mathbf{x}}_{2b}) \in T_\epsilon(\mathbf{X}, \mathbf{U}, \hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2)$. If there is more than one $\hat{\mathbf{x}}_{1b}$ satisfying the condition, w''_{2b} resolves the list. The cell number j^* of $\hat{\mathbf{x}}_{1b}$ determines $\mathbf{u}_{b+1} = \mathbf{U}(j^*)$. Thus by time instant $(b+2)N/2$, all three decoders have decoded $\mathbf{u}_{b+1}$.

Probability of Error: For our coding strategy, we will declare an error in block b ($b = 1, \ldots, B$) if one or more of the following events occur.

1) Event E_1: The source vector $\mathbf{x}_b$ is not a typical sequence with respect to P_X.
2) E_2: The encoder cannot find $j \in \{1, \ldots, M_0\}$ such that $\mathbf{U}(j)$ is jointly typical with $\mathbf{x}_{b+1}$.
3) E_3: Assuming $\mathbf{u}_b = \mathbf{U}(i), \mathbf{u}_{b+1} = \mathbf{U}(j)$, the encoder cannot find a $(\hat{\mathbf{x}}_{1b}, \hat{\mathbf{x}}_{2b})$ such that $(\mathbf{x}, \hat{\mathbf{x}}_{1b}, \hat{\mathbf{x}}_{2b}, \mathbf{u}_b)$ is jointly typical *and* $\hat{\mathbf{x}}_{1b}$ is in the jth cell of its codebook.
4) E_4: Decoder 2 is unable to decode $\hat{\mathbf{x}}_{1b}$ correctly with knowledge of $(\mathbf{x}_b, \hat{\mathbf{x}}_{2b})$ and w''_{2b}.

The probability of error for block b is

$$P_{eb} = P(E_1 \cup E_2 \cup E_3 \cup E_4).$$

Consider any $\epsilon > 0$. We can bound the probability of each each event for large N as follows. For large $N/2$, with high probability $\mathbf{x}_b$ is typical with respect to P_X. Thus, $P(E_1) < \epsilon/4$.

For $b = 1, \ldots, B-1$ and large $N/2$, there exists a codebook $\{\mathbf{U}(j), j \in \{1, \ldots, M_0\}\}$ such that with high probability, at least one codeword is jointly typical with $\mathbf{x}_{b+1}$ iff $M_0 > 2^{I(X;U)N/2}$. Hence for sufficiently large $N/2$, $P(E_2) < \epsilon/4$ if

$$R_0 > I(X; U). \tag{14}$$

To compute $P(E_3)$, we first note that given $\mathbf{u}_b = \mathbf{U}(i), \mathbf{u}_{b+1} = \mathbf{U}(j)$, we need to find an $\hat{\mathbf{x}}_{1b}$ from the jth cell of $\hat{\mathbf{X}}_1^i$ codebook (a cell has $2^{(R_1 - R_0)N/2}$ codewords) and an $\hat{\mathbf{x}}_{2b}$ from the $\hat{\mathbf{X}}_2^i$ codebook ($2^{R'_2 N/2}$ codewords) such that $(\hat{\mathbf{x}}_{1b}, \hat{\mathbf{x}}_{2b}) \in T_\epsilon(\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2 | \mathbf{X}, \mathbf{U})$. Using arguments similar to the proof in [1], we can show that this is possible with high

---

[3]We can assume $N >> k$, so that receiving $N/2 - k$ source samples is equivalent to receiving the entire block.

probability if

$$R_1 - R_0 > I(X; \hat{X}_1|U)$$
$$R_2^{'} > I(X; \hat{X}_1|U)$$
$$R_1 - R_0 + R_2^{'} > I(X; \hat{X}_1|U) + I(X; \hat{X}_2|U) \qquad (15)$$
$$+ I(\hat{X}_1; \hat{X}_2|XU)$$

Thus $P(E_3) < \epsilon/4$ if (15) is satisfied.

Finally, to compute $P(E_4)$, we need to measure the uncertainty of decoder 2 about $\hat{x}_{1b}$ when it knows $(\mathbf{x}_b, \hat{x}_{2b})$. Assuming there was no encoding error, i.e. $(E_1 \cup E_2 \cup E_3)^c$, the $\hat{\mathbf{X}}_{1b}$ chosen by the encoder is jointly typical with $(\mathbf{x}_b, \hat{x}_{2b})$. The probability that another random $\hat{\mathbf{X}}_{1b} \in T_\epsilon(\hat{\mathbf{X}}_1|\mathbf{U})$ is jointly typical with a random pair $(\mathbf{X}_b, \hat{\mathbf{X}}_{2b}) \in T_\epsilon(\mathbf{X}, \hat{\mathbf{X}}_2|\mathbf{U})$ is approximately $2^{-I(\hat{X}_1; X\hat{X}_2|U)N/2}$ for large $N/2$. We condition on U throughout since $\mathbf{u}_b$ is known to all decoders. Thus the number of other $\hat{\mathbf{X}}_1$ codewords that are jointly typical with the known pair $(\mathbf{x}_b, \hat{x}_{2b})$ is approximately

$$M_1 \cdot 2^{-I(\hat{X}_1; X\hat{X}_2|U)N/2} = 2^{(R_1 - I(\hat{X}_1; X\hat{X}_2|U))N/2} \qquad (16)$$

for large $N/2$. Thus if $R_1 > I(\hat{X}_1; X\hat{X}_2|U)$, $w_{2b}^{''}$ has to resolve a list whose size is given by (16); otherwise there is nothing to resolve. Hence we can have $P(E_4) < \epsilon/4$ if the rate $R_2^{''}$ of the extra index satisfies

$$R_2^{''} > \max\{0, R_1 - I(\hat{X}_1; X\hat{X}_2|U)\} \qquad (17)$$

From the above arguments, using the union bound we have

$$P_{be} < \epsilon, \quad b = 2, \ldots, B. \qquad (18)$$

It should be noted here that in the first step, we arbitrarily set $\mathbf{u}_1 = \mathbf{U}(1)$. In general, $\mathbf{u}_1$ will not be jointly typical with $\mathbf{x}_1$. Consequently, for the first block alone $P(E_2) = 1$. Thus if (14), (15) and (17) are satisfied, the average probability of error over the B blocks can be expressed as

$$P_e = \frac{1}{B}\sum_{b=1}^{B} P_{eb} < \frac{1}{B}(1 + \epsilon \ldots + \epsilon) = \frac{1}{B} + \frac{(B-1)\epsilon}{B} < 2\epsilon$$

for large enough B.

The central decoder 0 can either produce $\hat{\mathbf{X}}_{0b}$ as a function of $(\hat{\mathbf{X}}_{1b}, \hat{\mathbf{X}}_{2b})$, or $\mathbf{X}$ can be quantized to a $\hat{\mathbf{X}}_0$ codebook conditioned on the knowledge of decoder 0. No extra bits are needed in the first case. In the latter case, it can be shown [1] that the extra rate needed by the central decoder is $I(X; \hat{X}_0|\hat{X}_1, \hat{X}_2, U)$. This overhead needs to be shared between the rates R_1 and R_2. Note that the first situation is a special case of the second since $I(X; \hat{X}_0|\hat{X}_1, \hat{X}_2, U) = 0$ if $\hat{X}_0$ is a function of $\hat{X}_1, \hat{X}_2$.

We combine this shared overhead with the rates specified by (14), (15) and (17), and recognize that $R_2 = R_2^{'} + R_2^{''}$ to obtain the rate region of Theorem 1.

REFERENCES

[1] T. M. Cover and A. E. Gamal, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 851–857, November 1982.
[2] H. Witsenhausen, "On source networks with minimal breakdown degradation," *Bell Syst. Tech. J.*, vol. 59, pp. 1083–1087, July-August 1980.
[3] L. H. Ozarow, "On the source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, pp. 1909–1922, December 1980.
[4] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 721–726, November 1985.
[5] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 502–521, July 1987.
[6] Z. Zhang and T. Berger, "Multiple description source coding with no excess marginal rate," *IEEE Trans. Inform. Theory*, vol. 41, pp. 349–357, March 1995.
[7] F. Fu and R. W. Yeung, "On the rate-distortion region for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2012–2021, July 2002.
[8] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding for many channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2106–2114, September 2003.
[9] S. S. Pradhan, R. Puri, and K. Ramchandran, "n -channel symmetric multiple descriptions Part 1: (n; k) sourcechannel erasure codes," *IEEE Trans. Inform. Theory*, vol. 50, pp. 47–61, January 2004.
[10] R. Puri, S. S. Pradhan, and K. Ramchandran, "n-channel symmetric multiple descriptionspart 2: An achievable ratedistortion region," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1377–1392, April 2005.
[11] S. S. Pradhan, "On the role of feedforward in gaussian sources: Point-to-point source coding and multiple description source coding," *IEEE Trans. Inform. Theory*, vol. 53, pp. 331–349, January 2007.
[12] E. Martinian and G. W. Wornell, "Source Coding with Fixed Lag Side Information," *Proc. 42nd Annual Allerton Conference (Monticello, IL)*, 2004.
[13] R. Venkataramanan and S. S. Pradhan, "Source coding with feedforward: Rate-distortion theorems and error exponents for a general source," *IEEE Trans. Inform. Theory*, vol. 53, pp. 2154–2179, June 2007.
[14] T. Weissman and N. Merhav, "On competitive prediction and its relation to rate-distortion theory," *IEEE Trans. Inform. theory*, vol. IT-49, pp. 3185–3194, December 2003.
[15] T. Berger and Z. Zhang, "Minimum breakdown degradation in binary source encoding," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 807–814, November 1983.
[16] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic Press,, 1981.
[17] T. Cover and C. S. K. Leung, "An achievable rate-region for the multiple-access channel with feedback," *IEEE Trans. Inform. theory*, vol. IT-27, pp. 292–298, May 1981.
[18] F. M. J. Willems and E. C. Van der Muelen, "Partial feedback for the discrete memoryless multiple access channel," *IEEE Trans. Inform. theory*, vol. 29, pp. 287–290, March 1983.