Achievable Rates for Multiple Descriptions With Feed-Forward

Ramji Venkataramanan and S. Sandeep Pradhan, *Member, IEEE*

***Abstract*—The two-channel multiple descriptions problem for an independent and identically distributed (i.i.d.) source, with feed-forward to one or both side-decoders is considered. A single-letter achievable rate-region is derived; it enlarges the best known rate-region for multiple descriptions without feed-forward. The proof of the result uses a block-Markov superposition source coding strategy. In point-to-point source coding, feed-forward does not decrease the rate-distortion function of an i.i.d. source. In contrast, an example is provided to show that the derived region can be strictly larger than the optimal multiple description rate-distortion region without feed-forward.**

***Index Terms*—Feed-forward, multiple descriptions, rate-distortion region.**

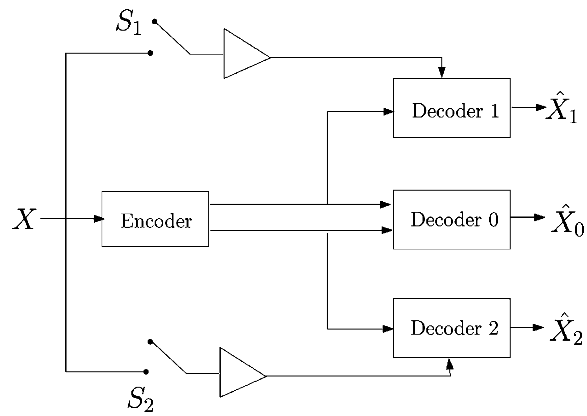

Fig. 1. The multiple descriptions problem.

I. Introduction

THE multiple descriptions problem, first posed by Gersho, Ozarow, Witsenhausen, and others, can be understood through the following example. 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 never reach 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 receiving either packet, but would like a better reconstruction if both packets are received. How should we compress the source into two different descriptions?

The multiple descriptions setup is shown in Fig. 1. In the standard problem, switches S_1 and S_2 are both open. X 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/sample 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, e.g., [1]–[14]. 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 standard lossy source coding problem, there is a source X that has to be reconstructed with some distortion D. The encoder takes 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 setup feed-forward with delay k.

Table I shows the time-line of events for a feed-forward system with block length five and a delay of one time unit. At time instant 5, the source has produced samples $(X_1, \ldots, X_5)$ which the encoder compresses into an index W, available instantaneously at the decoder. At time 6, the decoder reconstructs $\hat{X}_1$ using W, at time 7 it reconstructs $\hat{X}_2$ using W and X_1, and so on. In general given a block length N and a feed-forward delay k, we would like to characterize the rate versus distortion tradeoff. For a fixed k, define the fundamental limit of delay-k feed-forward by taking $N \to \infty$: the minimum achievable rate for a given distortion when the decoder has

Manuscript received April 21, 2009; revised August 31, 2010; accepted September 08, 2010. Date of current version March 16, 2011. This work was supported by the NSF by Grant CCF-0448115 (CAREER). The material in this paper was presented at the IEEE International Symposium on Information Theory, Toronto, ON, Canada, 2008.

R. Venkataramanan was with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48103 USA. He is now with Yale University, New Haven, CT 06520 USA (e-mail: rvenkata@umich.edu).

S. S. Pradhan is with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48103 USA (e-mail: pradhanv@eecs.umich.edu).

Communicated by E.-H. Yang, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2011.2112210

TABLE I
FEED-FORWARD WITH BLOCK LENGTH $N = 5$, DELAY $k = 1$

Time	1	2	3	4	5	6	7	8	9	10
Source	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Encoder	-	-	-	-	W	-	-	-	-	W
Extra info at decoder	-	-	-	-	-	-	X_1	X_2	X_3	X_4
Decoder						$\hat{X}_1$	$\hat{X}_2$	$\hat{X}_3$	$\hat{X}_4$	$\hat{X}_5$

perfect knowledge of all but the last k source samples. In other words, the rate-distortion function with delay-k feed-forward is the optimal rate-distortion tradeoff with block length N, where N can be arbitrarily large.

The notion of feed-forward is applicable to multiterminal problems as well. Fig. 1 shows a multiple descriptions system with feed-forward. Assume 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) as well as the source samples until time $n-k$. In this paper, we study 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 [15]–[18] as a variant of source coding with side information. For example, consider a field to be compressed and communicated from one node to another in a network. This field (e.g., an acoustic field) could propagate through the medium slowly and become available at the destination node 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 [19]. Examples illustrating the connection between feed-forward and prediction can be found in [19], [20]. 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 an equiprobable binary source; Bob, Carol, and Dave are interested in reconstructing the source sequence. Bob and Carol each want to reconstruct with the fraction of their errors being at most d, while Dave needs error-free reconstruction. Alice supplies information at rates R_1 and R_2 to Bob and Carol, respectively; Dave gets the information available to both Bob and Carol. Further assume that after reconstruction of each source sample, Alice reveals to Carol (but not to Bob and Dave) the actual value of the sample. 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. This example is studied in Section III.

In [15], a simple multiple-description coding scheme (based on scalar quantization) was presented for an independent and identically distributed (i.i.d.). Gaussian source with feed-forward to all decoders 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 a discrete memoryless source with feed-forward to one or both side-decoders. This rate region can be achieved with any finite feed-forward delay $k \geq 1$. In point-to-point source coding, feed-forward does not decrease the rate-distortion function of a discrete memoryless source with an additive, memoryless distortion measure [17], [19]. In contrast, for multiple descriptions, we show that the rate-distortion region of a discrete memoryless source can be strictly larger with feed-forward.

In Section II, we define the problem formally and state the main result. The prediction example described earlier is discussed in Section III. Section IV contains the proof of the main result, and Section V concludes the paper.

Notation: Upper-case letters will be used for random variables and lower-case letters for their realizations. Superscript notation such as X^N will be used to denote the random vector $(X_1, \dots, X_N)$. Entropy is measured in bits, and $h(.)$ denotes the binary entropy function.

II. PROBLEM STATEMENT AND MAIN RESULT

Consider a discrete memoryless source X with finite alphabet $\mathcal{X}$. We assume that the source samples $X_n, n = 1, 2, \dots$ are 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

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

The per-letter distortion measures d_m are assumed to have a finite upper-bound $d_{\max}$. The distortion on N-length sequences is the average of the per-letter distortions: for $m = 0, 1, 2$

$$d_m\left(x^N, \hat{x}_m^N\right) \triangleq \frac{1}{N} \sum_{n=1}^{N} d_m(x_n, \hat{x}_{mn}), \quad \forall x^N \in \mathcal{X}^N, \hat{x}_m^N \in \widehat{\mathcal{X}}_m^N. \tag{1}$$

A. Feed-Forward to Only One Decoder

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

Definition 1: A $(N, 2^{NR_1}, 2^{NR_2})$ multiple description code of block length N and rates (R_1, R_2), with delay k feed-forward to decoder 2, consists of:

1) Encoder mappings

$$e_m : \mathcal{X}^N \to \{1, \dots, 2^{NR_m}\}, \; m = 1, 2.$$

2) Mappings for decoders 0 and 1:

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

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

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

The encoder maps each N-length source sequence to a pair of indices in $\{1,\ldots,2^{NR_1}\} \times \{1,\ldots,2^{NR_2}\}$. The decoders receive their respective indices. Once the indices are received, reconstruction takes place *sequentially*, one sample at each time instant. In addition to its index, decoder 2 has access to the source samples until time $(n-k)$ to reconstruct the nth sample. Since decoders 1 and 0 do not receive any feed-forward, their reconstructions are completely determined by the indices they receive. Achievable rates are defined in the usual Shannon sense.

Definition 2: (R_1, R_2) is an achievable rate pair with feed-forward delay k for distortion (D_0, D_1, D_2) if for all $\epsilon > 0$ 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,

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

The rate-distortion region $R^k(D_0, D_1, D_2)$ is the set of achievable rate pairs with feed-forward delay k for distortion (D_0, D_1, D_2).

We emphasize that for a *fixed* k, the rate-distortion region is the set of rates achievable when the block length N can be arbitrarily large. As the delay k increases, the decoder has progressively less information available for decoding. Thus we have

$$\begin{aligned} R^1(D_0, D_1, D_2) \supseteq R^2(D_0, D_1, D_2) \supseteq \cdots \\ \supseteq R^\infty(D_0, D_1, D_2) \end{aligned}$$

where $R^\infty(D_0, D_1, D_2)$ is the rate-distortion region for multiple descriptions without feed-forward, i.e., the delay $k = \infty$.

Our main result is the following theorem, which specifies a set of rates that lie in $R^k(D_0, D_1, D_2)$ for all finite $k \ge 1$.

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

$$\begin{aligned} &R_1 > I(X;\hat{X}_1 U) \\ &R_2 > I(X;\hat{X}_2|U) + \max\{0, R_1 - I(X\hat{X}_2;\hat{X}_1|U)\} \\ &R_1 + R_2 > I(\hat{X}_1;\hat{X}_2|XU) \\ &\quad + \max\{0, R_1 - I(X\hat{X}_2;\hat{X}_1|U)\} \\ &\quad + I(X;\hat{X}_1 U) + I(X;\hat{X}_2|U) + I(X;\hat{X}_0|\hat{X}_1\hat{X}_2 U) \\ &Ed_m(X;\hat{X}_m) \le D_m, \quad m = 0,1,2. \end{aligned} \tag{2}$$

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 this rate region with the rates achievable for multiple descriptions without feed-forward. The multiple descriptions rate-distortion region (without feed-forward) is known only for certain special cases (see [2], [3], [5], [9], and [21]). The best known achievable region for the general two-channel multiple descriptions problem for an i.i.d. source is due to Zhang and Berger [6]. This region is an extension (using an auxiliary random variable) of the rate region obtained by El Gamal and Cover [3]. We reproduce the Zhang-Berger rate region below in a slightly modified, but equivalent, form.

Zhang-Berger Region [6]: 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

$$\begin{aligned} &R_1 > I(X;\hat{X}_1 U), \quad R_2 > I(X;\hat{X}_2 U) \\ &R_1 + R_2 > I(X;\hat{X}_1 U) + I(X;\hat{X}_2 U) \\ &\quad + I(\hat{X}_1;\hat{X}_2|XU) + I(X;\hat{X}_0|\hat{X}_1\hat{X}_2 U) \\ &Ed_m(X;\hat{X}_m) \le D_m, \quad m = 0,1,2. \end{aligned} \tag{3}$$

In general, $\hat{X}_1$ and $\hat{X}_2$ need to be conditionally dependent given X in order to satisfy the distortion constraint at the central decoder. This is achieved in the coding scheme of [6] in two ways. The source X is first quantized to U, which is sent to all the decoders. This requires a rate $I(X;U)$ to each decoder. The reconstructions of decoders 0, 1, 2 are produced conditioned on this cloud center U. The additional correlation needed between $\hat{X}_1$ and $\hat{X}_2$ is given by the term $I(\hat{X}_1;\hat{X}_2|XU)$ in the sum rate.

To see that Theorem 1 enlarges the no-feed-forward rate region (3), consider any set of random variables $U, \hat{X}_1, \hat{X}_2, \hat{X}_0$ jointly distributed with X. Set R_1 at its minimum value, i.e., $R_1 = I(X;\hat{X}_1 U)$. From (3), the minimum achievable Zhang-Berger rate to decoder 2 is

$$R_2^{\mathrm{ZB}} = I(X;\hat{X}_2 U) + I(\hat{X}_1;\hat{X}_2|XU) + I(X;\hat{X}_0|\hat{X}_1\hat{X}_2 U). \tag{4}$$

Let us compare this with the minimum R_2 with feed-forward prescribed by Theorem 1. From the structure of (2), we can have one of two situations:

a) $\max\{0, R_1 - I(X\hat{X}_2;\hat{X}_1|U)\} = 0$: Since $R_1 = I(X;\hat{X}_1 U)$, this happens when $I(X;\hat{X}_1 U) < I(X\hat{X}_2;\hat{X}_1|U)$ or equivalently, when $I(X;U) < I(\hat{X}_2;\hat{X}_1|XU)$. In this case, using Theorem 1 we see that

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

is achievable. Comparing with (4), this represents a savings of $I(X;U)$ bits/sample over the minimum no-feed-forward rate R_2^{ZB}. In other words, feed-forward has helped convey the cloud center U to user 2 without any additional rate.

b) $\max\{0, R_1 - I(X\hat{X}_2;\hat{X}_1|U)\} \ne 0$: Since $R_1 = I(X;\hat{X}_1 U)$, this occurs when $I(X;\hat{X}_1 U) > I(X\hat{X}_2;\hat{X}_1|U)$ or equivalently,

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

when $I(X;U) > I(\hat{X}_2;\hat{X}_1|XU)$. From Theorem 1, we obtain

$$\begin{aligned} R_2 &= I(X;\hat{X}_2|U) + I(X;\hat{X}_1U) - I(X\hat{X}_2;\hat{X}_1|U) \\ &\quad + I(\hat{X}_1;\hat{X}_2|XU) + I(X;\hat{X}_0|\hat{X}_1\hat{X}_2U) \\ &= I(X;\hat{X}_2|U) + I(X;U) + I(X;\hat{X}_0|\hat{X}_1\hat{X}_2U) \\ &= I(X;\hat{X}_2U) + I(X;\hat{X}_0|\hat{X}_1\hat{X}_2U) \end{aligned}$$

is achievable, a savings of $I(\hat{X}_2;\hat{X}_1|XU)$ bits/sample over the no-feed-forward rate given by (4). In this case, feed-forward to user 2 has eliminated the extra correlation term of the Zhang-Berger rate region.

Hence the savings in R_2 due to feed-forward is $\min\{I(X;U), I(\hat{X}_2;\hat{X}_1|XU)\}$. We may interpret the effect of feed-forward as reducing the rate needed to generate the required correlation between $\hat{X}_1$ and $\hat{X}_2$.

B. Feed-Forward to Both Decoders 1 and 2

Switches S_1 and S_2 in Fig. 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 now both decoder 1 and 2 are defined by a sequence of mappings. In addition to the index, both decoders 1 and 2 have access to the source samples until time $(n-k)$.

Achievable rates are defined 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 the convex hull of the union of these two regions is a (possibly larger) achievable rate-region.

A natural question to consider next is whether feed-forward to the central decoder alone is useful. In this setting, we can show that if one of the side-decoders needs to perfectly recover a function of the source, the optimal rate region is given by the El Gamal-Cover rate region [3]. In other words, feed-forward to the central decoder does not improve the optimal rate-distortion region for this special case. The proof of this fact is omitted since it is a simple extension of the proof of [9, Theorem 1]. For the general case, it is not clear how feed-forward to the central decoder can be exploited to achieve lower distortions at the side-decoders.

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 Hamming distortion measure is used. Therefore

$$d_m(x,\hat{x}_m) = \begin{cases} 1, & \hat{x}_m \neq x \\ 0, & \text{otherwise} \end{cases}, \quad m = 0,1,2.$$

Suppose decoders 1 and 2 want to reconstruct X with distortion d, while decoder 0 needs to reconstruct with average distortion 0.[2] We want to characterize the minimum sum-rate

$$r_{sum}(d) \triangleq \inf\{R_1 + R_2 : (R_1, R_2, 0, d, d) \text{ achievable}\}. \quad (5)$$

[2]From (1), note that average distortion d means that the expected normalized Hamming distance between a source sequence and its reconstruction is at most d as the block length goes to infinity, where the expectation is over all source sequences. Thus average distortion 0 indicates that the normalized Hamming distance should go to 0 with high probability.

A lower bound to $r_{sum}(d)$ without feed-forward was obtained in [6, Th. 3, Sect. VIII]:[3]

$$r_{sum}(d)_{no-ff} \geq 2 - h\left(\frac{4d + 1 - \sqrt{12d^2 - 4d + 1}}{2}\right). \quad (6)$$

Let us now assume only decoder 2 gets feed-forward with delay k. For $k = 1$, this is the prediction example discussed in Section I. Let U be a binary-valued random variable and fix the conditional distribution $P_{U,\hat{X}_1,\hat{X}_2,\hat{X}_0|X} = P_{U|X}P_{\hat{X}_1,\hat{X}_2|XU}P_{\hat{X}_0|XU\hat{X}_1\hat{X}_2}$ as follows.

Fix a parameter $D, 2d \leq D \leq 1$ and define

$$\begin{aligned} P_{U|X}(0|0) &= P_{U|X}(1|1) = 1 - D \\ P_{U|X}(0|1) &= P_{U|X}(1|0) = D. \end{aligned} \quad (7)$$

$P_{\hat{X}_1,\hat{X}_2|XU}$ is defined as

$$\begin{aligned} &P_{\hat{X}_1,\hat{X}_2|XU}(00|00) = P_{\hat{X}_1,\hat{X}_2|XU}(11|11) = 1 \\ &P_{\hat{X}_1,\hat{X}_2|XU}(01|01) = P_{\hat{X}_1,\hat{X}_2|XU}(10|01) = d/D \\ &P_{\hat{X}_1,\hat{X}_2|XU}(00|01) = 1 - 2d/D \\ &P_{\hat{X}_1,\hat{X}_2|XU}(01|10) = P_{\hat{X}_1,\hat{X}_2|XU}(10|10) = d/D \\ &P_{\hat{X}_1,\hat{X}_2|XU}(11|10) = 1 - 2d/D. \end{aligned} \quad (8)$$

$\hat{X}_0$ is a function of $(U, \hat{X}_1, \hat{X}_2)$

$$\hat{X}_0 = \begin{cases} \hat{X}_1 & \text{if } (\hat{X}_1 = \hat{X}_2) \\ 1 - U & \text{if } (\hat{X}_1 \neq \hat{X}_2) \end{cases} \quad (9)$$

It is easy to check that this joint distribution achieves the distortion triple $(D_1 = d, D_2 = d, D_0 = 0)$. Using this in Theorem 1, we can obtain an achievable rate-region when only decoder 2 receives feed-forward. The relevant information quantities are calculated below, with $h(\cdot)$ used to denote the binary entropy function.

$$\begin{aligned} I(X;U) &= H(U) - H(U|X) = 1 - h(D). \\ I(\hat{X}_2;\hat{X}_1|XU) &= H(\hat{X}_1|XU) - H(\hat{X}_1|\hat{X}_2XU) \\ &= Dh\left(\frac{d}{D}\right) - D\left(1 - \frac{d}{D}\right)h\left(\frac{d}{D-d}\right). \\ I(X;\hat{X}_2|U) &= I(X;\hat{X}_1|U) = H(\hat{X}_1|U) - H(\hat{X}_1|UX) \\ &= h(D-d) - Dh\left(\frac{d}{D}\right). \\ I(X;\hat{X}_0|\hat{X}_1\hat{X}_2U) &= 0. \end{aligned} \quad (10)$$

Equation (10) contains all the expressions required to compute the rate-region of Theorem 1. Thus for each d, we can select the value D to yield the best rate-constraint and obtain an achievable upper bound to $r_{sum}(d)$ in (5) (with feed-forward to only one decoder). This is plotted in graph (b) of Fig. 2 for distortions $d \geq 0.08$. Graph (a) is the lower bound (6) to $r_{sum}(d)$

[3]There appears to be a typo in the statement of the result in [6, Th. 3]. The correct version (given here) can be obtained from the proof of that theorem.

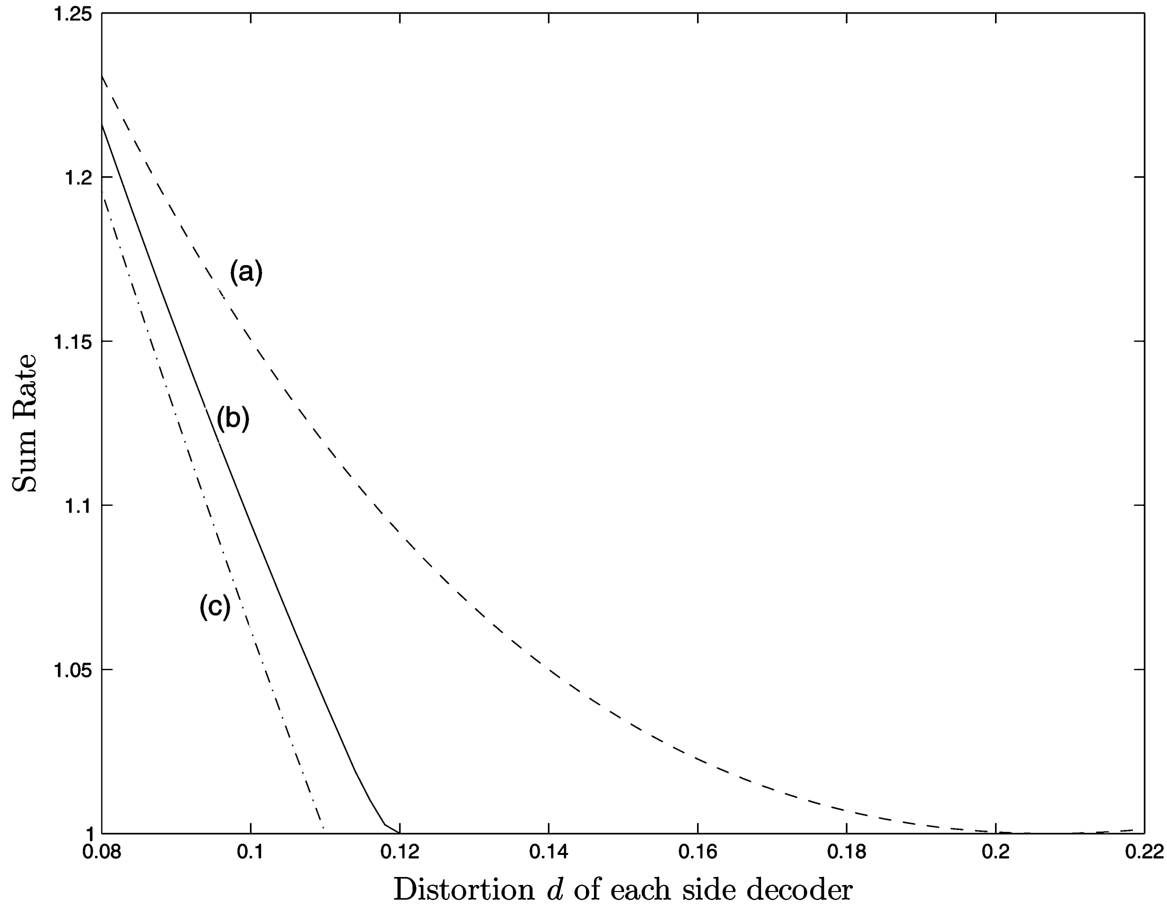

Fig. 2. (a) Lower bound (6) on $r_{sum}(d)$ without FF. (b) Achievable sum-rate with FF to one side-decoder. (c) Rate-distortion lower bound on $r_{sum}(d)$ with FF.

without feed-forward. We see that for all the distortions considered, feed-forward to one side-decoder yields achievable rates smaller than the optimal no feed-forward rate. This is in contrast to point-to-point source coding where feed-forward does not decrease the rate-distortion function of an i.i.d. source with an additive, memoryless distortion measure [17], [19]. Since decoders 1 and 2 produce reconstructions with distortion d, R_1 and R_2 each have to 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 Fig. 2.

Of particular interest is the situation when the sum rate $R_1 + R_2 = 1$. This is the case of *no excess rate* to the central decoder [5]. For this case, it was shown in [4] that without feed-forward, the minimum achievable distortion at each side-decoder is $(\sqrt{2}-1)/2 = 0.207$, which is also the value given by the lower bound (6). In comparison, with feed-forward to one side-decoder, we can achieve $d = 0.12$ with $R_1 + R_2 = 1$ (setting $D = 0.26$, we obtain $R_1 = 0.4986, R_2 = 0.5014$ from Theorem 1).

IV. Proof of Theorem 1

The source sequence is divided into a large number of blocks, say B blocks, with each block containing L source symbols. N, the total block length of the code is therefore equal to BL. For clarity, we will present the proof with delay 1 feed-forward. Thus source samples start being available at decoder 2 one time unit after it receives its index at time N. In other words, a sample produced by the source at time i is available to decoder 2 at time $i+N$. The extension of the proof to feed-forward with arbitrary delay k is straightforward.

To prove the theorem, we shall use the properties of strongly ϵ-typical sequences [22]. Length-L vectors $x^L, \hat{x}_1^L, \hat{x}_2^L$ are said to be jointly typical if their joint type is approximately $P_{X,\hat{X}_1,\hat{X}_2}$. The set of all jointly ϵ-typical tuples $(X^L, \hat{X}_1^L, \hat{X}_2^L)$ is denoted $T_\epsilon(\mathbf{X}, \hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2)$. The set of $\hat{X}_1^L$ sequences conditionally ϵ-typical with $X^L = \mathbf{x}$ is denoted $T_\epsilon(\hat{\mathbf{X}}_1|\mathbf{x})$. In the sequel, bold letters shall be used to denote random vectors, with their length understood to be L. We first present an outline of the coding scheme that explains the main ideas.

Outline: To exploit the feed-forward, we shall use a block-Markov superposition strategy [23], [24] covering pairs of adjacent blocks. While encoding the length-L source block $\mathbf{X}_b, 1 \le b \le B-1$, we would also like to give the decoders a coarse version of $\mathbf{X}_{b+1}$. This is done as follows. The codebook of user 1 is divided into 2^{LR_0} cells of equal size, as shown in Fig. 3. To encode $\mathbf{X}_b$, the encoder first quantizes $\mathbf{X}_{b+1}$ to $\mathbf{U}_{b+1}$ using a U-codebook of size 2^{LR_0}. If j is the chosen quantization index in the U-codebook, encoding for $\mathbf{X}_b$ is restricted to the jth cell of the codebook 1. This is depicted in Fig. 3—the encoder chooses $\hat{\mathbf{X}}_{1b}$ from the jth cell of codebook 1 and $\hat{\mathbf{X}}_{2b}$

TABLE II
TIME-LINE OF EVENTS AT ENCODER AND DECODER WITH FEED-FORWARD

Time instant	L	$2L$	$\ldots$	$BL = N$	$(B+1)L$	$(B+2)L$	$\ldots$	$(B+b)L$	$\ldots$
Source	$\mathbf{x}_1$	$\mathbf{x}_2$	$\ldots$	$\mathbf{x}_B$					
Encoder produces:	$\mathbf{u}_1$ -	$\mathbf{u}_2$ w_{11}, w_{21}	$\ldots$ $\ldots$	$\mathbf{u}_B$ $w_{1(B-1)}, w_{2(B-1)}$					
Decoder 1 has received:				$\mathbf{u}_1, w_{11}, \ldots$ $\ldots, w_{1(B-1)}$	-	-		-	
Decoder 1 produces:				$\hat{\mathbf{x}}_{11}$ $\mathbf{u}_2$	$\hat{\mathbf{x}}_{12}$ $\mathbf{u}_3$	$\hat{\mathbf{x}}_{13}$ $\mathbf{u}_4$	$\ldots$ $\ldots$	$\hat{\mathbf{x}}_{1(b+1)}$ $\mathbf{u}_{(b+1)}$	
Decoder 2 has received:				$\mathbf{u}_1, w_{21}, \ldots$ $\ldots, w_{2(B-1)}$	$\mathbf{x}_1$	$\mathbf{x}_2$	$\ldots$	$\mathbf{x}_b$	
Decoder 2 produces:				$\hat{\mathbf{x}}_{21}$	$\hat{\mathbf{x}}_{11}, \mathbf{u}_2$ $\hat{\mathbf{x}}_{22}$	$\hat{\mathbf{x}}_{12}, \mathbf{u}_3$ $\hat{\mathbf{x}}_{23}$	$\ldots$ $\ldots$	$\hat{\mathbf{x}}_{1b}, \mathbf{u}_{(b+1)}$ $\hat{\mathbf{x}}_{2(b+1)}$	

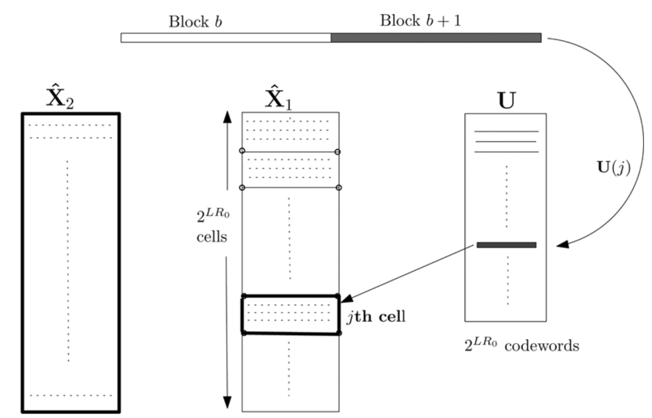

Fig. 3. Restricted encoding for codebook 1: $\mathbf{U}(j)$ is the codeword representing $\mathbf{X}_{b+1}$. The block b codeword for user 1 is chosen within cell j of the $\hat{\mathbf{X}}_1$ codebook.

from codebook 2 such that $(\mathbf{X}_b, \hat{\mathbf{X}}_{1b}, \hat{\mathbf{X}}_{2b})$ are jointly typical. The idea of restricted *decoding* using a nonrandom partition was introduced in [24] for a multiple-access channel with feedback. Here we use restricted encoding with partitioning to exploit feed-forward.

Table II shows the time-line of the information available at each terminal, at time-instants corresponding to the end of each block. The first two rows of the table show that at time $2L$, the source has produced the first two blocks $\mathbf{X}_1, \mathbf{X}_2$. At this time, the encoder quantizes $\mathbf{X}_2$ to $\mathbf{U}_2$ and then produces $(\mathbf{w}_{11}, \mathbf{w}_{21})$—the quantization indices corresponding to the first block—according to the procedure described in the previous paragraph. The encoding proceeds in this fashion until time $N = BL$, when the source has produced block $\mathbf{X}_B$, and the encoder has produced the indices $(\mathbf{w}_{11}, \ldots, \mathbf{w}_{1B-1})$ and $(\mathbf{w}_{21}, \ldots, \mathbf{w}_{2B-1})$. Instantly, the first set of indices is made available to decoder 1, and the second set to decoder 2. Both sets of indices are available to decoder 0.

At time $N = BL$, decoders 1, 2, and 0 reconstruct the first block, producing $\hat{\mathbf{X}}_{11}, \hat{\mathbf{X}}_{21}$ and $\hat{\mathbf{X}}_{01}$, respectively. At this time, decoders 1 and 0 also know $\mathbf{U}_2$ since it is determined by the cell-index of $\hat{\mathbf{X}}_{11}$. This is shown in the fifth row of the table. From time $(N+1) = (BL+1)$ onwards, decoder 2 starts receiving source symbols through feed-forward (recall that feed-forward delay is 1). Therefore, between times BL and $(B+1)L$, decoder 2 receives the block $\mathbf{X}_1$ through feed-forward At time $(B+1)L$, it decodes $\hat{\mathbf{X}}_{11}$ using $\mathbf{X}_1$ and $\hat{\mathbf{X}}_{21}$. Thus at time $(B+1)L$, all decoders know $\mathbf{U}_2$ since it is indexed by the cell of $\hat{\mathbf{X}}_{11}$. The decoding proceeds in this fashion as shown in Table II: at time $(B+b)L$, decoder 2 has received $\mathbf{X}_b$ through feed-forward and uses it to decode $\hat{\mathbf{X}}_{1b}$. Consequently, all decoders know $\mathbf{U}_{b+1}$, which they use to produce $\hat{\mathbf{X}}_{1(b+1)}, \hat{\mathbf{X}}_{2(b+1)}$ and $\hat{\mathbf{X}}_{0(b+1)}$.

$\mathbf{U}$ can be thought of as a cloud center, conditioned on which reconstructions are produced at the decoders. The coding strategy essentially uses the feed-forward to decoder 2 to convey $\mathbf{U}$ parsimoniously to the decoders. The detailed proof is given below.

Random Coding: Let $M_0 = 2^{LR_0}, M_1 = 2^{LR_1}$ and $M_2' = 2^{LR_2'}$. 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 L-vectors $\mathbf{U}$. For each $\mathbf{U}(i)$, choose a codebook of length-L vectors $\{\hat{\mathbf{X}}_1^i(1), \ldots, \hat{\mathbf{X}}_1^i(M_1)\}$, independently according to a uniform distribution over the set $T_\epsilon(\hat{\mathbf{X}}_1|\mathbf{U}(i))$. Similarly choose a codebook $\{\hat{\mathbf{X}}_2^i(1), \ldots, \hat{\mathbf{X}}_2^i(M_2')\}$ from $T_\epsilon(\hat{\mathbf{X}}_2|\mathbf{U}(i))$. We partition each $\hat{\mathbf{X}}_1^i$ codebook into M_0 disjoint cells, so that each cell has M_1/M_0 elements. We have assumed for simplicity that M_1/M_0 is an integer.

Encoding: We encode a source sequence of length $N = BL$ given by

$$[x_1, x_2, \ldots, x_{BL}] = [\mathbf{x}_1 \mathbf{x}_2 \ldots \mathbf{x}_B]$$

where $\mathbf{x}_b$ denotes the bth block of length L, for $b = 1, \ldots B$.

Step 0: Find $i \in \{1, \ldots, M_0\}$ such that $(\mathbf{x}_1, \mathbf{U}(i)) \in T_\epsilon(\mathbf{X}, \mathbf{U})$. Set $\mathbf{u}_1 = \mathbf{U}(i)$. A rate of R_0 to each side-decoder is necessary to convey $\mathbf{u}_1$. This is only needed for the first block, and is a negligible fraction of the total rate when the number of blocks B is large.

Step $b, b = 1, \ldots, B$: From the previous step, $\mathbf{u}_b$ is known. Say it is equal to $\mathbf{U}(i)$. For $1 \leq b < B$: observe the length-L block $\mathbf{x}_{b+1}$ and find a $j \in \{1, \ldots, M_0\}$ so 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 or if $b = B$, set $\mathbf{u}_{b+1} = \mathbf{U}(1)$. Thus we have $\mathbf{u}_b = \mathbf{U}(i), \mathbf{u}_{b+1} = \mathbf{U}(j)$.

Encode $\mathbf{x}_b$ as follows: 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* $\hat{\mathbf{X}}_1^i(w_{1b})$ 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 w'_{2b} to a random index in the $\hat{\mathbf{X}}_2^i$ codebook.

The encoding is depicted in Fig. 3. Note that we restrict ourselves to one cell within the $\hat{\mathbf{X}}_1^i$ codebook. Restricted encoding enables decoder 2 to take advantage of the feed-forward. Decoders 1 and 2 produce reconstructions $\hat{\mathbf{x}}_{1b}$ and $\hat{\mathbf{x}}_{2b}$ using w_{1b} and w'_{2b}, respectively. Later, decoder 2 learns $\mathbf{x}_b$ precisely through feed-forward and tries to decode $\hat{\mathbf{x}}_{1b}$ using $(\mathbf{x}_b, \hat{\mathbf{x}}_{2b})$. 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 appropriate codebook of rate $R_2^{''}$. The total rate R_2 sent to decoder 2 is thus $R_2^{'} + R_2^{''}$.

In summary, at time $N = BL$, the encoding is complete and the encoder sends $(\mathbf{u}_1, \{w_{1b}\}_{b=1}^B)$ to decoder 1 and $(\mathbf{u}_1, \{w'_{2b}, w''_{2b}\}_{b=1}^B)$ to decoder 2. The extra rate that may be needed for central decoder 0 is discussed at the end of the proof.

Decoding: At time instant BL, decoder 1 receives $(\mathbf{u}_1, w_{11}, \ldots, w_{1B})$ and decoder 2 receives $(\mathbf{u}_1, w'_{21}, \ldots w'_{2B}, w''_{21}, \ldots w''_{2B})$. The reconstruction at the two side-decoders, depicted in Table II, proceeds as follows. The generation of $\hat{\mathbf{x}}_{0b}, b = 1, \ldots, B$ is described at the end of the proof.

Step 1 (Executed at Time BL): $\mathbf{u}_1$ is known to all decoders. At time BL, the appropriate codebooks determined by $\mathbf{u}_1$ are used and reconstructions $\hat{\mathbf{x}}_{11}, \hat{\mathbf{x}}_{21}$ are produced using w_{11}, w'_{21}, respectively. In addition, decoder 1 also knows $\mathbf{u}_2$ at this time, since it is determined by the cell-index of $\hat{\mathbf{x}}_{11}$. From time $(N+1) = (BL+1)$ onwards, decoder 2 starts receiving source symbols through feed-forward. By time instant $(B+1)L$, decoder 2 has received the first source block $\mathbf{x}_1$ through feed-forward.

Step $b, b = 2, \ldots, B$ (Executed at Time $(B+b-1)L$): At the end of the previous step, $\hat{\mathbf{x}}_{1(b-1)}$ and $\hat{\mathbf{x}}_{2(b-1)}$ have been decoded by the respective decoders, and $\mathbf{u}_{(b-1)}$ is known at all decoders to be equal to $\mathbf{U}(i)$. By time instant $(B+b-1)L$, decoder 2 has received the source block $\mathbf{x}_{(b-1)}$ through feed-forward. It then decodes the codeword of decoder 1 using this information: it tries to find $\hat{\mathbf{x}}_{1(b-1)}$ from the $\hat{\mathbf{X}}_1^i$ codebook such that $(\mathbf{x}_{(b-1)}, \mathbf{u}_{(b-1)}, \hat{\mathbf{x}}_{1(b-1)}, \hat{\mathbf{x}}_{2(b-1)}) \in T_\epsilon(\mathbf{X}, \mathbf{U}, \hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2)$. If there is more than one $\hat{\mathbf{x}}_{1(b-1)}$ satisfying $w''_{2(b-1)}$ resolves the list.

The cell number j^* of $\hat{\mathbf{x}}_{1(b-1)}$ determines $\mathbf{u}_b = \mathbf{U}(j^*)$. Thus at time $(B+b-1)L$, all decoders know $\mathbf{u}_b$. At this time, the appropriate codebooks determined by $\mathbf{u}_b$ are used and reconstructions $\hat{\mathbf{x}}_{1b}, \hat{\mathbf{x}}_{2b}$ are produced using w_{1b}, w'_{2b}, respectively. Decoder 1 now knows $\mathbf{u}_{(b+1)}$ since it is determined by the cell-index of $\hat{\mathbf{x}}_{1b}$. The time-line of the decoding procedure is shown in the last row of Table II.

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}.

We bound the probability of each event for sufficiently large L as follows. Consider any $\epsilon > 0$. With high probability $\mathbf{x}_b$ is typical with respect to P_X. Thus $P(E_1) < \epsilon/B$.

For $b = 1, \ldots, B-1$, 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 $M_0 > 2^{LI(X;U)}$. Hence $P(E_2) < \epsilon/B$ if

$$R_0 > I(X;U). \tag{11}$$

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 $2^{L(R_1-R_0)}$ codewords) and an $\hat{\mathbf{x}}_{2b}$ from the $\hat{\mathbf{X}}_2^i$ codebook ($2^{LR_2^{'}}$ 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 of [3, Th. 1], it follows that this is possible with high probability (i.e., $P(E_3) < \epsilon/B$) if

$$\begin{aligned} R_1 - R_0 &> I(X;\hat{X}_1|U) \\ R_2^{'} &> I(X;\hat{X}_2|U) \\ R_1 - R_0 + R_2^{'} &> H(\hat{X}_1|U) + H(\hat{X}_2|U) \\ &\quad - H(\hat{X}_1, \hat{X}_2|XU) \\ &= I(X;\hat{X}_1|U) + I(X;\hat{X}_2|U) \\ &\quad + I(\hat{X}_1;\hat{X}_2|XU). \end{aligned} \tag{12}$$

Assuming there was no encoding error, i.e., $(E_1 \cup E_2 \cup E_3)^c$ holds, the $\hat{\mathbf{X}}_{1b}$ chosen by the encoder is jointly typical with $(\mathbf{x}_b, \hat{\mathbf{x}}_{2b})$. 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 $2^{-LI(\hat{X}_1;X\hat{X}_2|U)}$ for large L. Thus, conditioned on $\mathbf{u}_b$, the number of other $\hat{\mathbf{X}}_1$ codewords that are jointly typical with the pair $(\mathbf{x}_b, \hat{\mathbf{x}}_{2b})$ is approximately

$$M_1 \cdot 2^{-LI(\hat{X}_1;X\hat{X}_2|U)} = 2^{L(R_1 - I(\hat{X}_1;X\hat{X}_2|U))}. \tag{13}$$

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 (13). Hence we can have $P(E_4) < \epsilon/B$ if the rate $R_2^{''}$ of the extra index satisfies

$$R_2^{''} > \max\{0, R_1 - I(\hat{X}_1; X\hat{X}_2|U)\} \tag{14}$$

Assume (11), (12), and (14) are satisfied. From the arguments above and the union bound, we see that P_{be}, the probability of error in block b, satisfies $P_{be} < 4\epsilon/B, b = 1, \ldots, B$. The total probability of error over B blocks is

$$P_e = \sum_{b=1}^{B} P_{eb} < \sum_{b=1}^{B} \frac{4\epsilon}{B} < 4\epsilon.$$

Combining (11), (12) and (14), and recognizing that $R_2 = R_2^{'} + R_2^{''}$ we obtain the following rate constraints:

$$\begin{aligned} R_1 &> I(X;\hat{X}_1 U) \\ R_2 &> I(X;\hat{X}_2|U) + \max\{0, R_1 - I(X\hat{X}_2;\hat{X}_1|U)\} \\ R_1 + R_2 &> I(X;\hat{X}_2|U)\} + \max\{0, R_1 - I(X\hat{X}_2;\hat{X}_1|U)\} \\ &\quad + I(X;\hat{X}_1 U) + I(\hat{X}_1;\hat{X}_2|XU) \end{aligned} \tag{15}$$

Finally for $1 \leq b \leq B$, conditioned on the knowledge of central decoder 0, i.e., $(\mathbf{u}_b, \hat{\mathbf{x}}_{1b}, \hat{\mathbf{x}}_{2b}), \mathbf{x}_b$ can be quantized to $\hat{\mathbf{x}}_{0b}$. The extra rate required for this representation is $I(X; \hat{X}_0 | \hat{X}_1, \hat{X}_2, U)$. This overhead (to be conveyed to the central decoder) can be shared between the rates R_1 and R_2. Adding this overhead to the rate constraints specified in (15) gives the region of Theorem 1.

V. Conclusion

In the multiple descriptions problem, the distortion constraint of the central decoder dictates how reconstructions $\hat{X}_1$ and $\hat{X}_2$ need to be correlated. Feed-forward to one of the side-decoders can reduce the rate required to induce this correlation. The coding scheme presented in this paper uses feed-forward to one decoder only. It is worth exploring how one can do better in the presence of feed-forward to both side-decoders. This is especially interesting because it has been shown [15] that no excess rate is needed for an i.i.d. Gaussian source with feed-forward to both side-decoders, i.e., we can achieve the optimal rate-distortion function at all three decoders simultaneously.

Acknowledgment

The authors would like to thank the associate editor and the anonymous reviewers for their comments and suggestions, which led to a much improved manuscript.

References

[1] J. K. Wolf, A. D. Wyner, and J. Ziv, "Source coding for multiple descriptions," *Bell Syst. Tech. J.*, vol. 59, p. 1980, Oct. 1980.
[2] L. H. Ozarow, "On the source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, pp. 1909–1922, Dec. 1980.
[3] A. E. Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.
[4] T. Berger and Z. Zhang, "Minimum breakdown degradation in binary source encoding," *IEEE Trans. Inf. Theory*, vol. IT-29, pp. 807–814, Nov. 1983.
[5] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inf. Theory*, vol. IT-31, pp. 721–726, Nov. 1985.
[6] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-33, pp. 502–521, Jul. 1987.
[7] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 821–834, 1993.
[8] Z. Zhang and T. Berger, "Multiple description source coding with no excess marginal rate," *IEEE Trans. Inf. Theory*, vol. 41, pp. 349–357, Mar. 1995.
[9] F. Fu and R. W. Yeung, "On the rate-distortion region for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 48, pp. 2012–2021, Jul. 2002.
[10] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding for many channels," *IEEE Trans. Inf. Theory*, pp. 2106–2114, Sep. 2003.
[11] R. Puri, S. S. Pradhan, and K. Ramchandran, "n-channel symmetric multiple descriptions—Part 2: An achievable rate-distortion region," *IEEE Trans. Inf. Theory*, pp. 1377–1392, Apr. 2005.
[12] J. Chen, C. Tian, and S. Diggavi, "Multiple description coding for stationary Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2868–2881, Jun. 2009.
[13] C. Tian and J. Chen, "A novel coding scheme for symmetric multiple description coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5344–5365, Oct. 2010.
[14] H. Wang and P. Viswanath, "Vector Gaussian multiple description with two levels of receivers," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 401–410, 2009.
[15] S. S. Pradhan, "On the role of feedforward in Gaussian sources: Point-to-point source coding and multiple description source coding," *IEEE Trans. Inf. Theory*, pp. 331–349, Jan. 2007.
[16] E. Martinian and G. W. Wornell, "Source coding with fixed lag side information," in *Proc. 42nd Ann. Allerton Conf.*, 2004.
[17] R. Venkataramanan and S. S. Pradhan, "Source coding with feed-forward: Rate-distortion theorems and error exponents for a general source," *IEEE Trans. Inf. Theory*, pp. 2154–2179, Jun. 2007.
[18] N. Merhav and T. Weissman, "Coding for the feedback Gelfand-Pinsker channel and the feedforward Wyner-Ziv source," *IEEE Trans. Inf. Theory*, vol. 52, pp. 4207–4211, Sep. 2006.
[19] T. Weissman and N. Merhav, "On competitive prediction and its relation to rate-distortion theory," *IEEE Trans. Inf. Theory*, vol. IT-49, pp. 3185–3194, Dec. 2003.
[20] R. Venkataramanan and S. S. Pradhan, "On computing the feedback capacity of channels and the feed-forward rate-distortion function of sources," *IEEE Trans. Commun.*, vol. 58, no. 7, pp. 1889–1896, Jul. 2010.
[21] E. Ahmed and A. B. Wagner, "Erasure multiple descriptions," *IEEE Trans. Inf. Theory*, 2010, submitted for publication.
[22] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[23] T. Cover and C. S. K. Leung, "An achievable rate-region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. IT-27, pp. 292–298, May 1981.
[24] F. M. J. Willems and E. C. Van der Muelen, "Partial feedback for the discrete memoryless multiple access channel," *IEEE Trans. Inf. Theory*, vol. 29, pp. 287–290, Mar. 1983.

Ramji Venkataramanan received the B.Tech. degree from the Indian Institute of Technology, Madras, in 2002, and the Ph.D. degree in electrical engineering from the University of Michigan, Ann Arbor, in 2008.

He is currently a Postdoctoral Research Associate with Yale University, New Haven, CT. His research interests include information theory, coding, and stochastic network theory.

S. Sandeep Pradhan (M'10) received the M.E. degree from the Indian Institute of Science in 1996 and the Ph.D. degree from the University of California at Berkeley in 2001.

From 2002 to 2008, he was an Assistant Professor with the Department of Electrical Engineering and Computer Science, University of Michigan at Ann Arbor, where he is currently an Associate Professor. His research interests include sensor networks, multiterminal communication systems, coding theory, quantization, and information theory.

Dr. Pradham is the recipient of 2001 Eliahu Jury award given by the University of California at Berkeley for outstanding research in the areas of systems, signal processing, communications and control, the CAREER award given by the National Science Foundation (NSF), and the Outstanding achievement award for 2009 from the University of Michigan.