On the role of feedforward in Gaussian sources: Point-to-point source coding and multiple description source coding †

S. Sandeep Pradhan
Department of Electrical Engineering and Computer Science
University of Michigan at Ann Arbor
email: pradhanv@eecs.umich.edu.

January 10, 2005

Abstract

The model of source coding with noiseless feedforward deals with efficient representation (quantization) of information sources into indexes belonging to a finite set, where to reconstruct a source sample, the decoder in addition to this index, has access to all the previous noiseless source samples. This problem has applications in sensor networks, economics and control theory. In the fist part of this paper we consider a deterministic block coding scheme for independent and identically distributed (IID) Gaussian sources with noiseless feedforward. We show that this scheme is asymptotically optimal in terms of its rate-distortion function and the source coding error exponent.

In the second part of this paper we consider 2-channel multiple description source coding with noiseless feedforward. In particular we consider IID Gaussian sources and obtain the optimal rate-distortion region, by giving a deterministic scheme that is asymptotically optimal. The key result is that unlike the case where there is no feedforward, here there is no penalty to be paid for constraining the descriptions to be mutually refineable. That is when one of the channels is active, the decoder which operates on one of the descriptions achieves the optimal rate-distortion function, and when both channels are active, the joint decoder still attains the optimal rate-distortion function. This implies that for memoryless sources with additive distortion measures, unlike the case of point-to-point source coding where noiseless feedforward does not improve the optimal rate-distortion function, in the case of multiple description source coding, noiseless feedforward does indeed improve the optimal rate-distortion region. The proposed scheme is based on linear processing and uniform scalar quantization. We then show that the proposed scheme achieves the optimal multiple description source coding error exponents for the symmetric case where the rates of the descriptions are equal and the reconstruction distortion is only a function of the number of descriptions received.

1 Introduction

With the recent emergence of applications related to sensor networks [1], efficient encoding of information signals in a multiterminal setting has received special attention. One such problem is that of source coding with side information at the decoder [2], where the encoder wishes to represent a source X with an index belonging to a finite set to be transmitted to a decoder which has access to some correlated side information Y. The decoder wishes to obtain an estimate of the source with the help of the received index and the side information. In

†This work was supported in part by NSF ITR grant CCF 0329715. This work was presented in part at the IEEE International Symposium on Information Theory (ISIT), Chicago, IL, June 2004.

many applications involving estimation of some information field (such as seismic, acoustic), the signal to be estimated at the decoder is delay-sensitive, and might be available to the decoder either in a delayed and/or noisy form. Thus an efficient encoder must take into account all the information available at the decoder at the time of decoding a particular sample.

The implicit assumption in the model of source coding with side information is that the underlying sample pairs (X_i, Y_i) are instantaneously observed respectively at the encoder and the decoder. So after an encoding delay of l samples, when the decoder gets the message W (say being transmitted instantaneously using electromagnetic waves), it has access to the corresponding l samples of Y, so that the decoding can begin immediately. The time-line of the samples of the source, the message and the side information is depicted in Fig. 1 for $l = 5$. Note that in this model, for example, at the 6th time unit, the decoder reconstructs $\hat{X}_1, \ldots \hat{X}_5$ simultaneously as a function of W and $Y_1, \ldots Y_5$, though it may display them as shown in Fig. 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
Side info	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}
Decoder						$\hat{X}_1$	$\hat{X}_2$	$\hat{X}_3$	$\hat{X}_4$	$\hat{X}_5$

Figure 1: Time-line: instantaneous observations

What happens if the underlying signal field is traveling slowly (compared to the velocity of electromagnetic wave propagation) from the geographical location of the encoder to that of the decoder, so that a there is a delay between the instant when ith sample of X is observed at the encoder and the instant when corresponding ith sample of Y is observed at the decoder, with the additional constraint that the reconstruction be realtime. In that case, we need a different dynamic compression model which is depicted in Fig. 2. Here it is assumed that the signal field delay is 6 time units, so that for real-time reconstruction of the ith source sample, all the past samples of the side information are available. In other words, now the decoding operation consists of a sequence of functions such that the ith reconstruction is a function of W and $(i-1)$ side information samples. The encoding operation, however, remains as in [2], i.e., a mapping from the l-product source alphabet to an index set of size 2^{NR} where R is the rate of transmission. This general compression model takes this important physical signal delay into account in its real-time reconstruction. We refer to this model as source coding with feedforward [3, 4]. Note that in this problem, the encoder is not realtime and the decoder is realtime. In this

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
Side info	–	–	–	–	–	–	Y_1	Y_2	Y_3	Y_4
Decoder						$\hat{X}_1$	$\hat{X}_2$	$\hat{X}_3$	$\hat{X}_4$	$\hat{X}_5$

Figure 2: Time-line: delayed observations

work, as a first step, we consider an idealized version of this problem called source coding with noiseless feedforward [5]. In this model, we assume that noiseless source samples are available with a delay at the decoder, i.e. $Y = X$. From Fig. 2, it is clear that the model with $Y = X$ is meaningful only when the delay is at least $l+1$, where the block length is l. However, for a general Y, any delay leads to a valid problem. In the first part of this paper we consider Gaussian source coding with noiseless feedforward in a point-to-point setting. We provide a deterministic block-coding scheme with linear processing and scalar quantization that is asymptotically optimal in terms of achieving the rate-distortion function and the source coding error exponents.

In the second part we consider a multiterminal extension of this problem. In this paper we formulate the problem of multiple description source coding with noiseless feedforward and obtain the optimal rate-distortion region and the optimal error exponent region for the Gaussian sources. Here the encoder wishes to represent the source into 2 descriptions (for the case of 2-channel multiple descriptions) which are mutually refineable. The decoding structure involves three decoders having access to 3 different subsets of these two descriptions (except the null set). The goal is to produce these descriptions which simultaneously give as low distortions for the three decoders as possible. All the decoders are realtime while having access to delayed noiseless source samples. In other words, the two descriptions can be synergistically combined to give a reconstruction of the source that is more enhanced than that produced individually.

Related and prior Work: The problem of source coding with noiseless feedforward arose in a different context of competitive prediction in [4], where it was shown that for IID discrete sources with additive distortion measures, feedforward does not reduce the optimal rate-distortion function and does not increase the optimal error exponent with block coding. The model of source coding with general feedforward was defined in [3, 5] as a variant of the problem of source coding with side information at the decoder, and preliminary results of Gaussian source coding with noiseless feedforward were reported. In [6], an elegant variable-length coding strategy to achieve the optimal Shannon rate-distortion bound for any finite-alphabet IID source with feedforward was presented, along with a beautiful illustrative example. In [7], the optimal rate-distortion function and the random coding error exponents for general discrete sources were obtained for noiseless feedforward with arbitrary delay. Loosely speaking, for the case when the delay is $(l+1)$, the optimal rate-distortion function is given by the minimum of directed information [8, 9, 10] from the reconstruction sequence to the source sequence, which is given by

$$I(\hat{X}^l \to X^l) = \sum_{k=1}^{l} I(\hat{X}^k; X_k | X^{k-1}), \tag{1}$$

where $I(\cdot;\cdot)$ denotes mutual information [11]. It was also shown that feedforward does improve the optimal rate-distortion function and the optimal error exponents for sources with memory and/or with non-additive distortion measures.

The problem of source coding with feedforward is also related to source coding with a delay-dependent distortion function [12] and causal source coding [13] and real-time source coding (see [14] and references therein). Further, note that the notion of directed information is also very similar to certain measures of linear dependence used in economics to determine causality of certain interaction between two time series [15, 16, 17]. For example, one may wish to measure the amount by which prime interest rates affects unemployment rate. Hence the model of source coding with feedforward might have further connections to economics and control theory.

Multiple description source coding without feedforward has been studied in great detail in the literature [18, 19, 20, 21, 22, 23, 24, 25] motivated by applications to robust information transmission over packet erasure networks. The optimal rate-distortion region for Gaussian sources with mean squared error distortion measure was determined by [19] for the case of two descriptions. A design of multiple description quantizers was proposed in [22]. For a recent tutorial on the constructive approaches see [26]. In [27], multiple description source coding with side information at the decoder is considered.

In summary, the main contribution of this paper are the following:

(1) We obtain a deterministic block coding scheme which achieves the optimal Shannon rate-distortion function for IID Gaussian sources with mean squared error distortion measure.

(2) We show that this scheme is optimal in terms of its source coding error exponent.

(3) We give a formulation of the problem of multiple description source coding with noiseless feedforward and evaluate the optimum rate-distortion region for Gaussian sources with mean squared error distortion measure for two channels with a deterministic block coding theorem. In particular it is shown that there is

no penalty to be paid for constraining the descriptions to be mutually refineable. That is, each of the three decoders can achieve their respective optimal point-to-point rate-distortion functions simultaneously.

(4) In the case of symmetric multiple descriptions, we show that the above deterministic coding is optimal in terms of its multiple description source coding error exponents. In particular, it is shown that each of the three decoders can achieve their respective optimal point-to-point source coding error exponents simultaneously.

In this paper, we consider only fixed-rate source coding error exponents along with block quantization systems. However, we refer to them as just error exponents for conciseness. We remark here that although in the point-to-point setting, the presence of feedforward does not improve upon Shannon's rate-distortion function and the optimal source coding error exponents for IID sources with additive distortion measures, there is a dramatic reduction in the complexity of the encoding and decoding operations. Further, in the multiterminal setting, even for IID sources with additive distortion measures, feedforward does improve the optimal rate-distortion region. We also note that the problem of source coding with feedforward can be considered as a functional dual [11, 28, 29, 30, 31] of the problem of channel coding with feedback. Hence the block-coding schemes that are proposed for source coding with feedforward in this paper can be considered as some sort of duals to the schemes considered in [32, 33, 34] (also see [35]) for channel coding with feedback both for the point-to-point as well as the multiterminal settings.

The paper is organized as follows. In Section 2 we formulate the two problems as mentioned above. In Section 3 we consider point-to-point source coding with noiseless feedforward, and in Section 4 we consider 2-channel multiple description source coding with noiseless feedforward, and Section 5 concludes the paper.

2 Problem Formulation

Consider a stationary discrete memoryless source X with a probability distribution $p(x)$ with some alphabet $\mathcal{X}$, and a reconstruction alphabet $\hat{\mathcal{X}}$. The encoder observes a sequence of independent realizations of the source from the given distribution. Associated with the source, there is a distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$. The distortion measure for a pair of sequences of length l is the average of the distortions of l samples: $d(\mathbf{x}, \hat{\mathbf{x}}) = (1/l)\sum_{i=1}^{l} d(x_i, \hat{x}_i)$, where x_i and $\hat{x}_i$ denote the ith samples of $\mathbf{x}$ and $\hat{\mathbf{x}}$ respectively.

Definition 1: A code with parameters (l, Θ, Φ) for source coding with noiseless feedforward would involve an encoding function:

$$F : \mathcal{X}^l \to \{1, 2, \ldots, \Theta\}, \tag{2}$$

and a sequence of decoding functions for $k = 1, 2, \ldots, l$,

$$G_k : \{1, 2, \ldots, \Theta\} \times \mathcal{X}^{k-1} \to \hat{\mathcal{X}}, \tag{3}$$

such that $Ed(\mathbf{X}, \mathbf{G}(\mathbf{X})) \leq \Phi$, where $\mathbf{G}(\mathbf{X})$ denotes the l-length reconstruction vector.

That is, the decoder receives the index transmitted by the encoder, and to reconstruct the ith sample (for $i = 1, 2, \ldots, l$), it has access to all the past samples of the source till $(i-1)$. The goal is to minimize $E\left[d\left(\mathbf{X}, \mathbf{G}(\mathbf{X})\right)\right]$ for a given rate $R = (1/l)\log\Theta$.

Definition 2: A tuple (R, D) is said to be achievable for source coding with feedforward if $\forall \nu > 0$, there exists a code for sufficiently large l, with parameters (l, Θ, Φ) such that $\frac{1}{l}\log\Theta < R + \nu$ and $\Phi < D + \nu$.

Let $R_{ff}(D)$ denotes the infimum of R such that (R, D) is achievable. A schematic of this problem is shown in Fig. 1. In this paper we consider a special case of IID Gaussian source with zero-mean and variance σ^2, and

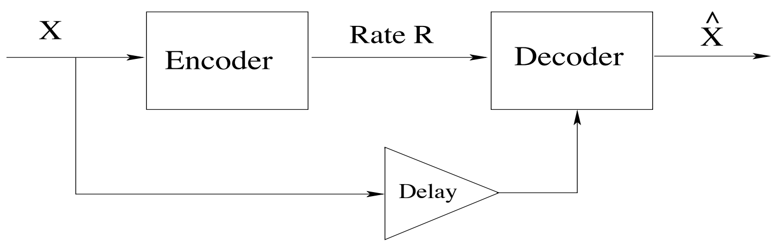

Figure 3: Source coding with feedforward: the decoder, to reconstruct any source sample, has access to all the previous samples in addition to the quantized version of the source.

with mean squared error as the distortion measure.

Consider multiple description source coding in the presence of noiseless feedforward as shown in Fig. 4. The

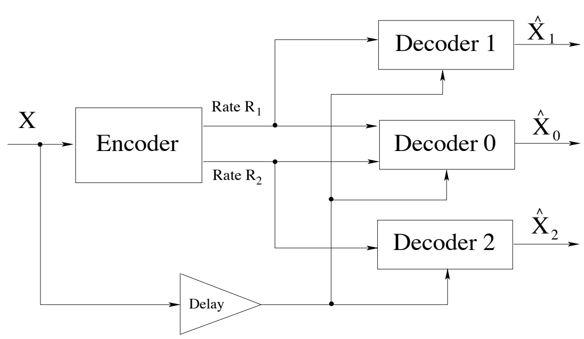

Figure 4: A schematic of 2-channel multiple description source coding with noiseless feedforward: delay=$l+1$, blocklength of the encoder=l.

general problem of 2-channel multiple description source coding with noiseless feedforward can be formulated as follows. Let X_k for $k = 1, 2, \ldots$ be a sequence of IID random variables drawn according to a known distribution $p(x)$, and let $\mathcal{X}$ denote its alphabet. We are given three reconstruction alphabets $\hat{\mathcal{X}}_0$, $\hat{\mathcal{X}}_1$ and $\hat{\mathcal{X}}_2$. There are three distortion measures: $d_i : \mathcal{X} \times \hat{\mathcal{X}}_i \to \mathbb{R}^+$ for $i = 0, 1, 2$. The distortion measure on l-sequences is given by the average per-symbol distortion.

Definition 3: A code with parameters $(l, \Theta_1, \Theta_2, \Phi_0, \Phi_1, \Phi_2)$ for the problem of 2-channel multiple descriptions source coding with noiseless feedforward would involve two encoding functions:

$$F_i : \mathcal{X}^l \to \{1, 2, \ldots, \Theta_i\} \text{ for } i = 1, 2, \tag{4}$$

and three sequences of decoding functions for $k = 1, 2, \ldots, l$:

$$G_{ik} = \{1, 2, \ldots, \Theta_i\} \times \mathcal{X}^{k-1} \to \hat{\mathcal{X}}_i, \quad \text{for } i = 1, 2, \text{ and} \tag{5}$$

$$G_{0k} = \{1, 2, \ldots, \Theta_1\} \times \{1, 2, \ldots, \Theta_2\} \times \mathcal{X}^{k-1} \to \hat{\mathcal{X}}_0, \tag{6}$$

such that for $i = 0, 1, 2$,

$$Ed(\mathbf{X}, \mathbf{G}_i(\mathbf{X})) \leq \Phi_i. \tag{7}$$

In other words, for $i = 0, 1, 2$, the ith decoder has access to previous noiseless source samples to reconstruct the present sample. The encoders are not real-time, whereas the decoders are so.

Definition 4: A tuple $(R_1, R_2, D_0, D_1, D_2)$ is said to be achievable if for arbitrary $\nu > 0$, there exists for sufficiently large l, a code with parameters $(l, \Theta_1, \Theta_2, \Phi_0, \Phi_1, \Phi_2)$ such that

$$\Theta_i \leq 2^{l(R_i+\nu)} \quad \text{and } \Phi_j \leq D_j + \nu \text{ for } i = 1, 2 \text{ and } j = 0, 1, 2. \tag{8}$$

The goal is to find the optimal rate-distortion region which is given by the convex closure of the set of all achievable tuples. For the rest of the paper we will devote our attention to IID Gaussian sources with mean squared error distortion measure.

3 Point-to-point source coding with noiseless feedforward

Notation: The proofs of all lemmas in this paper appear in the Appendix. All logarithms are with respect to the natural number e. Upper case letters denote random variables and matrices. Bold faced letters denote vectors.

Recall that the optimal Shannon rate-distortion function $R(D)$ at distortion D, *without feedforward* for IID Gaussian sources with zero mean and variance σ^2 is given by

$$R(D) = \frac{1}{2}\log\frac{\sigma^2}{D}. \tag{9}$$

In the following we describe a deterministic fixed-rate block-coding scheme involving an encoder-decoder pair with a structure as given above. Consider a parametrized version of this function where for any $\beta > 1$, the rate and distortion point $(R(\beta), D(\beta))$ on this function is given by $(\log(\beta), \sigma^2/\beta^2)$. When $\beta \approx 1$, then rate is nearly zero and when $\beta \approx \infty$, then distortion is nearly zero. In the following we show that $\forall \epsilon' > 0$ and $\epsilon > 0$ and $\beta > 1$, there exists a sufficiently large number $l(\beta, \epsilon, \epsilon')$, such that the proposed scheme achieves distortion of $D \leq \sigma^2/\beta^2 + \epsilon$ with rate $R = (1+\epsilon')\log(\beta)$ when the block-length l associated with the scheme is chosen such that $l > l(\beta, \epsilon, \epsilon')$.

3.1 Encoder and decoder

Note that the encoder has to quantize l source samples taking into account the fact that to reconstruct any source sample, all the past source samples are available. Further, after the reconstruction of each source sample, the decoder observes the corresponding sample for the reconstruction of the next sample. Hence the decoder has access to the exact value of the quantization noise unlike the case when there is no feedforward, where the decoder never learns about the quantization noise. The encoder has to exploit this and communicate important information to the decoder through this quantization noise. In essence, the encoder has to anticipate the information that would be requested by the decoder, and slowly convey it through the quantization noise of each source sample, in addition to the index transmitted.

Fix $\beta > 1$, and sufficiently small $\epsilon > 0$ and $\epsilon' > 0$. Consider the following function of l source samples:

$$Y = X_1 + \frac{X_2}{\beta} + \frac{X_3}{\beta^2} + \frac{X_4}{\beta^3} + \ldots + \frac{X_l}{\beta^{l-1}}, \tag{10}$$

where l denotes the block-length of the scheme.

It can be easily shown that Y is Gaussian with mean 0 and variance given by

$$E(Y^2) = \frac{\sigma^2(1-\beta^{-2l})}{1-\beta^{-2}} \leq \frac{\sigma^2\beta^2}{\beta^2-1}. \tag{11}$$

For the first block of l source samples, Y will be a function of $\{X_1, X_2, \ldots, X_l\}$. Consider a uniform scalar quantizer with M levels and bounded between $-\Delta/2$ and $\Delta/2$, where Δ and M will be determined later. Thus the step size of this quantizer is Δ/M. The encoder quantizes only Y using the above quantizer and the index of the cell containing it is sent to the decoder without entropy coding. Let $\hat{Y}$ denotes the quantized version of Y, where the set of mid-points of the quantization cells denote the alphabet of $\hat{Y}$.

The decoder reconstruction is given by the following scheme:

$$\hat{X}_i = \beta(\hat{X}_{i-1} - cX_{i-1}), \tag{12}$$

for $i = 2, \ldots, l$ and $\hat{X}_1 = c\hat{Y}$, where c is a constant which will be determined later. The encoder and decoder start over for the next block of l samples.

The intuition behind this strategy is that since the later source samples get more help in terms of feedforward, we give a boost for the initial source samples by choosing the coefficients in the linear combination to follow a geometric progression.

3.2 Asymptotic Analysis

Theorem 1: For the proposed quantization scheme, if we choose $c = \frac{\beta^2-1}{\beta^2}$, $M = \beta^{l(1+\epsilon')}$, and $\Delta = \beta^{l\epsilon'}$, then the average expected distortion satisfies

$$\lim_{l\to\infty} \frac{1}{l}\sum_{k=1}^{l} E(X_k - \hat{X}_k)^2 = \frac{\sigma^2}{\beta^2}. \tag{13}$$

Proof: Let us denote the quantization noise as $Q = \hat{Y} - Y$. Now we calculate the average expected distortion which is given by

$$\frac{1}{l}\left[\sum_{i=1}^{l} E(X_i - \hat{X}_i)^2\right]. \tag{14}$$

Note that each of the terms in this average expected distortion, constitutes three components: a) distortion due to quantization, b) distortion due to the collapse of the l-length source vector into a scalar Y, and c) a crossterm induced by the first two components. To simplify the exposition, we consider these three constituents separately, and evaluate them.

Let for $k = 1, 2, \ldots, l$,

$$S_k = (1-c)X_k - c\sum_{m=k+1}^{l} \frac{X_m}{\beta^{m-k}}. \tag{15}$$

Now the kth term in this summation can be written as

$$\begin{aligned} E(X_k - \hat{X}_k)^2 &= E(S_k - c\beta^{k-1}Q)^2 && (16)\\ &\le E(S_k^2) + c^2\beta^{2(k-1)}E(Q^2) + 2|c|\beta^{k-1}\left[E(S_k^2)E(Q^2)\right]^{1/2}. && (17) \end{aligned}$$

Note that

$$E(S_k^2) = (1-c)^2\sigma^2 + \frac{c^2\sigma^2\left(1-\beta^{-2(l-k)}\right)}{\beta^2-1}. \tag{18}$$

Hence the average distortion can be written as

$$\frac{1}{l}\sum_{k=1}^{l} E(X_k - \hat{X}_k)^2 \le (1-c)^2\sigma^2 + \frac{c^2\sigma^2}{\beta^2-1} - \frac{\beta^2c^2\sigma^2\left(1-\beta^{-2l}\right)}{l(\beta^2-1)^2} + \frac{c^2E(Q^2)(\beta^{2l}-1)}{l(\beta^2-1)} + \xi, \tag{19}$$

where ξ denotes the crossterm. Our strategy is to make the contribution of the quantization noise in the average distortion to go to zero asymptotically as a function of the block-length l. This then implies that ξ also goes to zero as $l \to \infty$. We then optimize the rest of the terms (which are asymptotically non-zero) with respect to c. Let us choose $M = \beta^{l(1+\epsilon')}$, $\Delta = \beta^{l\epsilon'}$. Let $\mathbb{E}$ denote the event that $|Y| > \Delta/2$. Thus we have

$$E(Q^2|\mathbb{E}^c) \leq \frac{\Delta^2}{M^2} = \frac{1}{\beta^{2l}}. \tag{20}$$

We have the following lemma for the complementary event.

Lemma 1:

$$P(\mathbb{E})E(Q^2|\mathbb{E}) \leq \frac{8(E(Y^2))^{3/2}}{\sqrt{2\pi}\Delta} \exp\left[-\frac{\Delta^2}{32E(Y^2)}\right]. \tag{21}$$

Using this lemma we note that

$$\begin{aligned}
\frac{\beta^{2l}}{l}E(Q^2) &\leq \frac{\beta^{2l}}{l}E(Q^2|\mathbb{E}^c) + \frac{\beta^{2l}}{l}P(\mathbb{E})E(Q^2|\mathbb{E}) && (22)\\
&\leq \frac{1}{l} + \frac{8\sigma^3\beta^3\beta^{l(2-\epsilon')}}{l\sqrt{2\pi}(\beta^2-1)^{3/2}} \exp\left[-\frac{\beta^{2l\epsilon'}(\beta^2-1)}{32\sigma^2\beta^2}\right] && (23)\\
&\to 0 \text{ as } l \to \infty. && (24)
\end{aligned}$$

Further, it can be easily shown that $(1-c)^2\sigma^2 + \frac{c^2\sigma^2}{\beta^2-1}$ achieves its minimum of $\frac{\sigma^2}{\beta^2}$ when $c = (\beta^2-1)/\beta^2$. Next by noting that

$$E(S_k^2) \leq \sigma^2, \quad \forall k = 1, 2, \ldots, l, \tag{25}$$

and using (24) we provide an upper bound on the cross term ξ as follows:

$$\begin{aligned}
\xi &= \frac{2c}{l}\sum_{k=1}^{l}\beta^{k-1}\left[E(S_k^2)E(Q^2)\right]^{1/2} && (26)\\
&\leq \frac{2c\sigma}{l}\sqrt{E(Q^2)}\sum_{k=1}^{l}\beta^{k-1} && (27)\\
&\leq \frac{2c\sigma}{l(\beta-1)}\beta^l\sqrt{E(Q^2)} && (28)\\
&\to 0 \text{ as } l \to \infty. && (29)
\end{aligned}$$

Hence

$$\lim_{l\to\infty}\frac{1}{l}\sum_{k=1}^{l}E(X_k - \hat{X}_k)^2 = \frac{\sigma^2}{\beta^2}. \tag{30}$$

Thus we have shown that the proposed scheme is asymptotically optimal in terms of achieving the optimal rate-distortion function.

□

Before proceeding further, in summary, we note that with noiseless feedforward, for arbitrary rates, the proposed scheme uses uniform bounded scalar quantization followed by linear processing to asymptotically (as a function of blocklength) achieve the optimal rate-distortion function. Although the fraction of the average distortion contributed by the quantization noise is asymptotically zero, enough bits have to be expended to make this contribution decay to zero sufficiently fast as a function of blocklength. Hence the parameters of the scalar quantizer have been chosen such that this behavior is manifested.

3.3 Source coding error exponents with noiseless feedforward

In this section we consider the source coding error exponent of the proposed scheme. Note that for the above scheme the cross-term ξ goes to zero for large blocklength l. To make the large deviation analysis tractable, we consider a small modification to the above quantization scheme which results in the cross term being exactly equal to 0.

Definition 5: For a given rate R and target distortion D, a real number E is said to be an achievable error exponent for source coding with feedforward, if $\forall \nu > 0$, sufficiently small, there exists a source code with feedforward with parameters (l, Θ, Φ) such that

$$\frac{1}{l}\log\Theta \leq R+\nu, \quad \text{and} \quad -\frac{1}{l}\log P\left[\frac{1}{l}\sum_{k=1}^{l}(X_k-\hat{X}_k)^2 > D\right] \geq E-\nu. \tag{31}$$

The optimal source coding error exponent $E_{ff}(R,D)$ is defined as the infimum of E such that E is an achievable error exponent for rate R and distortion D.

Recall that the source coding error exponent [36, 37, 38] (also see [39] and the references therein) *without feedforward* for the same source is given by

$$E(R,D) = \frac{1}{2}\left[\frac{D2^{2R}}{\sigma^2} - 1 - \log\left(\frac{D}{\sigma^2}\right) - 2R\right]. \tag{32}$$

Modified encoder and decoder: Fix $\beta > 1$ and block-length l as before, and let

$$Y' = aX_1 + \frac{X_2}{\beta} + \frac{X_3}{\beta^2} + \ldots + \frac{X_l}{\beta^{l-1}}, \tag{33}$$

where a is some constant which will be determined later. The encoder quantizes Y' using a uniform scalar quantizer with M levels and bounded between $-\Delta/2$ and $\Delta/2$. Let $\hat{Y}'$ denote the quantized version of Y' with mid-points of the cells being used as before. Let $M = \beta^{l(1+\epsilon')}$ and $\Delta = \beta^{l\epsilon'}$ as before. The decoder is given by

$$\hat{X}_i = \beta(\hat{X}_{i-1} - cX_{i-1}), \tag{34}$$

for $i = 3, 4, \ldots, l$, $\hat{X}_1 = (1/a)\hat{Y}'$, and $\hat{X}_2 = \beta ac(\hat{X}_1 - X_1)$. Let $c = (\beta^2 - 1)/\beta^2$, and $Q = \hat{Y}' - Y'$.

Our goal is to choose a such that the crossterm is exactly equal to 0. Further, the asymptotic behavior of this scheme is the same as that of the scheme of the previous section. It is evident that only the reconstruction of X_1 is different in these two schemes. Hence, for the modified scheme define $S'_k = S_k$ for $k = 2, 3, \ldots, l$ and

$$S'_1 = -\frac{1}{a}\sum_{m=2}^{l}\frac{X_m}{\beta^{m-1}}. \tag{35}$$

Now by noting that $E(X_k - \hat{X}_k)^2 = E(S'_k - c\beta^{k-1}Q)^2$ for $k = 2, 3, \ldots, l$, and $E(X_1 - \hat{X}_1)^2 = E(S'_1 - (1/a)Q)^2$, we have

$$-\frac{2S'_1 Q}{a} = \frac{2Q}{a^2}\sum_{m=2}^{l}\frac{X_m}{\beta^{m-1}}, \text{ and} \tag{36}$$

$$-\sum_{k=2}^{l} 2cS'_k Q\beta^{k-1} = -2cQ\sum_{m=2}^{l}\frac{X_m}{\beta^{m-1}}. \tag{37}$$

Hence to make the crossterm equal to 0, we choose

$$a = \sqrt{\frac{\beta^2}{\beta^2 - 1}}. \tag{38}$$

Now, one can note that the asymptotic performance remains unchanged as

$$\lim_{l\to\infty}\frac{1}{l}\sum_{k=1}^{l}E(X_k - \hat{X}_k)^2 = \frac{\sigma^2}{\beta^2}. \tag{39}$$

Now we present the main result of this section.

Theorem 2: The above scheme achieves the optimal source coding error exponent with feedforward, which is given by

$$E(R, D) = \frac{De^{2R}}{2\sigma^2} - \frac{1}{2} - \frac{1}{2}\log\frac{D}{\sigma^2} - R. \tag{40}$$

Proof: First we prove that (i) $\forall R > \log(\beta)$, the probability that a source word is not reconstructed with average distortion less than $\frac{\theta\sigma^2}{\beta^2}$, for any $\theta > 1$, decays exponentially fast with block-length, and the exponent depends only on θ, and is independent of R, σ^2 and β, and (ii) the exponent for $R = \log(\beta)$ is 0 $\forall\theta > 1$.

Now, the source distortion has only two components, one contributed by quantization and the other by the dimensional collapse. In other words

$$\frac{1}{l}\sum_{k=1}^{l}(X_k - \hat{X}_k)^2 < \frac{1}{l}\sum_{k=1}^{l}(S'_k)^2 + \frac{\beta^{2l}}{l}Q^2. \tag{41}$$

By using the following two lemmas we get an exponential bound on the first component and a doubly exponential bound on the second component.

Lemma 2: $\forall \theta > 1$, we have

$$P\left[\frac{1}{l}\sum_{i=1}^{l}(S_i')^2 > \frac{\theta\sigma^2}{\beta^2}\right] \leq e^{-l[E^*(\theta)]+o(l)}, \tag{42}$$

where

$$E^*(\theta) = \left(\frac{\theta-1}{2}\right) - \frac{1}{2}\log\theta, \tag{43}$$

and $o(x)$ denotes a function where $\forall q > 0$, $\exists x_0 > 0$, such that $\forall x > x_0$, implies $0 \leq o(x) < qx$. Note that $E^*(\theta)$ is a monotone increasing function of θ for $\theta > 1$.

Lemma 3: For any $\delta > 0$, sufficiently small,

$$P\left[\frac{\beta^{2l}}{l}Q^2 > \delta\right] \leq s_1 e^{-s_2 e^{l s_3}}, \tag{44}$$

where s_1, s_2 and s_3 are functions of β, l and ϵ':

$$s_1 = \frac{4\sigma\beta}{\sqrt{\pi}\beta^{l\epsilon'}\sqrt{\beta^2-1}}, \quad s_2 = \frac{\beta^2-1}{16\beta^2\sigma^2} \quad \text{and} \quad s_3 = 2\epsilon'\log\beta. \tag{45}$$

Note that $s_i > 0$ for $i = 1, 2, 3$. Now, the error exponents of the proposed scheme can be evaluated as follows. Fix some $\theta > 1$, and a sufficiently small $\delta > 0$ such that $\frac{(\theta^2-1)\sigma^2}{\beta^2} > \delta$. Consider the following set of inequalities:

$$\begin{aligned}
P\left[\frac{1}{l}\sum_{i=1}^{l}(X_i - \hat{X}_i)^2 > \frac{\theta\sigma^2}{\beta^2}\right] &\leq P\left[\frac{1}{l}\sum_{i=1}^{l}S_i^2 + \frac{\beta^{2l}}{l}Q^2 > \frac{\theta\sigma^2}{\beta^2}\right] && (46)\\
&\leq P\left[\frac{1}{l}\sum_{i=1}^{l}S_i^2 > \frac{\theta\sigma^2}{\beta^2} - \delta\right] + P\left[\frac{\beta^{2l}}{l}Q^2 > \delta\right] && (47)\\
&\leq e^{-l\left[E^*\left(\theta - \frac{\delta\beta^2}{\sigma^2}\right)\right]+o(l)} + s_1 e^{-s_2 e^{l s_3}}. && (48)
\end{aligned}$$

Since the above inequality is true $\forall\delta$ such that $\frac{(\theta^2-1)\sigma^2}{\beta^2} > \delta > 0$, the source coding error exponent is given by $E^*(\theta)$. Further note that this exponent is valid for $R > \log(\beta)$. When $R = \log(\beta)$, we have $s_3 = 0$, implying that the exponent is 0.

We will now restate this in the standard form. Fix a target distortion D and rate R, such that $R > \frac{1}{2}\log\left[\frac{\sigma^2}{D}\right]$. Fix also a perturbation parameter $\nu > 0$, and sufficiently small. We now need to choose parameters θ and β such that 1) θ is as large as possible, 2) $\frac{\theta\sigma^2}{\beta^2} = D$, and 3) $R > \log\beta$.

Choosing $\beta = e^{R-\nu}$ we have

$$\theta = \frac{De^{2(R-\nu)}}{\sigma^2}. \tag{49}$$

Hence, $\forall\nu > 0$, and sufficiently small, we have

$$\lim_{l\to\infty} -\frac{1}{l}\log P\left[\frac{1}{l}\sum_{i=1}^{l}(X_i - \hat{X}_i)^2 > D\right] \geq \frac{De^{2(R-\nu)}}{2\sigma^2} - \frac{1}{2} - \frac{1}{2}\log\frac{D}{\sigma^2} - (R-\nu). \tag{50}$$

This implies that for distortion D and rate R such that $R > \frac{1}{2}\log(\sigma^2/D)$, the source coding error exponent for the proposed scheme is equal to

$$\frac{De^{2R}}{2\sigma^2} - \frac{1}{2} - \frac{1}{2}\log\frac{D}{\sigma^2} - R. \tag{51}$$

The optimality follows from the fact that in the case of point-to-point source coding, noiseless feedforward does not increase the error exponent for IID sources with additive distortion measures [4, 7].

□

At a high level, note that the error event, i.e., the event that a source word is not reconstructed with distortion less than D, is the union of two error events. The first error event, which can be considered as the encoder error event, happens when Y' goes beyond the bound of the scalar quantizer. The second event, which can be considered as the decoder error event, happens when the first event does not occur and the distortion due to dimensional collapse exceeds D. The first event decays doubly exponentially fast (which follows from the proof of Lemma 3) and the second event decays exponentially fast with blocklength.

It is also worth noting the following interesting properties of the proposed deterministic scheme. The index of the quantization cell containing Y' is transmitted to the decoder without any entropy coding, yet the proposed scheme approaches the optimal rate-distortion bound. This implies that the entropy of the quantization index per source sample asymptotically approaches R nats for all sufficiently large l.

4 2-channel Multiple description source coding with noiseless feedforward

First, recall from [19] that the optimal rate-distortion region for 2-channel multiple description source coding *without feedforward* for the IID Gaussian source with variance σ^2 is given by $D_1 \geq \sigma^2 e^{-2R_1}$, $D_2 \geq \sigma^2 e^{-2R_2}$ and

$$D_0 \geq \begin{cases} \frac{\sigma^2 e^{-2(R_1+R_2)}}{1-(\sqrt{\Pi}-\sqrt{\Xi})^2} & \text{if } \Pi \geq \Xi \\ \sigma^2 e^{-2(R_1+R_2)} & \text{otherwise,} \end{cases} \tag{52}$$

where $\sigma^4\Pi = (\sigma^2 - D_1)(\sigma^2 - D_2)$ and $\sigma^4\Xi = D_1 D_2 - \sigma^4 e^{-2(R_1+R_2)}$.

In this paper we restrict our attention to Gaussian sources with mean squared error criterion, i.e., $X_k \sim \mathcal{N}(0,\sigma^2)$, with $\hat{\mathcal{X}}_i = \mathbb{R}$ and $d_i(x,\hat{x}) = (x-\hat{x})^2$ for $i = 0,1,2$. In the following we obtain the optimal rate-distortion region for this problem. We provide a deterministic scheme that achieves the boundary of this region. We also evaluate the multiple description source coding error exponents for the proposed scheme for the symmetric case.

4.1 Asymptotic Analysis

Theorem 3: For an IID Gaussian source with variance σ^2 and squared error distortion measure, the optimal rate-distortion region for the problem of 2-channel multiple description source coding with noiseless feedforward is given by

$$D_1 \geq \sigma^2 e^{-2R_1}, \quad D_2 \geq \sigma^2 e^{-2R_2}, \quad D_0 \geq \sigma^2 e^{-2(R_1+R_2)}. \tag{53}$$

In the following we provide a direct and a converse coding theorems. One of the implications of this theorem is that although for memoryless sources with additive distortion measures, the presence of noiseless feedforward does not reduce the optimal rate-distortion function in the point-to-point source coding, it does improve the optimal rate-distortion region for such sources in the multiterminal setting. Further, there is no penalty to be paid for constraining the descriptions to be mutually refineable for the Gaussian sources. It remains to be seen whether the following proposed method can be extended for general n-channel multiple descriptions where $n > 2$.

Deterministic direct coding theorem: Consider a parametrized version of the boundary of the above rate-distortion region. Let $\beta_1 > 1$ and $\beta_2 > 1$. We propose to achieve the rate-distortion tuple:

$$\left(\log(\beta_1), \log(\beta_2), \frac{\sigma^2}{\beta_1^2\beta_2^2}, \frac{\sigma^2}{\beta_1^2}, \frac{\sigma^2}{\beta_2^2}\right). \tag{54}$$

In other words, we show that $\forall \epsilon' > 0$ and $\epsilon > 0$, and $\beta_1 > 1$, $\beta_2 > 1$, there exists a sufficiently large number $l(\beta_1, \beta_2, \epsilon', \epsilon)$ such that the proposed scheme will achieve distortions of $D_1 \leq \sigma^2/\beta_1^2 + \epsilon$, $D_2 \leq \sigma^2/\beta_2^2 + \epsilon$, $D_0 \leq \sigma^2/(\beta_1^2\beta_2^2) + \epsilon$, with rates $R_i = (1 + \epsilon') \log(\beta_i)$ for $i = 1, 2$.

Fix $\beta_1 > 1$, $\beta_2 > 1$ and sufficiently small $\epsilon > 0$ and $\epsilon' > 0$. Consider the following linear combinations of $X_1, X_2, \ldots, X_l$ for all large even l:

$$Y_1 = X_1 + \frac{X_2}{\beta_1} + \frac{X_3}{\beta_1^2} + \frac{X_4}{\beta_1^3} + \frac{X_5}{\beta_1^4} + \frac{X_6}{\beta_1^5} + \ldots + \frac{X_l}{\beta_1^{l-1}}, \tag{55}$$

$$Y_2 = -X_1 + \frac{X_2}{\beta_2} - \frac{X_3}{\beta_2^2} + \frac{X_4}{\beta_2^3} - \frac{X_5}{\beta_2^4} + \frac{X_6}{\beta_2^5} + \ldots + \frac{X_l}{\beta_2^{l-1}}. \tag{56}$$

Note that for $i = 1, 2$, Y_i is Gaussian with zero mean and variance $\frac{\sigma^2(1-\beta_i^{-2l})}{1-\beta_i^{-2}}$. The variance of Y_i is bounded from above by $\sigma^2\beta_i^2/(\beta_i^2 - 1)$. Now for $i = 1, 2$, consider a uniform scalar quantizer E_i with M_i levels and bounded between $-\Delta_i/2$ and $\Delta_i/2$, where M_i and Δ_i will be determined later. The encoder simply scalar quantizes Y_i with E_i, and the index of the cell containing Y_i is sent on the ith channel without entropy coding for $i = 1, 2$. Note that the step size of E_i is Δ_i/M_i. Let $\hat{Y}_i$ denote the quantized version of Y_i, where the midpoints of the quantization cells denote the alphabet of $\hat{Y}_i$ for $i = 1, 2$.

Consider the following three decoders. Let $\hat{X}_{ik}$ denote the reconstruction of Decoder-i for the kth source sample. Decoder-1 uses the following reconstructions:

$$\hat{X}_{1k} = \beta_1(\hat{X}_{1,(k-1)} - c_1 X_{k-1}) \quad \text{for } k = 2, 3, \ldots, l, \tag{57}$$

and Decoder-2 uses the following reconstructions:

$$\hat{X}_{2k} = -\beta_2(\hat{X}_{2,(k-1)} - c_2 X_{k-1}) \quad \text{for } k = 2, 3, \ldots, l, \tag{58}$$

$$\hat{X}_{11} = c_1 \hat{Y}_1, \quad \text{and} \quad \hat{X}_{21} = -c_2 \hat{Y}_2. \tag{59}$$

Decoder-0 uses the following reconstructions:

$$\hat{X}_{0k} = \frac{c_3}{c_1}\hat{X}_{1k} + \frac{c_4}{c_2}\hat{X}_{2k} \quad \text{for } k = 1, 2, 3, \ldots, l. \tag{60}$$

In the following we optimize the distortions over c_1, c_2, c_3 and c_4.

The intuition behind this strategy of encoding is the following. Since later samples get more help from noiseless feedforward, we want to give a boost for the initial samples as in the case of point-to-point source coding. Hence the weights in the linear combination are chosen to satisfy a geometric progression. Further, it turns out that the asymptotic performance of such a system for point-to-point transmission does not depend on the sign of these weights. So to help Decoder-0, we choose the signs of these weights in such a way that the distortion due to the dimensional collapse for Decoder-0 is minimized. Further, as in the case of point-to-point source coding, here too, the encoder has to slowly communicate important information to the three decoders through the quantization noises of each source sample, in addition to the two descriptions transmitted, while maintaining the ability of these descriptions to be mutually refineable.

Since we are quantizing only two random variables, Y_1 and Y_2, there are three components in the distortions of the decoders (as in the previous section): the first due to the quantization noise, the second due to the dimensional collapse, and the third due to the crossterm. In the following we carry out the analysis of the distortions for the three decoders. At a high level, our strategy is to make the contribution due to the quantization and the crossterm to go to zero as $l \to \infty$. One more point which is worth noting here is that we do not use entropy coding while attaining the optimal rate-distortion region as in the previous section.

Let us denote the quantization noise as $Q_i = \hat{Y}_i - Y_i$. Using the analysis of Section 3, by choosing $c_1 = (\beta_1^2 - 1)/\beta_1^2$, $M_1 = \beta_1^{l(1+\epsilon')}$ and $\Delta_1 = \beta_1^{l\epsilon'}$, it can be shown that

$$\lim_{l \to \infty} \frac{1}{l} \sum_{k=1}^{l} E(X_k - \hat{X}_{1k})^2 = \frac{\sigma^2}{\beta_1^2}. \tag{61}$$

Using similar analysis it can be shown that

$$\lim_{l\to\infty} \frac{1}{l}\sum_{k=1}^{l} E(X_k - \hat{X}_{2k})^2 = \frac{\sigma^2}{\beta_2^2}, \tag{62}$$

by choosing $c_2 = (\beta_2^2 - 1)/\beta_2^2$, $M_2 = \beta_2^{l(1+\epsilon')}$ and $\Delta_2 = \beta_2^{l\epsilon'}$. Note that for the chosen quantizers, $E(Q_i^2)\frac{\beta_i^{2l}}{l} \to 0$ as $l \to \infty$.

Now let us consider the average distortion for Decoder-0. Let for $k = 1, 2, \ldots, l$,

$$S_{0k} = (1 - c_3 - c_4)X_k - \sum_{m=k+1}^{l} \left(\frac{c_3}{\beta_1^{m-k}} + (-1)^{m-k}\frac{c_4}{\beta_2^{m-k}} \right) X_m. \tag{63}$$

Note that

$$E(S_{0k}^2) = (1 - c_3 - c_4)^2\sigma^2 + \frac{\sigma^2 c_3^2(1 - \beta_1^{-2(l-k)})}{\beta_1^2 - 1} + \frac{\sigma^2 c_4^2(1 - \beta_2^{-2(l-k)})}{\beta_2^2 - 1} - \frac{2\sigma^2 c_3 c_4 \left(1 - (-\beta_1\beta_2)^{-(l-k)}\right)}{\beta_1\beta_2 + 1}. \tag{64}$$

Hence note that

$$\frac{1}{l}\sum_{k=1}^{l} E(S_{0k})^2 \leq (1 - c_3 - c_4)^2\sigma^2 + \frac{\sigma^2 c_3^2}{\beta_1^2 - 1} + \frac{\sigma^2 c_4^2}{\beta_2^2 - 1} - \frac{2c_3c_4\sigma^2}{\beta_1\beta_2 + 1} + \frac{2c_3c_4\sigma^2\beta_1\beta_2}{l(\beta_1^2\beta_2^2 - 1)} \tag{65}$$

$$\to (1 - c_3 - c_4)^2\sigma^2 + \frac{\sigma^2 c_3^2}{\beta_1^2 - 1} + \frac{\sigma^2 c_4^2}{\beta_2^2 - 1} - \frac{2c_3c_4\sigma^2}{\beta_1\beta_2 + 1} \quad \text{as } l \to \infty. \tag{66}$$

It can now be shown that this limiting expression achieves its minimum value when c_3 and c_4 are chosen as

$$c_3 = \frac{(\beta_1^2 - 1)(\beta_1\beta_2 + 1)}{\beta_1^2\beta_2(\beta_1 + \beta_2)}, \text{ and } c_4 = \frac{(\beta_2^2 - 1)(\beta_1\beta_2 + 1)}{\beta_1\beta_2^2(\beta_1 + \beta_2)}, \tag{67}$$

which results in

$$\lim_{l\to\infty} \frac{1}{l}\sum_{k=1}^{l} E(S_{0k})^2 \leq \frac{\sigma^2}{\beta_1^2\beta_2^2}. \tag{68}$$

Now let us look at the contribution of the quantization noise in the average distortion:

$$\frac{1}{l}\sum_{k=1}^{l} E\left[c_3\beta_1^{k-1}Q_1 + (-1)^k c_4\beta_2^{k-1}Q_2\right]^2 \leq \frac{c_3^2 E(Q_1^2)\beta_1^{2l}}{l(\beta_1^2 - 1)} + \frac{c_4^2 E(Q_2^2)\beta_2^{2l}}{l(\beta_2^2 - 1)} + \frac{2c_3c_4(\beta_1\beta_2)^l\sqrt{E(Q_1^2)E(Q_2^2)}}{l(\beta_1\beta_2 - 1)} \tag{69}$$

$$\to 0 \quad \text{as } l \to \infty, \tag{70}$$

where we have used the fact that as $l \to \infty$, $(\beta_i^{2l}E(Q_i^2))/l \to 0$ for $i = 1, 2$. Looking at the cross term in the average distortion,

$$\frac{2}{l}\sum_{k=1}^{l} \left[c_3\beta_1^{k-1}E(S_{0k}Q_1) + (-1)^k c_4\beta_2^{k-1}E(S_{0k}Q_2)\right] \leq \frac{2\vartheta}{l}\left[\frac{c_3\sqrt{E(Q_1^2)}\beta_1^l}{\beta_1 - 1} + \frac{c_4\sqrt{E(Q_2^2)}\beta_2^l}{\beta_2 - 1}\right] \tag{71}$$

$$\to 0 \quad \text{as } l \to \infty, \tag{72}$$

where we have used the fact that

$$E(S_{0k}^2) \leq (1 - c_3 - c_4)^2 + \frac{\sigma^2 c_3^2}{\beta_1^2 - 1} + \frac{\sigma^2 c_4^2}{\beta_2^2 - 1} \triangleq \vartheta. \tag{73}$$

Hence it follows that in the distortion for Decoder-0, the contribution due to the quantization noises Q_1 and Q_2 also goes to zero as $l \to \infty$, and the contribution due to the cross terms (i.e., terms like $E(S_{0k}Q_i)$ for $i = 1, 2$) also goes to zero as $l \to \infty$. Using these results it follows that

$$\lim_{l \to \infty} \frac{1}{l} \sum_{k=1}^{l} E(X_k - \hat{X}_{0k})^2 \leq \frac{\sigma^2}{\beta_1^2 \beta_2^2}. \tag{74}$$

Hence we have shown that the tuple $(R_1, R_2, \sigma^2 2^{-2(R_1+R_2)}, \sigma^2 2^{-2R_1}, \sigma^2 2^{-2R_2})$ is achievable.

Converse: The converse coding theorem follows immediately by noting the following fact [4, 7]. In the point-to-point case, the optimal rate-distortion function for a memoryless source with additive distortion measure with noiseless feedforward is the same as the Shannon rate-distortion function.

4.2 Multiple description source coding error exponent region for the symmetric case

To obtain the error exponents for the above system, we need a more refined argument. In the following we consider the case of symmetric multiple description source coding where $R_1 = R_2 = R$ and $D_1 = D_2$, and evaluate the optimal multiple description source coding error exponents. Hence let $\beta_1 = \beta_2 = \beta$. The corresponding analysis for the more the general asymmetric case is highly cumbersome and intractable at the time when this paper is written. The key idea is to make the crossterms in the average distortions exactly equal to zero for all the three decoders, only then the large deviation analysis becomes tractable.

Definition 6: For a given rate pair R_1, R_2 and a tuple of target distortion D_i for $i = 0, 1, 2$, a tuple (E_0, E_1, E_2) is said to be an achievable error exponent tuple for multiple description source coding with noiseless feedforward, if $\forall \nu > 0$, sufficiently small, there exists a multiple description source code with noiseless feedforward with parameters $(l, \Theta_1, \Theta_2, \Phi_0, \Phi_1, \Phi_2)$ such that for $i = 1, 2$ and $j = 0, 1, 2$,

$$\frac{1}{l} \log \Theta_i < R_i + \nu, \quad \text{and} \quad -\frac{1}{l} \log P \left[\frac{1}{l} \sum_{k=1}^{l} (X_k - \hat{X}_{jk})^2 > D_j \right] \geq E_j - \nu. \tag{75}$$

The optimal error exponent region $\mathcal{E}_{mdff}$ denotes the convex closure of the set of such achievable error exponent tuples.

Consider the following modified encoder and decoders.

Encoder and decoder: Fix $\beta > 1$, even l, $M = \beta^{l(1+\epsilon')}$, and $\Delta = \beta^{l\epsilon'}$. The encoder is given in the following. Let

$$Y_1' = aX_1 + bX_2 + cX_3 + \frac{X_4}{\beta} + \frac{X_5}{\beta^2} + \frac{X_6}{\beta^3} + \ldots + \frac{X_l}{\beta^{l-3}}, \tag{76}$$

$$Y_2' = -aX_1 + bX_2 - cX_3 + \frac{X_4}{\beta} - \frac{X_5}{\beta^2} + \frac{X_6}{\beta^3} - \ldots, + \frac{X_l}{\beta^{l-3}}, \tag{77}$$

where constants a, b, and c will be determined later. The joint encoder quantizes Y_1' and Y_2' using two uniform scalar quantizers with M levels and bounded between $-\Delta/2$ and $\Delta/2$. Let $\hat{Y}_i'$ denote the quantized version of Y_i' for $i = 1, 2$ with midpoints of the corresponding quantization cells denoting the alphabets of $\hat{Y}_i'$. Then the encoder sends the index of the cells containing Y_i' on the ith channel. The ith decoder for $i = 1, 2$ is given by

$$\hat{X}_{ik} = (-1)^{i-1}\beta\left(\hat{X}_{i,(k-1)} - \frac{\beta^2 - 1}{\beta^2}X_{k-1}\right), \quad \text{for} \quad k = 4, 5, \ldots, l, \text{ and} \tag{78}$$

$$\hat{X}_{i1} = (-1)^{i-1}d\hat{Y}_i', \quad \hat{X}_{i2} = (-1)^{i-1}e\left(\frac{1}{ad}\hat{X}_{i1} - X_1\right), \quad \hat{X}_{i3} = (-1)^{i-1}f\left(\frac{1}{be}\hat{X}_{i2} - X_2\right), \tag{79}$$

where the constants d, e, and f will be determined later. The reconstructions for Decoder-0 is given by

$$\hat{X}_{0k} = \frac{\beta^2 + 1}{2\beta^2}\left[\hat{X}_{1k} + \hat{X}_{2k}\right], \tag{80}$$

for $k = 4, 5, \ldots, l$, and

$$\hat{X}_{01} = \frac{g}{d}[\hat{X}_{11} + \hat{X}_{21}] \quad \hat{X}_{02} = \frac{h}{e}[\hat{X}_{12} + \hat{X}_{22}] \quad \hat{X}_{03} = \frac{i}{f}[\hat{X}_{13} + \hat{X}_{23}], \tag{81}$$

where g, h, i are another set of constants to be determined later. In the sequel we will choose these 9 constants such that the crossterms in the distortions of the three decoders are exactly equal to 0.

Note that in (76,77) we have modified the coefficients of the first three source samples. One might ask why three? It turns out that if we had modified the coefficients of only the first two samples, then there would be 6 unknowns and 7 equations. For the case of three, the number of unknowns is equal to the number of equations, which is 9. For the case of 4, there would be 12 unknowns and 11 equations, and so on. Hence for the case of 4, we have more degrees of freedom in choosing these coefficients.

The following theorem gives the main result of this subsection.

Theorem 4: The above scheme achieves the optimal multiple description source coding error exponent region $\mathcal{E}_{mdff}(R, R, D_0, D_1, D_1)$, which is given by the set of all tuples (E_0, E_1, E_2) such that for $i = 1, 2$,

$$E_i \le \frac{D_1 e^{2R}}{2\sigma^2} - \frac{1}{2} - \frac{1}{2}\log\frac{D_1}{\sigma^2} - R, \tag{82}$$

$$E_0 \le \frac{D_0 e^{4R}}{2\sigma^2} - \frac{1}{2} - \frac{1}{2}\log\frac{D_0}{\sigma^2} - 2R. \tag{83}$$

Proof: First let us consider the average distortion of Decoder-1. Define for $k = 4, 5, \ldots, l$,

$$S'_{1k} = \frac{X_k}{\beta^2} - \frac{\beta^2 - 1}{\beta^2} \sum_{m=k+1}^{l} \frac{X_m}{\beta^{m-k}}, \tag{84}$$

and

$$S'_{11} = (1 - ad)X_1 - dbX_2 - dcX_3 - \sum_{m=4}^{l} \frac{dX_m}{\beta^{m-3}}, \tag{85}$$

$$S'_{12} = (1 - be)X_2 - ecX_3 - \sum_{m=4}^{l} \frac{eX_m}{\beta^{m-3}}, \tag{86}$$

$$S'_{13} = (1 - fc)X_3 - \sum_{m=4}^{l} \frac{fX_m}{\beta^{m-3}}. \tag{87}$$

Now note that

$$\frac{1}{l}\sum_{k=1}^{l} E(X_k - \hat{X}_{1k})^2 = \frac{1}{l}\sum_{k=1}^{l} E[(S'_{1k})^2] + \left[(d^2 + e^2 + f^2) + \sum_{k=4}^{l} \frac{(\beta^2 - 1)^2}{\beta^4} \beta^{2(k-3)}\right] E(Q_1^2) + 2E(Q_1 \xi'_1) \tag{88}$$

where ξ'_1 denotes the crossterm, which can be evaluated as

$$\xi'_1 = -d(1-ad)X_1 + (d^2 b - e(1-be))X_2 + (d^2 c + e^2 c - f(1-fc))X_3 + \left[d^2 + e^2 + f^2 - \frac{\beta^2 - 1}{\beta^2}\right] \sum_{k=4}^{l} \frac{X_k}{\beta^{k-3}}. \tag{89}$$

Hence to make $\xi'_1 = 0$, we need

$$ad = 1, \quad d^2 + e^2 = \frac{e}{b}, \quad d^2 + e^2 + f^2 = \frac{f}{c}, \quad d^2 + e^2 + f^2 = \frac{\beta^2 - 1}{\beta^2}. \tag{90}$$

Similarly, for Decoder-2, by defining for $k = 4, 5, \ldots, l$,

$$S'_{2k} = \frac{X_k}{\beta^2} - \frac{\beta^2 - 1}{\beta^2} \sum_{m=k+1}^{l} (-1)^{m-k} \frac{X_m}{\beta^{m-k}}, \quad \text{and} \tag{91}$$

$$S'_{21} = (1 - ad)X_1 + dbX_2 - dcX_3 - \sum_{m=4}^{l} (-1)^{m-3} \frac{dX_m}{\beta^{m-3}}, \tag{92}$$

$$S'_{22} = (1 - be)X_2 + ecX_3 - \sum_{m=4}^{l} (-1)^{m-4} \frac{eX_m}{\beta^{m-3}}, \quad \text{and} \quad S'_{23} = (1 - fc)X_3 - \sum_{m=4}^{l} (-1)^{m-3} \frac{fX_m}{\beta^{m-3}}, \tag{93}$$

it can be shown that to make the corresponding cross term in the average distortion of Decoder-2 exactly equal to 0, we again need the conditions given in (90).

Next let us look at the average distortion for Decoder-0. Let us define for $k = 4, 5, \ldots, l$,

$$S'_{0k} = \frac{X_k}{\beta^4} - \frac{\beta^4 - 1}{2\beta^4} \sum_{m=k+1}^{l} \frac{X_m}{\beta^{m-k}} (1 + (-1)^{m-k}), \tag{94}$$

and

$$S'_{01} = (1-2ga)X_1 - 2gcX_3 - \sum_{m=4}^{l} \frac{gX_m}{\beta^{m-3}}(1+(-1)^{m-3}), \tag{95}$$

$$S'_{02} = (1-2hb)X_2 - \sum_{m=4}^{l} \frac{hX_m}{\beta^{m-3}}(1+(-1)^{m-4}), \tag{96}$$

$$S'_{03} = (1-2ic)X_3 - \sum_{m=4}^{l} \frac{iX_m}{\beta^{m-3}}(1+(-1)^{m-3}). \tag{97}$$

Using these definitions, we can rewrite the average distortion of Decoder-0 as

$$\frac{1}{l}\sum_{k=1}^{l} E(X_k - \hat{X}_{0k})^2 = \frac{1}{l}\sum_{k=1}^{l} E[(S'_{0k})^2] + \kappa'_0 + 2E(Q_1\xi'_{01}) + 2E(Q_2\xi'_{02}), \tag{98}$$

where

$$\xi'_{01} = -g(1-2ga)X_1 - h(1-2hb)X_2 + (2g^2c - i(1-2ic))X_3 + \left[h^2 - \frac{\beta^4-1}{4\beta^6}\right]\sum_{m=4}^{l}\frac{X_m}{\beta^{m-3}}(1+(-1)^{m-4}) \tag{99}$$

$$+\left[(g^2+i^2) - \frac{\beta^4-1}{4\beta^4}\right]\sum_{m=4}^{l}\frac{X_m}{\beta^{m-3}}(1+(-1)^{m-3}), \tag{100}$$

$$\xi'_{02} = g(1-2ga)X_1 - h(1-2hb)X_2 - (2g^2c - i(1-2ic))X_3 + \left[h^2 - \frac{\beta^4-1}{4\beta^6}\right]\sum_{m=4}^{l}\frac{X_m}{\beta^{m-3}}(1+(-1)^{m-4}) \tag{101}$$

$$-\left[(g^2+i^2) - \frac{\beta^4-1}{4\beta^4}\right]\sum_{m=4}^{l}\frac{X_m}{\beta^{m-3}}(1+(-1)^{m-3}), \tag{102}$$

and κ'_0 denotes the term involving moments of the quantization noises. Note that we have used the following two facts in deriving the above expressions for ξ'_{01} and ξ'_{02}:

$$\frac{\beta^4-1}{2\beta^4}\sum_{m=4}^{l} S'_{0m}\beta^{m-3} = \frac{\beta^4-1}{4\beta^6}\left[\sum_{m=4}^{l}\frac{X_m}{\beta^{m-3}}(1+(-1)^{m-4}) + \sum_{m=5}^{l}\frac{X_m}{\beta^{m-5}}(1+(-1)^{m-5})\right], \tag{103}$$

$$\frac{\beta^4-1}{2\beta^4}\sum_{m=4}^{l}(-1)^{m-4} S'_{0m}\beta^{m-3} = \frac{\beta^4-1}{4\beta^6}\left[\sum_{m=4}^{l}\frac{X_m}{\beta^{m-3}}(1+(-1)^{m-4}) - \sum_{m=5}^{l}\frac{X_m}{\beta^{m-5}}(1+(-1)^{m-5})\right]. \tag{104}$$

Thus to make the cross terms exactly equal to 0, we need

$$2ga = 1, \quad 2hb = 1, \quad g^2 + i^2 = \frac{i}{2c}, \quad g^2 + i^2 = \frac{\beta^4-1}{4\beta^4}, \quad h^2 = \frac{\beta^4-1}{4\beta^6}. \tag{105}$$

Consider the following lemma.

Lemma 4: There exists a solution, where all the 9 variables are real and finite, to the 9 equations as defined in (90) and (105), $\forall \beta > 1$.

Distortion analysis of Decoder-1: Let us consider the rest of the terms in the average distortion of Decoder-1. The condition that $\xi_1' = 0$ can be rewritten as

$$S_{11}' = -\frac{\beta^2 - 1}{d\beta}\left[\sum_{m=4}^{l} \beta^{m-4} S_{1m}'\right] - eS_{12}' - fS_{13}'. \tag{106}$$

Using this relation we have the following lemma:

Lemma 5: $\forall \theta > 1$, we have

$$P\left[\frac{1}{l}\sum_{k=1}^{l}[S_{1k}']^2 > \frac{\theta\sigma^2}{\beta^2}\right] \leq e^{-l[E^*(\theta)]+o(l)}. \tag{107}$$

Using the analysis of the previous section, one can show that $P(\mathbb{E}_i')$ goes to zero doubly exponentially fast as a function of the block-length l, where $\mathbb{E}_i'$ denotes the event that $Y_i' > \Delta/2$. This in turn implies that $\forall \nu > 0$,

$$\lim_{l\to\infty} -\frac{1}{l}\log P\left[\frac{1}{l}\sum_{k=1}^{l}(X_k - \hat{X}_{1k})^2 > D\right] \geq \frac{De^{2(R-v)}}{2\sigma^2} - \frac{1}{2} - \frac{1}{2}\log\frac{D}{\sigma^2} - (R-\nu), \tag{108}$$

which is the desired result. A similar result can be obtained for Decoder-2.

Distortion analysis of Decoder-0: Now we can analyze the rest of the terms in the average distortion of Decoder-0. Note that the condition given by $\xi_{01}' = 0$, and $\xi_{02}' = 0$ as discussed above, can be rewritten as

$$gS_{01}' + hS_{02}' + iS_{03}' = -\frac{\beta^4 - 1}{2\beta^4}\sum_{m=4}^{l} S_{0m}'\beta^{m-3}, \quad \text{and} \tag{109}$$

$$-gS_{01}' + hS_{02}' - iS_{03}' = -\frac{\beta^4 - 1}{2\beta^4}\sum_{m=4}^{l} (-1)^{m-4} S_{0m}'\beta^{m-3}. \tag{110}$$

Using these two relations, we can get the following result:

Lemma 6: $\forall \theta > 1$, we have

$$P\left[\frac{1}{l}\sum_{k=1}^{l}[S_{0k}']^2 > \frac{\theta\sigma^2}{\beta^4}\right] \leq e^{-l[E^*(\theta)]+o(l)}. \tag{111}$$

Now looking at the contribution to the average distortion from the quantization noise, we get

$$\kappa_0' = \frac{1}{l}\left[(g^2 + i^2)(Q_1 - Q_2)^2 + h^2(Q_1 + Q_2)^2 + \sum_{k=4}^{l}\left[\frac{\beta^4 - 1}{2\beta^4}\beta^{k-4}\left(Q_1 + (-1)^{k-4}Q_2\right)\right]^2\right]. \tag{112}$$

We have the following lemma about $P(\kappa_0 > \delta)$, $\forall \delta > 0$, the probability of the event regarding the quantization noise:

Lemma 7: $\forall \delta > 0$ sufficiently small, we have

$$P[\kappa_0' > \delta] \leq s_{01}e^{-s_{02}e^{ls_{03}}}, \tag{113}$$

where

$$s_{01} = \frac{2\sqrt{2}\beta^{-l\epsilon'}}{\sqrt{\pi}}\sqrt{a^2+b^2+c^2+\frac{1}{\beta^2-1}}, \quad s_{02} = \frac{1}{8\left(a^2+b^2+c^2+\frac{1}{\beta^2-1}\right)}, \quad s_{03} = 2\epsilon' \log\beta. \tag{114}$$

Using Lemma 6 and 7, it follows that $\forall \nu > 0$,

$$\lim_{l\to\infty} -\frac{1}{l}\log P\left[\frac{1}{l}\sum_{k=1}^{l}(X_k - \hat{X}_{0k})^2 > D\right] \geq \frac{De^{(4R-2v)}}{2\sigma^2} - \frac{1}{2} - \frac{1}{2}\log\frac{D_0}{\sigma^2} - (2R-\nu), \tag{115}$$

which is the desired result.

□

As in the point-to-point case, here too, the error event associated with each decoder is essentially the union of two error events. In the symmetric case, for Decoder-0, exactly half of the terms that contribute to the distortion due to dimensional collapse vanish because of the alternating signs. Since the contribution of the quantization noise to the average distortion is asymptotically zero, it is as if a single description of rate $2R$ is handed over to Decoder-0 with noiseless feedforward.

5 Conclusions

We have obtained a block coding scheme for Gaussian sources in the point-to-point setting for source coding with noiseless feedforward. This scheme achieves the optimal rate-distortion function and the optimal source coding error exponents. Further, we have formulated the problem of multiple description source coding with noiseless feedforward, and studied the case of Gaussian sources in detail. It is rather surprising that a simple linear processing and uniform scalar quantization can achieve the boundary of the optimal rate-distortion region for the problem of 2-channel multiple description source coding with noiseless feedforward for Gaussian sources with mean squared error criterion. The take-home message is that feedforward does indeed improve the optimal rate-distortion region and the optimal error exponents for memoryless sources with additive distortion measures in multiterminal settings. Several questions are in order. Can this method be extended to the case of more than 2 descriptions? What is the information-theoretic characterization of the optimal rate-distortion region for general discrete memoryless sources for multiple descriptions with noiseless feedforward? Can the proposed method be extended for noisy feedforward?

Acknowledgments: The author gratefully acknowledges the discussions he had on this topic with Professor Prakash Ishwar, Dr. Emin Martinian, Professor David Neuhoff, Professor Kannan Ramchandran, Professor Sekhar Tatikonda and Dr. Vinay Vaishampayan.

Appendix

Proof of Lemma 1: First consider the following expression for some $\psi > 0$,

$$\int_{\psi}^{\infty} (x-\psi)^2 \exp\left(-\frac{x^2}{2}\right) dx = -xe^{-x^2/2}\Big|_{\psi}^{\infty} + \int_{\psi}^{\infty} e^{-x^2/2}dx + \psi e^{-x^2/2}\Big|_{\psi}^{\infty} + \psi^2 \int_{\psi}^{\infty} e^{-x^2/2}dx \quad (116)$$

$$= (\psi^2+1)\int_{\psi}^{\infty} e^{-x^2/2}dx - \psi e^{-\psi^2/2} \quad (117)$$

$$= (\psi^2+1)\sqrt{2\pi}\text{erfc}_*(\psi) - \psi e^{-\psi^2/2} \quad (118)$$

$$\leq \frac{(\psi^2+1)}{\psi}e^{-\psi^2/2} - \psi e^{-\psi^2/2} \quad (119)$$

$$= \frac{1}{\psi}e^{-\psi^2/2}, \quad (120)$$

where we have used the fact [40] that $\text{erfc}_*(\psi) \leq \frac{1}{\psi\sqrt{2\pi}}\exp(-\frac{\psi^2}{2})$. Now using this result, and by noting that when $|Y| > \Delta/2$ it is quantized to either $(\Delta/2 - \Delta/(2M))$ or $(\Delta/(2M) - \Delta/2)$ depending on whether Y is positive or negative respectively, we have

$$P(\mathbb{E})E(Q^2|\mathbb{E}) = 2\int_{\Delta/2}^{\infty} \left[x - \left(\frac{\Delta}{2} - \frac{\Delta}{2M}\right)\right]^2 \frac{1}{\sqrt{2\pi E(Y^2)}} \exp\left[-\frac{x^2}{2E(Y^2)}\right] dx \quad (121)$$

$$\leq 2\int_{\frac{\Delta}{2}(1-\frac{1}{M})}^{\infty} \left[x - \left(\frac{\Delta}{2} - \frac{\Delta}{2M}\right)\right]^2 \frac{1}{\sqrt{2\pi E(Y^2)}} \exp\left[-\frac{x^2}{2E(Y^2)}\right] dx \quad (122)$$

$$\leq \frac{4(E(Y^2))^{3/2}M}{\sqrt{2\pi}\Delta(M-1)} \exp\left[-\frac{\Delta^2(M-1)^2}{8E(Y^2)M^2}\right] \quad (123)$$

$$\leq \frac{8(E(Y^2))^{3/2}}{\sqrt{2\pi}\Delta} \exp\left[-\frac{\Delta^2}{32E(Y^2)}\right]. \quad (124)$$

□

Proof of Lemma 2: Let $Z = \sum_{i=1}^{l}(S_i')^2$. Using the Chebyshev inequality we have

$$P\left[Z > \frac{\theta l \sigma^2}{\beta^2}\right] \leq \exp\left[-\frac{\lambda\theta l\sigma^2}{\beta^2}\right] E[\exp(\lambda Z)], \quad (125)$$

for any $\theta > 1$, $\lambda > 0$. Now let us evaluate the second term on the right hand side. Note that using (36,37), we can obtain the following relation among S_k' for $k = 1, 2, \ldots, l$

$$S_1' + \sum_{k=2}^{l}\sqrt{\beta^2-1}\beta^{k-2}S_k' = 0. \quad (126)$$

Using this, Z can be rewritten as

$$Z = \sum_{i=2}^{l}(S_i')^2 + \left[\sum_{i=2}^{l}\sqrt{\beta^2-1}\beta^{i-2}S_i'\right]^2 = \mathbf{S}^T\left[I + (\beta^2-1)KK^T\right]\mathbf{S}, \quad (127)$$

where ^{T} denotes transposition, $\mathbf{S}$ is the column vector formed by $S'_2, S'_3, \ldots, S'_l$, I is the identity matrix of size $(l-1)$ and K is the column vector given by $K = [1 \;\; \beta \;\; \beta^2 \;\; \ldots \beta^{l-2}]^T$.

Using the relation between the vectors $S'_2, S'_3, \ldots, S'_l$ and $X_2, \ldots, X_l$, it can be noted that

$$\mathbf{S} = \Theta \mathbf{X}, \tag{128}$$

where $\mathbf{X}$ is the column vector formed by $X_2, X_3, \ldots, X_l$, and Θ is an upper triangular square matrix with all the diagonal elements equal to $1/\beta^2$ and its (ij)th element is given by $\Theta_{ij} = (1-\beta^2)/\beta^{(j-i+2)}$ for $1 \le i \le (l-2)$, and $2 \le j \le (l-1)$ and $i < j$.

Now, the second term of the right hand side of (125) can be evaluated as

$$\begin{aligned} E[\exp \lambda Z] &= \frac{1}{(2\pi)^{(l-1)/2}\sigma^{l-1}} \int \exp\left[-\frac{1}{2\sigma^2}\mathbf{X}^T\left[I - 2\lambda\sigma^2\Theta^T\left[I + (\beta^2-1)KK^T\right]\Theta\right]\mathbf{X}\right] d\mathbf{X} && (129) \\ &= \left|I - 2\lambda\sigma^2\Theta^T\Theta - 2\lambda\sigma^2(\beta^2-1)\Theta^T KK^T\Theta\right|^{-\frac{1}{2}}, && (130) \end{aligned}$$

where the only conditions on λ are that $\lambda > 0$ and the matrix $(I - 2\lambda\sigma^2\Theta^T\Theta - 2\lambda\sigma^2(\beta^2-1)\Theta^T KK^T\Theta)$ be positive definite.

Let us first evaluate $K^T\Theta$ by noting that its ith term for $1 \le i \le (l-2)$, is given by

$$(1-\beta^2)\left[\frac{1}{\beta^{i+1}} + \frac{\beta}{\beta^i} + \ldots + \frac{\beta^{i-1}}{\beta^2}\right] + \beta^{i-1} = \beta^{i-1} - \beta^{i-1}\left[1 - \beta^{-2i}\right] = \frac{1}{\beta^{i+1}}. \tag{131}$$

Hence, we have $K^T\Theta = (1/\beta^2)N^T$, where N is the column vector given by $N^T = [1 \;\; 1/\beta \;\; 1/\beta^2 \;\; \ldots \;\; 1/\beta^{l-2}]$.

Further, one can evaluate $\Theta^T\Theta$ by computing the ith diagonal term as

$$\begin{aligned} \frac{1}{\beta^4} + (\beta^2-1)^2\left[\frac{1}{\beta^6} + \frac{1}{\beta^8} + \ldots + \frac{1}{\beta^{2(i+1)}}\right] &= \frac{1}{\beta^4} + \frac{(\beta^2-1)}{\beta^4}\left[1 - \frac{1}{\beta^{2i-2}}\right] && (132) \\ &= \frac{1}{\beta^2} - \frac{(\beta^2-1)}{\beta^{2(i+1)}}, && (133) \end{aligned}$$

and by computing the (ij)th term for $1 \le i, j \le (l-1)$ and $i < j$ as

$$\begin{aligned} -\frac{(\beta^2-1)}{\beta^{j-i+4}} + (\beta^2-1)^2\left[\frac{1}{\beta^{j-i+6}} + \frac{1}{\beta^{j-i+8}} + \ldots + \frac{1}{\beta^{i+j+2}}\right] &= \frac{(1-\beta^2)}{\beta^{j-i+4}} + \frac{(\beta^2-1)}{\beta^{j-i+4}}\left[1 - \beta^{-2(i-1)}\right] && (134) \\ &= \frac{(1-\beta^2)}{\beta^{i+j+2}}. && (135) \end{aligned}$$

Using these results, the matrix $\Theta\Theta^T$ can be rearranged to be expressed as $\Theta\Theta^T = \frac{1}{\beta^2}I - \frac{(\beta^2-1)}{\beta^4}NN^T$. Collecting all the above results regarding $K^T\Theta$ and $\Theta^T\Theta$, we have

$$I - 2\lambda\sigma^2\Theta^T\Theta - 2\lambda\sigma^2(\beta^2-1)\Theta^T KK^T\Theta = \left(1 - \frac{2\lambda\sigma^2}{\beta^2}\right) I. \tag{136}$$

This implies that the only condition on λ is that $0 < \lambda < \frac{\beta^2}{2\sigma^2}$. Thus, the second term of the right hand side of (125) can be expressed as

$$E[\exp(\lambda Z)] = \left|\left(1 - \frac{2\lambda\sigma^2}{\beta^2}\right) I\right|^{-\frac{1}{2}} = \left(1 - \frac{2\lambda\sigma^2}{\beta^2}\right)^{-(l-1)/2}. \tag{137}$$

Now using this we have,

$$P\left[Z > \frac{\theta l \sigma^2}{\beta^2}\right] \leq \left[\exp\left(\frac{-\lambda\theta l\sigma^2}{\beta^2}\right)\right]\left[1 - \frac{2\lambda\sigma^2}{\beta^2}\right]^{-(l-1)/2} \tag{138}$$

$$= e^{-l[\bar{E}(\lambda,\theta)]+o(l)}, \tag{139}$$

where

$$\bar{E}(\lambda,\theta) = \log\left[e^{\frac{\lambda\theta\sigma^2}{\beta^2}}\sqrt{1 - \frac{2\lambda\sigma^2}{\beta^2}}\right]. \tag{140}$$

To get the best upper bound on the probability $P[Z > \frac{\theta l \sigma^2}{\beta^2}]$ among the above class, we maximize $\bar{E}(\lambda,\theta)$ with respect to λ to get:

$$P\left[Z > \frac{\theta l \sigma^2}{\beta^2}\right] \leq e^{-l[E^*(\theta)]+o(l)}, \tag{141}$$

where

$$E^*(\theta) = \left(\frac{(\theta-1)}{2}\right) - \frac{1}{2}\log\theta, \tag{142}$$

and the maximization is achieved when $\lambda = \frac{\beta^2(\theta-1)}{2\sigma^2\theta}$, which is clearly smaller than $\beta^2/(2\sigma^2)$.

$\square$

Proof of Lemma 3: Define $\mathbb{E}'$ as the event that $|Y'| > \Delta/2$. Now for sufficiently large l, we have

$$P\left[\frac{\beta^{2l}}{l}Q^2 > \delta\right] \leq P\left[\frac{\beta^{2l}}{l}Q^2 > \delta \middle| (\mathbb{E}')^c\right] + P[\mathbb{E}'] \tag{143}$$

$$= P[\mathbb{E}'], \tag{144}$$

where we have used the fact that the square of step size of the quantizer satisfies the following relation:

$$|Q^2| \leq \frac{\Delta^2}{M^2} = \frac{1}{\beta^{2l}} < \frac{\delta l}{\beta^{2l}}, \tag{145}$$

for sufficiently large l. Note that

$$E(Y')^2 \leq \frac{\sigma^2(\beta^2+1)}{\beta^2-1} \leq \frac{2\sigma^2\beta^2}{\beta^2-1}. \tag{146}$$

Now let us analyze the probability $P(\mathbb{E}')$:

$$P(\mathbb{E}') = \Pr\left[|Y'| > \frac{\beta^{l\epsilon'}}{2}\right] = 2\mathrm{erfc}_*\left[\frac{\beta^{l\epsilon'}}{2\sqrt{E(Y')^2}}\right] \tag{147}$$

$$\leq \quad 2\mathrm{erfc}_* \left[\frac{\beta^{l\epsilon'}\sqrt{\beta^2-1}}{2\sqrt{2}\sigma\beta} \right] \tag{148}$$

$$\leq \quad s_1 e^{-s_2 e^{l s_3}}, \tag{149}$$

where

$$s_1 = \frac{4\sigma\beta}{\sqrt{\pi}\beta^{l\epsilon'}\sqrt{\beta^2-1}}, \quad s_2 = \frac{\beta^2-1}{16\beta^2\sigma^2} \quad \text{and} \quad s_3 = 2\epsilon' \log\beta, \tag{150}$$

and we have used the following fact [40]:

$$\mathrm{erfc}_*(x) \leq \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right). \tag{151}$$

Note that this bound is meaningful only when $\epsilon' > 0$.

□

Proof of Lemma 4: Since h is already determined, we can reduce the above system of equations to the following system of 5 equations consisting of 5 independent variables c, d, e, f, i:

$$d^2 + e^2 = 2eh, \quad d^2 + e^2 + f^2 = \frac{f}{c}, \quad d^2 + e^2 + f^2 = \frac{\beta^2-1}{\beta^2}, \quad d^2 + 4i^2 = \frac{2i}{c}, \quad d^2 + 4i^2 = \frac{\beta^4-1}{\beta^4}. \tag{152}$$

By repeating this procedure, we arrive at one single nonlinear equation in c as given by

$$\frac{\beta^4-1}{\beta^4} - \frac{(\beta^4-1)^2}{\beta^8}c^2 + \frac{\beta^2(\beta^2-1)}{\beta^2+1}\left[1 - \frac{\beta^2-1}{\beta^2}c^2\right]^2 = \frac{\beta^2-1}{\beta^2} - \frac{(\beta^2-1)^2}{\beta^4}c^2, \tag{153}$$

which can be simplified as a quadratic equation in c^2 as

$$\beta^6(\beta^2-1)^2c^4 - c^2(\beta^2-1)(2\beta^8 + 2\beta^4 + 3\beta^2 + 1) + \beta^4(\beta^6 + \beta^2 + 1) = 0. \tag{154}$$

The other 8 variables can be obtained by solving the above equation for c as

$$h = \sqrt{\frac{\beta^4-1}{4\beta^6}}, \quad b = \frac{\beta^3}{\sqrt{\beta^4-1}}, \quad f = \frac{\beta^2-1}{\beta^2}c, \quad i = \frac{\beta^4-1}{2\beta^4}c, \quad e = \frac{\sqrt{\beta^2-1}}{\beta\sqrt{\beta^2+1}}\left[\beta^2 - (\beta^2-1)c^2\right], \tag{155}$$

$$d = \sqrt{\frac{\beta^4-1}{\beta^4} - \frac{(\beta^4-1)^2}{\beta^8}c^2}, \quad g = \frac{1}{2}\sqrt{\frac{\beta^4-1}{\beta^4} - \frac{(\beta^4-1)^2}{\beta^8}c^2}, \quad a = \frac{1}{\sqrt{\frac{\beta^4-1}{\beta^4} - \frac{(\beta^4-1)^2}{\beta^8}c^2}}. \tag{156}$$

By observing all of this information, we can deduce that the minimum number of constraints we need to put on the solution of the quadratic equation (154) such that all 9 variables are real and finite is 2, and are given by (i) $c^2 \geq 0$ and (ii) $\frac{\beta^4}{\beta^4-1} \geq c^2$. In the following we will show that both of these constraints are met.

First note that the coefficient of c^2 in (154), given by the polynomial $-(\beta^2-1)(2\beta^8+2\beta^4+3\beta^2+1) < 0$ for all $\beta > 1$. Next, for c^2, the solution of (154), to be real for all $\beta > 1$, we need to satisfy the following condition:

$$\left[2\beta^8 + 2\beta^4 + 3\beta^2 + 1\right]^2 > 4\beta^{10}(\beta^6 + \beta^2 + 1), \tag{157}$$

which is true if and only if $\forall\beta > 1$,

$$4\beta^{12} + 8\beta^{10} + 8\beta^8 + 12\beta^6 + 13\beta^4 + 6\beta^2 + 1 > 0, \tag{158}$$

which is true. This is the required condition for the solution c^2 to be real. Further, since $(\beta^2 - 1) > 0$, and $\beta^6 + \beta^2 + 1 > 0$ for all $\beta > 1$, and using the above argument, we deduce that c^2, the solution of (154), is always positive for $\beta > 1$. Thus there is a real c which satisfies (154), i.e., (154) has two positive roots and two negative roots. In other words,

$$c^2 = \frac{(2\beta^8 + 2\beta^4 + 3\beta^2 + 1) \pm \sqrt{(2\beta^8 + 2\beta^4 + 3\beta^2 + 1)^2 - 4\beta^{10}(\beta^6 + \beta^2 + 1)}}{2\beta^6(\beta^2 - 1)}. \tag{159}$$

Next we need to show that $1 > ((\beta^4 - 1)c^2)/\beta^4$ for one of the above two values of c^2. In the following we choose the smaller of the two values of c^2. The required condition on c^2 can be rewritten as

$$\sqrt{(2\beta^8 + 2\beta^4 + 3\beta^2 + 1)^2 - 4\beta^{10}(\beta^6 + \beta^2 + 1)} > (2\beta^8 + 2\beta^4 + 3\beta^2 + 1) - \frac{2\beta^{10}}{\beta^2 + 1}. \tag{160}$$

First note that the right hand side of the above equation satisfies

$$\begin{aligned} (2\beta^8 + 2\beta^4 + 3\beta^2 + 1) - \frac{2\beta^{10}}{\beta^2 + 1} &= \frac{2\beta^8 + 2\beta^6 + 5\beta^4 + 4\beta^2 + 1}{\beta^2 + 1} && (161) \\ &> 0, && (162) \end{aligned}$$

$\forall\beta > 1$. Using this, and the fact that the term inside the square root is positive for $\beta > 1$, the required condition is implied if $\forall\beta > 1$

$$(2\beta^8 + 2\beta^4 + 3\beta^2 + 1)^2 - 4\beta^{10}(\beta^6 + \beta^2 + 1) > \left[(2\beta^8 + 2\beta^4 + 3\beta^2 + 1) - \frac{2\beta^{10}}{\beta^2 + 1}\right]^2. \tag{163}$$

This equation can be simplified as

$$2\beta^4 + \beta^2 > 0, \tag{164}$$

which is true if and only if for all $\beta > 1$.

In summary there exists a finite c that satisfies (154) such that $c^2 > 0$ and $1 > ((\beta^4 - 1)c^2)/\beta^4$, which implies that d is real. This implies that the rest of the constants that satisfy the 9 equations are all real and finite.

$\square$

Proof of Lemma 5: By defining

$$A = \left[e,\ f,\ \frac{\beta^2-1}{\beta},\ (\beta^2-1),\ \beta(\beta^2-1),\ \ldots,(\beta^2-1)\beta^{l-5}\right], \tag{165}$$

we note that

$$S'_{11} = -\frac{1}{d}A^T\mathbf{S}'_1, \tag{166}$$

where $\mathbf{S}'_1 = [S'_{12}\ S'_{13}\ \ldots S'_{1l}]$. We can write $\mathbf{S}'_1$ as a function of the source samples as $\mathbf{S}'_1 = \hat{\Theta}\mathbf{X}$, where $\mathbf{X} = [X_2, X_3, \ldots, X_l]$ and $\hat{\Theta}$ is an upper triangular matrix with (ij)th term given by $\hat{\Theta}_{ij} = (1-\beta^2)/\beta^{j-i+2}$ for $j > i > 2$, the (ii)th term is given by $\hat{\Theta}_{ii} = 1/\beta^2$ for $i > 2$. The first two rows of $\hat{\Theta}$ are respectively given by $[(1-be), -ec, -e\hat{N}^T/\beta]$ and $[0, (1-fc), -f\hat{N}^T/\beta]$, where $\hat{N}^T = [1\ 1/\beta\ 1/\beta^2 \ldots 1/\beta^{l-4}]$. Hence we have

$$\sum_{k=1}^{l}[S'_{1k}]^2 = \mathbf{X}^T\left[\hat{\Theta}^T\hat{\Theta} + \frac{1}{d^2}\hat{\Theta}^T AA^T\hat{\Theta}\right]\mathbf{X}. \tag{167}$$

Using arguments similar to Lemma 2, it can be shown that

$$\hat{\Theta}^T\hat{\Theta} = \begin{bmatrix} (1-be)^2 & -ec(1-be) & -\frac{e(1-be)}{\beta}\hat{N}^T \\ -ec(1-be) & e^2c^2 + (1-fc)^2 & \frac{(e^2c-f(1-fc))}{\beta}\hat{N}^T \\ -\frac{e(1-be)}{\beta}\hat{N} & \frac{(e^2c-f(1-fc))}{\beta}\hat{N} & \frac{1}{\beta^2}I + \left[\frac{e^2+f^2}{\beta^2} - \frac{\beta^2-1}{\beta^4}\right]\hat{N}\hat{N}^T \end{bmatrix}, \text{ and} \tag{168}$$

$\hat{\Theta}^T AA^T\hat{\Theta}$ is given by

$$\begin{bmatrix} (1-be)^2e^2 & (1-fc)(1-be)fe - e^3c(1-be) & \left(\frac{\beta^2-1}{\beta^3} - \frac{e^2+f^2}{\beta}\right)(1-be)e\hat{N}^T \\ (1-fc)(1-be)fe - e^3c(1-be) & (1-fc)^2f^2 + e^4c^2 - 2e^2cf(1-fc) & \left(\frac{\beta^2-1}{\beta^3} - \frac{e^2+f^2}{\beta}\right)(f(1-fc) - e^2c)\hat{N}^T \\ \left(\frac{\beta^2-1}{\beta^3} - \frac{e^2+f^2}{\beta}\right)(1-be)e\hat{N} & \left(\frac{\beta^2-1}{\beta^3} - \frac{e^2+f^2}{\beta}\right)(f(1-fc) - e^2c)\hat{N} & \left[\frac{\beta^2-1}{\beta^3} - \frac{e^2+f^2}{\beta}\right]^2\hat{N}\hat{N}^T \end{bmatrix}. \tag{169}$$

Hence it can be shown that $\hat{\Theta}^T\hat{\Theta} + \frac{1}{d^2}\hat{\Theta}^T AA^T\hat{\Theta}$ is a diagonal matrix with the first diagonal element given by $(1-be)$, the second diagonal element given by $(1-fc)$ and the rest of the elements of the diagonal are equal to $1/\beta^2$. Now following Chebyshev inequality, $\forall\theta > 1$, and $\lambda > 0$, we have

$$P\left[\frac{1}{l}\sum_{k=1}^{l}[S'_{1k}]^2 > \frac{\theta\sigma^2}{\beta^2}\right] \leq e^{-\frac{\lambda\theta l\sigma^2}{\beta^2}}\left[1 - \frac{2\lambda\sigma^2}{\beta^2}\right]^{-(l-3)/2}\left[(1-2\lambda\sigma^2(1-be))(1-2\lambda\sigma^2(1-fc))\right]^{-1/2} \tag{170}$$

$$\leq e^{l[E^*(\theta)]+o(l)}, \tag{171}$$

by choosing $\lambda = \frac{\beta^2(\theta-1)}{2\sigma^2\theta}$ as in Lemma 2. But, in the above set of inequalities, we have implicitly assumed that the value of $\lambda > 0$ is such that the following condition is met:

$$I - 2\lambda\sigma^2\left(\hat{\Theta}^T\hat{\Theta} + \frac{1}{d^2}\hat{\Theta}^T AA^T\hat{\Theta}\right) > 0. \tag{172}$$

This implies that λ has to be chosen such that the following constraints are met

$$\lambda < \frac{\beta^2}{2\sigma^2}, \quad \lambda < \frac{1}{2\sigma^2(1-be)}, \quad \lambda < \frac{1}{2\sigma^2(1-fc)}. \tag{173}$$

Since we do not want to sacrifice our freedom in choosing θ, we need to prove that $\beta^2 < 1/(1-be)$, and $\beta^2 < 1/(1-fc)$. In the following we do just that.

First note that since $\beta > 1$, we have $\frac{\beta^4}{\beta^4-1} < \frac{\beta^2}{\beta^2-1}$. Using this relation, and the values of b and e from Lemma 4, we deduce that $be > 0$. Next observe that $\forall\beta > 1$,

$$\begin{aligned} 1-be &= 1-\frac{\beta^4}{\beta^2+1}+\frac{c^2\beta^2(\beta^2-1)}{(\beta^2+1} && (174)\\ &= \frac{1}{\beta^2+1}\left(\beta^2(c^2-1)(\beta^2-1)+1\right) && (175)\\ &> 0, && (176) \end{aligned}$$

which requires $c^2 > 1\ \forall\beta > 1$. Substituting the value of c^2 from (159), this condition is equivalent to $2\beta^2+1 > 0$ which is true $\forall\beta > 1$.

Thus the required condition is satisfied if and only if $\beta^2(1-be) < 1$. By substituting the values of b and e in terms of β and c^2, it can be shown that we require the following condition

$$c^2 < \frac{\beta^6-\beta^4+1}{\beta^4(\beta^2-1)}. \tag{177}$$

By substituting the value of c^2 as obtained in (159) in the above equation, it can be shown that the required condition is equivalent to the condition $\beta^4+3\beta^2+1 > 0$, which is true $\forall\beta > 1$. Hence we have shown that $\beta^2 < 1/(1-be)$.

Now the other required condition is $\beta^2 < 1/(1-fc)$. By noting that $fc = \frac{c^2(\beta^2-1)}{\beta^2}$, and $c^2 < \frac{\beta^2}{\beta^2-1}$ (as shown in the above paragraph), we can deduce that $0 < (1-fc) < 1$. Substituting the value of f as obtained from Lemma 4, it can be shown that this is equivalent to the condition that $c^2 > 1$, which is true $\forall\beta > 1$ as shown above. In essence, the only condition on λ is that $0 < \lambda < \frac{\beta^2}{2\sigma^2}$.

□

Proof of Lemma 6: Solving the two equations given by (109,110), and using (103,104) we get,

$$\begin{aligned} S'_{01} &= -\frac{i}{g}S'_{03}-\frac{\beta^4-1}{4g\beta^4}\sum_{m=5}^{l}S'_{0m}\beta^{m-3}(1+(-1)^{m-5}) && (178)\\ &= -E^T\mathbf{X}_0, && (179) \end{aligned}$$

and similarly

$$S'_{02} = -\frac{\beta^4-1}{2h\beta^7}\Psi_1^T\mathbf{X}_1, \tag{180}$$

where $\mathbf{X}_0^T = [X_3, X_5, \ldots, X_{l-1}]$, $\mathbf{X}_1^T = [X_4, X_6, \ldots, X_l]$, and

$$E^T = \left[\frac{-i(1-2ic)}{g}, \left(\frac{2i^2}{\beta^2 g} - \frac{\beta^4-1}{2g\beta^6}\right)\Psi_0^T\right], \tag{181}$$

where $\Psi_0^T = [1\ 1/\beta^2\ 1/\beta^4\ \ldots\ 1/\beta^{l-6}]$, and $\Psi_1^T = [1\ 1/\beta^2\ 1/\beta^4\ \ldots\ 1/\beta^{l-4}]$. Next, we can express $\mathbf{S}'_{00} = G_0\mathbf{X}_0$ and $\mathbf{S}'_{01} = G_1\mathbf{X}_1$, where $(\mathbf{S}'_{00})^T = [S'_{03}\ S'_{05}\ \ldots\ S'_{0,l-1}]$, $(\mathbf{S}'_{01})^T = [S'_{04}\ S'_{06}\ \ldots\ S'_{0l}]$, and G_1 is an upper triangular matrix with diagonal entries given by $1/\beta^4$ and (ij)th entry is given by $-(\beta^4-1)/\beta^{2(j-i)+4}$ for $j > i$, and G_1 is also an upper triangular matrix with the first diagonal entry given by $(1-2ic)$, and the rest of the diagonal entries given by $1/\beta^4$, and (ij)th entry is given by $-(\beta^4-1)/\beta^{2(j-i)+4}$ for $j > i$ and $i \geq 2$, and $(1j)$th entry is given by $-2i/\beta^{2j-2}$ for $j \geq 2$. Hence

$$\sum_{k=1}^{l}[S'_{0k}]^2 = \mathbf{X}_0^T(EE^T + G_0G_0^T)\mathbf{X}_0 + \mathbf{X}_1^T\left[\frac{(\beta^4-1)^2}{4h^2\beta^{14}}\Psi_1\Psi_1^T + G_1G_1^T\right]\mathbf{X}_1 \tag{182}$$

Now, let us evaluate the above summation. First note that

$$G_1G_1^T = \frac{1}{\beta^4}I - \frac{\beta^4-1}{\beta^8}\Psi_1\Psi_1^T. \tag{183}$$

Hence

$$\left[\frac{(\beta^4-1)^2}{4h^2\beta^{14}}\Psi_1\Psi_1^T + G_1G_1^T\right] = \frac{1}{\beta^4}I. \tag{184}$$

Now let us look at the second part of the right hand side of (182). By noting that

$$EE^T = \begin{bmatrix} \frac{i^2(1-2ic)^2}{g^2} & \frac{i(1-2ic)}{g}\left[\frac{\beta^4-1}{2\beta^6 g} - \frac{2i^2}{\beta^2 g}\right]\Psi_0^T \\ \frac{i(1-2ic)}{g}\left[\frac{\beta^4-1}{2\beta^6 g} - \frac{2i^2}{\beta^2 g}\right]\Psi_0 & \left[\frac{\beta^4-1}{2\beta^6 g} - \frac{2i^2}{\beta^2 g}\right]^2\Psi_0\Psi_0^T \end{bmatrix} \tag{185}$$

and

$$G_0G_0^T = \begin{bmatrix} (1-2ic)^2 & \frac{-2i(1-2ic)}{\beta^2}\Psi_0^T \\ \frac{-2i(1-2ic)}{\beta^2}\Psi_0 & \frac{1}{\beta^4}I + \left[\frac{4i^2}{\beta^4} - \frac{\beta^4-1}{\beta^8}\right]\Psi_0\Psi_0^T \end{bmatrix}, \tag{186}$$

and using the properties of g, i and c, we get $EE^T + G_0G_0^T$ equal to a diagonal matrix with the first diagonal entry given by $(1-2ic)$ and the rest given by $1/\beta^4$. Let $\bar{\Theta}$ denote the diagonal matrix of size $(l-1)$ with the first diagonal entry given by $(1-2ic)$ and the rest given by $1/\beta^4$. Now we can evaluate the required probability using Chebyshev inequality as follows: $\forall\lambda > 0$, we have

$$\begin{aligned}
P\left[\frac{1}{l}\sum_{k=1}^{l}[S'_{0k}]^2 > \frac{\theta\sigma^2}{\beta^4}\right] &= P\left[\sum_{k=1}^{l}[S'_{0k}]^2 > \frac{\theta l\sigma^2}{\beta^4}\right] && (187)\\
&\leq \exp\left[-\frac{\lambda\theta l\sigma^2}{\beta^4}\right]E\left[e^{\lambda\sum_{k=1}^{l}[S'_{0k}]^2}\right] && (188)\\
&= \exp\left[-\frac{\lambda\theta l\sigma^2}{\beta^4}\right]\left[1-2\lambda\sigma^2(1-2ic)\right]^{-1/2}\left[1-\frac{2\lambda\sigma^2}{\beta^4}\right]^{-(l-2)/2} && (189)\\
&= e^{-l\left(\log\left[\sqrt{1-\frac{2\lambda\sigma^2}{\beta^4}}e^{\frac{\lambda\theta\sigma^2}{\beta^4}}\right]\right)+o(l)} && (190)\\
&\leq e^{-l[E^*(\theta)]+o(l)}, && (191)
\end{aligned}$$

by choosing $\lambda = \frac{\beta^4(\theta-1)}{2\sigma^2\theta}$. But as in Lemma 5, we have assumed that the following condition is met

$$1 - 2\sigma^2\lambda\bar{\Theta} > 0, \tag{192}$$

which is equivalent to the following conditions:

$$\lambda < \frac{\beta^4}{2\sigma^2}, \quad \lambda < \frac{1}{2\sigma^2(1-2ic)}. \tag{193}$$

As in Lemma 5, to avoid the loss of freedom in the choices for θ, we prove that for the values of i and c obtained in Lemma 4, we have $\beta^4 < 1/(1-2ic)$. Using arguments similar to those used in the proof of Lemma 5, for the value of c^2 obtained from (159), it can be shown that $0 < (1-2ic) < 1$. Hence we need to prove that $\beta^4(1-2ic) < 1$. This is equivalent to $c^2 > 1$. This condition is satisfied $\forall \beta > 1$ as proved in Lemma 5. In essence the only condition that λ needs to satisfy is $0 < \lambda < \frac{\beta^4}{2\sigma^2}$.

□

Proof of Lemma 7: Consider the following set of inequalities:

$$\begin{aligned} \kappa_0' &\leq \frac{1}{l}\left[\frac{1}{4}(Q_1 - Q_2)^2 + \frac{1}{4}(Q_1+Q_2)^2 + \frac{\beta^{2l}}{4l(\beta^2-1)}[Q_1^2+Q_2^2] + \frac{\beta^{2l}}{2l}|Q_1||Q_2|\right] && (194)\\ &\leq \left[\frac{1}{l}(Q_1^2+Q_2^2) + \frac{1}{(\beta^2-1)}\left[\frac{\beta^{2l}}{l}Q_1^2 + \frac{\beta^{2l}}{l}Q_2^2\right] + \sqrt{\frac{\beta^{2l}}{l}Q_1^2}\sqrt{\frac{\beta^{2l}}{l}Q_2^2}\right]. && (195) \end{aligned}$$

Now let us evaluate the required probability as follows: $\forall \delta > 0$, we have

$$\begin{aligned} P[\kappa_0' > \delta] &\leq P[\kappa_0' > \delta|(\mathbb{E}_1')^c \cap (\mathbb{E}_2')^c] + P(\mathbb{E}_1' \cup \mathbb{E}_2') && (196)\\ &\leq P\left[\left(\frac{1}{l}(Q_1^2+Q_2^2) + \frac{\beta^{2l}}{l(\beta^2-1)}\left[Q_1^2+Q_2^2\right] + \sqrt{\frac{\beta^{4l}}{l^2}Q_1^2Q_2^2}\right) > \delta \,\middle|\, (\mathbb{E}_1')^c \cap (\mathbb{E}_2')^c\right] + 2P(\mathbb{E}_1'). && (197) \end{aligned}$$

Now note that the two quantizers have been chosen such that given the event $(\mathbb{E}_1')^c \cap (\mathbb{E}_2')^c$, we have that $|Q_i| \leq \Delta/M = \beta^{-l}$, for $i = 1, 2$, which implies that $|\beta^l Q_i| < 1$. Hence $\forall \tau > 0$, $\exists l_0$ such that $\forall l > l_0$, the following are true

$$\frac{1}{\min\{(\beta^2-1),1\}}\frac{\beta^{2l}Q_1^2}{l} < \tau, \quad \text{and} \quad \frac{1}{\min\{(\beta^2-1),1\}}\frac{\beta^{2l}Q_2^2}{l} < \tau. \tag{198}$$

Choose τ such that $5\tau = \delta$, hence the first term in (197) is exactly zero. The second term goes to zero doubly exponentially fast as a function of block-length l using Lemma 3. By noting that

$$E(Y_{01}')^2 \leq a^2 + b^2 + c^2 + \frac{1}{\beta^2-1}, \tag{199}$$

the desired result follows directly.

□

References

[1] J. M. Kahn, R. H. Katz, and K. S. J. Pister, "Mobile networking for smart dust," in *ACM/IEEE Intl. Conf. on Mobile Computing and Networking, Seattle, WA*, August 1999.

[2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 1–10, January 1976.

[3] S. S. Pradhan, "Source coding with feedforward," in *CSPL Technical Report No. TR 345, Univ. of Michigan*, December 2003.

[4] T. Weissman and N. Merhav, "On competitive prediction and its relation to rate-distortion theory," *IEEE Transactions on Information theory*, vol. IT-49, pp. 3185–3194, December 2003.

[5] S. S. Pradhan, "Source coding with feedforward: Gaussian sources," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT), Chicago, IL*, June/July 2004.

[6] E. Martinian and G. W. Wornell, "Source Coding with Fixed Lag Side Information," *Proc. of the 42nd Annual Allerton Conference (Monticello, IL)*, 2004.

[7] R. Venkataramanan and S. S. Pradhan, "Source coding with feedforward," in *Proc. IEEE Inform. Theory Workshop(ITW), San Antonio, TX*, October 2004.

[8] J. Massey, "Causality, Feedback and Directed Information," *Proceedings of the 1990 Symposium on Information Theory and its Applications (ISITA-90)*, pp. 303–305, 1990.

[9] G. Kramer, *Directed Information for channels with Feedback.* PhD thesis, Swiss Federal Institute of Technology, Zurich, 1998.

[10] S. Tatikonda and S. Mitter, "The Capacity of Channels with Feedback- Part 1: The General Case," submitted to *IEEE Transactions on Information Theory*, October 2001.

[11] T. M. Cover and J. A. Thomas, *Elements of information theory.* New York: Wiley, 1991.

[12] S. I. Krich, "Coding for a Delay-Dependent Fidelity Criterion," *IEEE Transactions on Information Theory*, pp. 77–85, January 1974.

[13] D. Neuhoff and R. Gilbert, "Causal source codes," *IEEE Trans. Inform. Theory*, pp. 701–713, Sept. 1982.

[14] D. Teneketzis, "Optimal real-time encoding-decoding of Markov sources in noisy environments," *Proc. Math. Theory of Networks and Sys. (MTNS), Leuwen, Belgium*, 2004.

[15] J. Geweke, "Measurements of linear dependence and feedback between multiple time series," *J. Amer. Stat. Assoc.*, vol. 77, pp. 304–313, 1982.

[16] J. Rissanen and M. Wax, "Measures of mutual and causal dependence between two time series," *IEEE Transactions on Information Theory*, vol. 33, pp. 598–561, July 1987.

[17] P. E. Caines and C. W. Chan, "Feedback between stationary processes," *IEEE. Trans. Automatic control*, vol. 20, pp. 498–508, 1975.

[18] A. El Gamal and T. M. Cover, "Achievable Rates for Multiple Descriptions," *IEEE Transactions on Information Theory*, vol. 28, pp. 851–857, November 1982.

[19] L. Ozarow, "On a Source Coding Problem with Two Channels and Three Receivers," *Bell System Technical Journal*, vol. 59, pp. 1909–1921, December 1980.

[20] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. on Inform. Theory*, vol. IT-31, pp. 721–726, November 1985.

[21] Z. Zhang and T. Berger, "New Results in Binary Multiple Descriptions," *IEEE Transactions on Information Theory*, vol. 33, pp. 502–521, July 1987.

[22] V. A. Vaishampayan, "Design of Multiple Description Scalar Quantizers," *IEEE Transactions on Information Theory*, vol. 3, pp. 821–834, May 1993.

[23] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple Description Coding for Many Channels," *IEEE Trans. on Inform. Theory*, pp. 2106–2114, September 2003.

[24] S. S. Pradhan, R. Puri, and K. Ramchandran, "n-channel Symmetric Multiple Descriptions- Part I: (n,k) Source-channel Erasure Codes," *IEEE Transactions on Inform. Theory*, vol. 50, pp. 47–61, January 2004.

[25] R. Puri, S. S. Pradhan, and K. Ramchandran, "n-channel Symmetric Multiple Descriptions- Part II: An Achievable Rate-distortion Region," *to appear in IEEE Transactions on Inform. Theory*, 2005.

[26] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal processing magazine*, vol. 18, pp. 74–93, September 2001.

[27] S. Diggavi and V. Vaishampayan, "Multiple description source coding with side information," in *Proc. IEEE Inform. Theory Workshop(ITW), San Antonio, TX*, October 2004.

[28] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete memoryless sources.* Academic Press, New York, 1981.

[29] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1629–1638, June 2002.

[30] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. on Information Theory*, vol. 49, pp. 1181–1203, May 2003.

[31] R. J. Barron, B. Chen, and G. W. Wornell, "The Duality between Information Embedding and Source Coding with Side Information and Some Applications," *IEEE Transactions on Information Theory*, vol. 49, pp. 1159–1180, May 2003.

[32] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback- part II: Band-limited signals," *IEEE Trans. on Information Theory*, vol. IT-12, pp. 183–189, April 1966.

[33] A. D. Wyner, "On the Schalkwijk-Kailath coding scheme with a peak energy constraint," *IEEE Trans. Inform. Theory*, vol. 14, pp. 129–135, 1968.

[34] L. H. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Transactions on Information Theory*, vol. IT-30, pp. 623–629, July 1984.

[35] G. Kramer, "Feedback strategies for white Gaussian interference networks," *IEEE Transactions on Information Theory*, vol. 48, pp. 1423–1438, June 2002.

[36] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. on Information Theory*, vol. IT-20, pp. 197–199, March 1974.

[37] R. Blahut, "Hypothesis testing and information theory," *IEEE Trans. on Information Theory*, vol. IT-20, pp. 405–417, July 1974.

[38] S. Ihara and M. Kubo, "Error exponent for coding of memoryless Gaussian sources with a fidelity criterion," *IEICE Trans. Fundamentals*, vol. 10, pp. 1891–1897, October 2000.

[39] T. Weissman and N. Merhav, "Tradeoffs between the excess-code-length exponent and the excess-distortion exponent in lossy source coding," *IEEE Transactions on Information Theory*, vol. 48, pp. 396–415, February 2002.

[40] H. L. Van Trees, *Detection estimation and modulation theory.* Wiley: New York, 1968.