Dithered Lattice-Based Quantizers for Multiple Descriptions *

Yael Frank-Dayan
Dept. of EE, Tel-Aviv Univ.
yaelfd@eng.tau.ac.il

Ram Zamir
Dept. of EE, Tel-Aviv Univ.
zamir@eng.tau.ac.il

to appear in the IEEE tr. Info. Theory, 2001-2002

Abstract

Multiple description source coding is aimed at achieving graceful degradation in reconstruction with respect to losing portions of the code, with the cost of some redundancy. We examine multiple description schemes which use entropy coded dithered lattice quantizers (ECDQ). We propose two techniques, one based on successive refinement (SR), and the other a dithered and periodic version of the multiple description scalar quantizer (MDSQ) with distributed cells proposed by Vaishampayan. Similarly to the single description case, both techniques are universal in nature, and are equivalent to additive noise channels. This allows to derive analytical expressions for the rate-distortion performance for general sources, and to compare them to the optimal rate regions at both high and low resolutions. Among other results, we establish that while the dithered MDSQ scheme loses only the space filling loss of the scalar lattice at any resolution, the SR based scheme loses an additional 0.5 bit at any lattice dimension. Possible improvements, such as "refinement time sharing" and "dependent dithering", are discussed.

Key Words: multiple descriptions, successive refinement, dithered quantization, lattice quantizers, universal quantization.

*This research was supported in part by the Israel Science Foundation administered by the Israel Academy of Science. This work was presented in part at the Data Compression Conference, Snowbird, Utah, March 2000.

1 Introduction

The multiple description (MD) problem refers to a scenario in which only a part of the source code produced by the transmitter arrives at the receiver. Under these conditions, it is advantageous to break the code into several "descriptions" of the source sample vector. A good multiple description code is one that can use any subset of these descriptions to reconstruct the source sample, and achieve a smaller distortion as the number of descriptions that arrive at the receiver grows.

The formulation of the multiple description problem in the information theoretic literature, [22], is usually confined to two descriptions, a memoryless source, and a distortion measure(s). The source is encoded into two descriptions which are sent over two lossless channels of rates R_1 and R_2. There are three receivers, one for each channel and a third "central" receiver that receives both channels. The "side" receivers reconstruct the source with distortions d_1 and d_2, whereas the "central" receiver reconstructs the source with a smaller distortion d_0. A theoretic solution to the MD problem, based on Shannon's random coding approach, was published in [6]. For specific cases, the converse proof was found in [1], [15], [13] and [23]; yet, the complete solution for the MD problem is still unknown.

The practical interest in this problem generated quantization schemes for compression with multiple descriptions. Early solutions were based on the natural correlation in the signal that remains even after compression and allows interpolation of missing segments; see [14] and [12] on speech compression, and [11] on image compression. The more recent theory of quantized over-complete expansions gives a general framework for these ideas, and shows how to control the amount of added redundancy [9]. However, these schemes are efficient only for moderate ratios of $\frac{d_1}{d_0}$ and usually prioritize the "side" distortions. Correlating transforms achieve a wide range of side-to-central distortion ratios, without using redundant sampling or expansion [8]. However, they require the source to have memory or varying variance, so they are not applicable in the case of a memoryless source.

As opposed to the linear interpolation / transform approach, the multiple description scalar quantizer (MDSQ) of Vaishampayan [19] uses a non-linear mapping to relate the side and central reconstructions. The "side" descriptions are produced by an index assignment function that operates on the "central" scalar quantizer. The Lagrangian formulation of the optimization problem takes into account the "central" and "side" distortions, and adapts the encoder to the source's distribution. This scheme is limited to the symmetric case ($d_1 = d_2,\ R_1 = R_2$) and to scalar quantization, and it covers only a discrete set of d_1/d_0 ratios starting from 6dB (staggered quantizers). An asymptotic analysis in high resolution conditions for this scheme was performed in [20] and, more importantly, the results were compared to the optimal solution of Ozarow. Many variations on this scheme were proposed in recent years; see, e.g., [16, 2, 21].

A special case of the MD problem, where d_1 (or d_2) is set to the maximum, is successive refinement (SR) [5]. This problem has a complete information theoretic solution as well as practical ones. A sub-optimum method for multiple descriptions based on SR was suggested in [27]. Roughly, using the successive refinement approach to solve a multiple description problem, one would start out with coarse descriptions and derive the "central" description as a refinement of the "side" descriptions.

In this paper we use Entropy Coded Dithered Quantizers (ECDQ), with subtractive dither, [25], to examine and compare two encoding schemes for the multiple description problem: one based on the approach of successive refinement, and the other on the MDSQ. Our goal is to study these schemes, to analyze them, and, as a result, to reveal significant aspects of the MD problem. Through the use of the ECDQ approach, we were also able to generalize the quantization portion of the schemes so that they would be universal, i.e., source independent. The statistics of the source are captured mainly by the entropy coding portion of the ECDQ, which can be implemented by some universal lossless compression scheme (e.g., a Lempel-Ziv-based algorithm). At low resolution quantization we incorporate post-filters into the ECDQ; tuning these filters requires knowledge of the second order statistics of the source.

The first scheme we introduce, the two stage dithered quantizer (TSDQ), suggests an SR based solution. The encoder operates in two stages: the first stage produces two coarse encodings of the source, and the second step produces a refinement to the reconstruction that is obtained from the two coarse encodings. Each description contains one coarse codeword and a part of the refinement code. The main building blocks of the encoder are three ECDQs. The advantage of this scheme is its flexibility; however, it suffers an inherent loss in comparison with the optimum.

The second scheme we introduce combines a limiting periodic case of the MDSQ of Vaishampayan [19] with the ECDQ method. Our main contribution here is the universal structure of the quantizer, the incorporation of post-filters at low resolution, and the exact analysis (not asymptotic) of the performance for any resolution.

A basic property of the single description ECDQ is its rate-distortion equivalence with an additive noise channel (see properties (1.) and (2.) in Section 2.1 below). Applying the ECDQ to the multiple description schemes, we found that both the TSDQ and the dithered MDSQ schemes become equivalent to sets of additive noise channels. Using this equivalence, we compare the performance of the proposed encoders to the optimum solutions for the quadratic memoryless Gaussian [15] and the non Gaussian [23] cases.

The total rate ($R_1 + R_2$) achieved by the entropy coded dithered periodic MDSQ (PMDSQ) exceeds the theoretical bounds by $\log(\frac{2\pi e}{12}) \simeq 0.5$ bit per sample (corresponding to 1.5dB increase in distortion per each description) at high resolution for any "smooth" source. For a general source at any resolution, the excess rate is at most $\log(2\pi e/12) + 2D(X; X^*)$, where $D(X; X^*)$ denotes the divergence of the source from Gaussianity.

The two stage *scalar* dithered quantizer, while being much simpler in design, exceeds the bound by at most $\frac{3}{2}\log(2\pi e/12) + 0.5$ bit per sample at high resolution conditions for any "smooth" source; for a Gaussian source and balanced descriptions this bound on the redundancy holds at all resolutions. Even if the TSDQ uses a multiple dimension lattice ECDQ, it's redundancy, at the limit of large dimension, is still ~ 0.5 bit per sample for balanced descriptions $d_1 = d_2 \gg d_0$. For un-balanced descriptions, or for low ratios d_1/d_0, this 0.5 bit loss is eliminated. An interesting modification of TSDQ - *refinement time sharing* - may improve the performance in the balanced case by up to $\sim$ 1dB.

The paper is organized as follows: Section 2 summarizes the basic properties of the single description ECDQ, and some of the important theoretic results for the MD problem which we use as a basis for comparison. Section 3 introduces the two stage dithered quantizer (TSDQ) and shows an equivalent set of additive noise channels. This section also discusses two variants of the basic TSDQ: refinement time sharing and dependent dithering. Section 4 starts out by describing the periodic MDSQ (PMDSQ), and continues by proposing an equivalent additive noise "sum channel" for the entropy coded dithered PMDSQ. Both sections compare the entropy rates of the schemes with the information theoretic bounds.

2 Background

2.1 Entropy Coded Dithered Quantization: Basic Properties

To better understand the motivation for using lattice quantizers and ECDQ, [25, 24, 26], we briefly review definitions and properties of the ECDQ. A K dimensional lattice quantizer is formed from a lattice L_K. The quantizer $Q_K(\cdot)$ maps every vector $\mathbf{x} \in R^K$ into the nearest lattice point $\mathbf{l}_i \in L_K$; the region of all K-vectors mapped into a lattice point $\mathbf{l}_i \in L_K$ is the Voronoi region,

$$V(\mathbf{l}_i) = \{\mathbf{x} \in R^K : \|\mathbf{x} - \mathbf{l}_i\| \leq \|\mathbf{x} - \mathbf{l}_j\|, \text{ for all } j \neq i\}.$$

The dither $\mathbf{Z}$ is a K-dimensional random vector, independent of the source, and uniformly distributed over the basic cell of the lattice, V_0, which is the Voronoi region of the lattice point $\mathbf{0}$. We assume that the encoder and the decoder share common randomness, so the dither vector is available to both of them (in this sense $\mathbf{Z}$ is a "pseudo-random" vector). The normalized second moment of the lattice, G_K, characterizes the second moment of the dither vector

$$\frac{1}{K}E\|\mathbf{Z}\|^2 = G_K V^{2/K}$$

where V denotes the volume of V_0. Both the entropy encoder and decoder are conditioned on the dither's sample $\mathbf{Z}$. We assume that the entropy coder is complex enough, such that it's rate is close to the input entropy. The lattice quantizer with dither represents the source vector $\mathbf{x}$ by a vector $\mathbf{u}$, $\mathbf{u} = Q_K(\mathbf{x} + \mathbf{z}) - \mathbf{z}$, where $\mathbf{z}$ is a sample of the random vector $\mathbf{Z}$.

The resulting properties of the ECDQ are:

1. The quantization error vector $\mathbf{U} - \mathbf{X}$ is independent of $\mathbf{X}$ and is distributed as $-\mathbf{Z}$. In particular, the mean squared quantization error is given by the second moment of the dither, independently of the source distribution, i.e.,

$$d = \frac{1}{K}E\|\mathbf{U} - \mathbf{X}\|^2 = \frac{1}{K}E\|\mathbf{Z}\|^2 = G_K V^{2/K}.$$

2. The coding rate of the ECDQ, i.e., the conditional entropy of the dithered quantizer given the dither, is equal to the mutual information between the input and output of an additive noise channel $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, where $\mathbf{N}$, the channel's noise, has the same pdf as $-\mathbf{Z}$, [25]:

$$H(Q_K(\mathbf{X} + \mathbf{Z})|\mathbf{Z}) = I(\mathbf{X}; \mathbf{Y}) = h(\mathbf{X} + \mathbf{N}) - h(\mathbf{N}) \tag{1}$$

where $H(\cdot|\mathbf{Z})$ denotes conditional entropy given the dither, $I(\cdot;\cdot)$ denotes mutual information, and $h(\cdot)$ denotes differential entropy.

3. When optimal pre and post filters are used, then the redundancy over $R(d)$, the rate-distortion function of the source, satisfies a "divergence bound", [26]:

$$\frac{1}{K} H(Q_K|\mathbf{Z}) \leq R(d) + \frac{1}{2}\log 2\pi e G_K + \frac{1}{K} D(\mathbf{X}; \mathbf{X}^*) \tag{2}$$

where for short we use Q_K to denote $Q_K(\mathbf{X} + \mathbf{Z})$, and $D(\mathbf{X}; \mathbf{X}^*)$ denotes the divergence of a source sample X from Gaussianity

$$D(\mathbf{X}; \mathbf{X}^*) = \int f_{\mathbf{X}}(\alpha) \log(\frac{f_{\mathbf{X}}(\alpha)}{f_{\mathbf{X}^*}(\alpha)}) d\alpha$$

where $f_{\mathbf{X}^*}(\cdot)$ denotes the density function of a Gaussian source $\mathbf{X}^*$ with the same mean and covariance matrix as $\mathbf{X}$. This result is important for low resolution quantization. At high resolution the filter degenerates, and a tighter characterization can be obtained, where the divergence term in (2) is eliminated and the inequality becomes an equality.

4. For optimal lattice quantizers, i.e., lattice quantizers with the minimal normalized second moment G_K, the autocorrelation of the quantizer noise is "white" [24], i.e.

$$R_{\mathbf{Z}} = E(\mathbf{Z}\mathbf{Z}^t) = \sigma^2 I \tag{3}$$

where I is the $K \times K$ identity matrix, $\sigma^2 = G_K^{opt} V^{\frac{2}{K}}$ is the second moment of the lattice, and $G_K^{opt} = \min_{Q_K} \frac{1}{K} \frac{\int_{V_0} \|\mathbf{x}\|^2 d\mathbf{x}}{V^{1+\frac{2}{K}}}$ is the minimal normalized second moment of a K-dimensional lattice. Thus, the samples of the quantization noise are uncorrelated and have the same power.

2.2 The Theoretic Rate Region for Multiple Descriptions

As shown in [23], Ozarow's "Shannon theoretic" solution [15, 6] for the multiple description problem for a Gaussian source $X \sim N(0, \sigma_x^2)$ and mean squared error

(MSE) distortion measure can be written in the form

$$R_1 \geq \tfrac{1}{2}\log\left(\frac{\sigma_x^2}{d_1}\right) \tag{4a}$$

$$R_2 \geq \tfrac{1}{2}\log\left(\frac{\sigma_x^2}{d_2}\right) \tag{4b}$$

$$R_1 + R_2 \geq \tfrac{1}{2}\log\left(\frac{\sigma_x^4}{d_1 d_2}\right) + \delta \tag{4c}$$

where δ is defined by

$$\delta = \begin{cases} \frac{1}{2}\log\left(\frac{1}{1-\rho^2}\right), & d_0 \leq d_{0max} \\ 0, & d_0 \geq d_{0max}\ , \end{cases} \tag{4d}$$

$\rho = \rho(\sigma_x^2, d_0, d_1, d_2)$ denotes the negative correlation coefficient between the two additive Gaussian noises in the optimum test channel for a Gaussian source, and

$$d_{0max} = \frac{d_1 d_2}{d_1 + d_2 - d_1 d_2/\sigma_x^2}\ . \tag{4e}$$

This rate region characterizes the performance of an optimum block coding scheme of unbounded dimension. See [15, 6, 23] for further details.

In [23] inner and outer bounds on the achievable region $\mathcal{R}_X(d_1, d_2, d_0)$ for a general memoryless (non Gaussian) source X based on the Gaussian region are developed. The bounds have the form

$$\mathcal{R}^*(\sigma_x^2, d_0, d_1, d_2) \subseteq \mathcal{R}_X(d_0, d_1, d_2) \subseteq \mathcal{R}^*(P(X), d_0, d_1, d_2) \tag{5}$$

where $R^*(\sigma^2, d_1, d_2, d_0)$ denotes the region of a Gaussian source $N(0, \sigma^2)$ as given in (4), and σ_x^2 and $P(X)$ denote the variance and the entropy power of X, i.e., $P(X) = \frac{1}{2\pi e}2^{2h(X)}$. This result was also examined in high resolution conditions and it was shown that if $P(X) > 0$, then the rate region $\mathcal{R}_X(d_0, d_1, d_2)$ approaches the outer bound $\mathcal{R}^*(P(X), d_0, d_1, d_2)$, as $d_1, d_2 \to 0$, i.e., it contains all rate pairs satisfying

$$R_1 \geq \tfrac{1}{2}\log\left(\frac{P(X)}{d_1}\right) + \epsilon \tag{6a}$$

$$R_2 \geq \tfrac{1}{2}\log\left(\frac{P(X)}{d_2}\right) + \epsilon \tag{6b}$$

$$R_1 + R_2 \geq \tfrac{1}{2}\log\left(\frac{P^2(X)}{d_1 d_2}\right) + \delta_{HR}(d_0, d_1, d_2) + \epsilon \tag{6c}$$

where $\delta_{HR}(d_0, d_1, d_2) \triangleq \lim_{\sigma_x^2 \to \infty} \delta(\sigma_x^2, d_0, d_1, d_2)$, and $\epsilon \to 0$ as $d_1, d_2 \to 0$. We use these rate-region characterizations to evaluate our coding schemes.

3 The Two Stage Dithered Quantizer

In this section we introduce a vector quantizer for two descriptions that is composed of three ECDQs. The encoding consists of two stages. In the first stage two independent low-resolution descriptions are generated using two ECDQs with statistically independent dithers. In the second stage, a refinement is generated by encoding the estimation error of the source from the two low resolution codes with a third ECDQ. This refinement code is divided between the two descriptions and is used only by the combined ("central") decoder.

3.1 Description of Scheme

Let $Q_1(\cdot), Q_2(\cdot)$ and $Q_3(\cdot)$ denote K dimensional lattice quantizers and let $\mathbf{z}_1, \mathbf{z}_2$ and $\mathbf{z}_3$ denote samples of the corresponding K-dimensional random variables $\mathbf{Z}_1, \mathbf{Z}_2$ and $\mathbf{Z}_3$. The latter are statistically independent and each is uniformly distributed over the basic cell of the associated lattice quantizer. The lattices have a "white" quantization noise covariance matrix of the form $\sigma_i^2 I = E(\mathbf{Z}_i \mathbf{Z}_i^t),\ i = 1, 2, 3$, where σ_i^2 is the second moment of the lattice quantizer $Q_i(\cdot)$.

The two marginal code words are c_1 and c_2 which are the entropy encodings of $Q_1(\mathbf{X}+\mathbf{Z}_1)$ and $Q_2(\mathbf{X}+\mathbf{Z}_2)$ conditioned on $\mathbf{Z}_1$ and $\mathbf{Z}_2$. The source is assumed to have a zero mean and a covariance matrix of the form $\sigma_x^2 I$. Let $\mathbf{X}_{1,2}$ denote the optimal linear estimation of $\mathbf{X}$, in the MSE sense, from the two marginal reconstructions after subtracting the dither, i.e.,

$$\mathbf{x}_{1,2} = a(Q_1(\mathbf{x}+\mathbf{z}_1) - \mathbf{z}_1) + b(Q_2(\mathbf{x}+\mathbf{z}_2) - \mathbf{z}_2).$$

Linear minimum-MSE estimation was chosen because of its simplicity, and since it is universal over the class of sources with fixed variance. Furthermore, as $K \to \infty$ it becomes optimal due to the asymptotic Gaussianity of lattice quantization noise [24].

The scalars a and b in (3.1) are given by:

$$a = \frac{\sigma_x^2 \sigma_2^2}{\sigma_x^2\sigma_1^2 + \sigma_x^2\sigma_2^2 + \sigma_1^2\sigma_2^2} \qquad b = \frac{\sigma_x^2 \sigma_1^2}{\sigma_x^2\sigma_1^2 + \sigma_x^2\sigma_2^2 + \sigma_1^2\sigma_2^2}.$$

They were chosen to minimize $E(\mathbf{X}_{1,2} - \mathbf{X})^2$, where the expectation is taken over X, Z_1 and Z_2, taking into account the whiteness of the quantization noise, $Q_i(\mathbf{X} + \mathbf{Z}_i) - \mathbf{Z}_i - \mathbf{X},\ i = 1, 2$, and the mutual independence of the noises and the source [26]. We denote the estimation error $\mathbf{X} - \mathbf{X}_{1,2}$ by $\mathbf{W}$. The "refinement" codeword, c_3, is the entropy coding of the dithered quantization of $\mathbf{W}$, i.e. $Q_3(\mathbf{W} + \mathbf{Z}_3)$. Finally, the encoder produces two codewords C_1 and C_2, each of these words is a union of a marginal codeword and a part of c_3. The encoding scheme can be seen in Figure 1. The coding rates of c_1, c_2, c_3 are

$$\begin{aligned} KR_1' &= H(Q_1|\mathbf{Z}_1) \qquad KR_2' = H(Q_2|\mathbf{Z}_2) \\ KR_3 &= H(Q_3|\mathbf{Z}_3, Q_2, \mathbf{Z}_2, Q_1, \mathbf{Z}_1). \end{aligned} \tag{7}$$

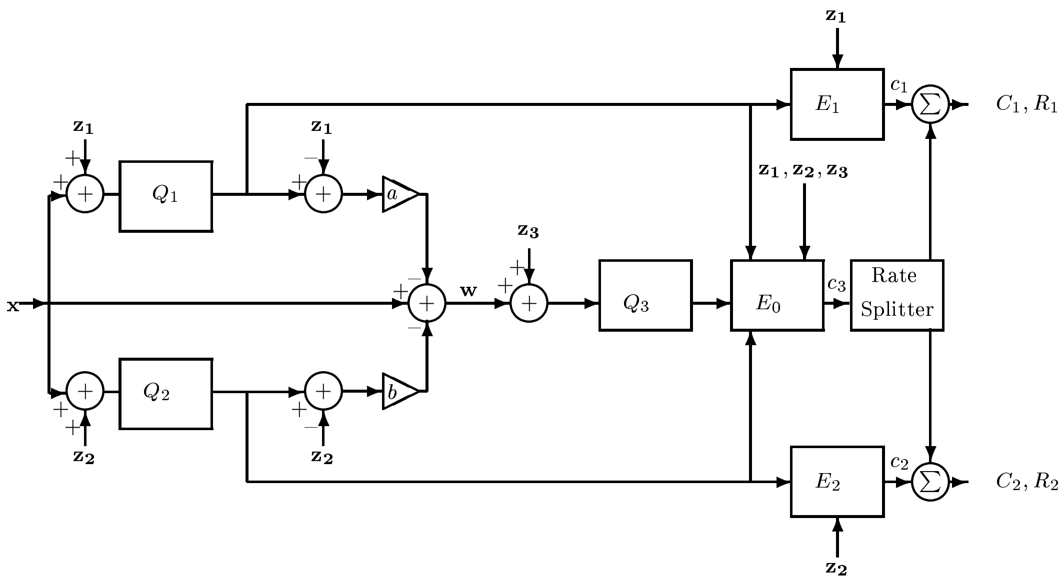

Figure 1: The TSDQ encoder. E denotes lossless (entropy) encoder.

Unlike the equations for R_1' and R_2' which are familiar from the single description ECDQ, the equation regarding R_3 is unique to multiple descriptions in the sense that it is conditioned not only on the dither $\mathbf{Z}_3$, but also on the outcomes and the dithers of the two "side decoders". The resulting rate pair , i.e., that of the two descriptions, is

$$R_1 = R_1' + \alpha R_3 \quad \text{and} \quad R_2 = R_2' + (1-\alpha)R_3, \tag{8}$$

where $0 \leq \alpha \leq 1$ is a free parameter. Hence, each "marginal receiver" receives only one of the codewords (C_1 or C_2), whereas the "central receiver" receives all three code words.

With slight manipulation of [6, Section 6], it can be shown that for fixed $Q_1(\cdot), Q_2(\cdot)$ and $Q_3(\cdot)$, this set of rate pairs $\{R_1, R_2\}$ is the minimum achievable [3]. See also [7].

The two "marginal receivers" reconstruct the source using the optimal linear estimation in the MSE sense ("post filtering"):

$$\hat{\mathbf{x}}_i = \alpha_i(Q_i(\mathbf{x}+\mathbf{z}_i) - \mathbf{z}_i) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_i^2}(Q_i(\mathbf{x}+\mathbf{z}_i) - \mathbf{z}_i) \qquad i = 1, 2. \tag{9}$$

The resulting distortion is

$$d_i = \frac{\sigma_i^2 \sigma_x^2}{\sigma_i^2 + \sigma_x^2} \qquad i = 1, 2. \tag{10}$$

Similarly, the "central receiver" uses all three codewords to reconstruct the source vector as

$$\hat{\mathbf{x}}_0 \;=\; a(Q_1(\mathbf{x}+\mathbf{z}_1) - \mathbf{z}_1) + b(Q_2(\mathbf{x}+\mathbf{z}_2) - \mathbf{z}_2) + \alpha_0(Q_3(\mathbf{w}+\mathbf{z}_3) - \mathbf{z}_3) \tag{11}$$

where $\alpha_0 = \frac{\sigma_x^2\sigma_1^2\sigma_2^2}{\sigma_x^2\sigma_1^2\sigma_2^2+\sigma_3^2(\sigma_1^2\sigma_2^2+\sigma_1^2\sigma_x^2+\sigma_2^2\sigma_x^2)}$ and the resulting distortion is

$$d_0 = \frac{\sigma_x^2\sigma_1^2\sigma_2^2\sigma_3^2}{\sigma_x^2\sigma_1^2\sigma_2^2 + \sigma_3^2(\sigma_1^2\sigma_2^2 + \sigma_1^2\sigma_x^2 + \sigma_2^2\sigma_x^2)} \ . \tag{12}$$

Note that σ_i^2, being the second moment of the lattice, can be tuned to guarantee distortion d_i, and is usually larger than d_i, with equality at high resolution conditions. Note also that the the scheme is "universal" in the sense that Q_1, Q_2, and Q_3 do not depend on the source statistics but its variance.

3.2 Performance Analysis

The following theorem provides a simple information theoretic characteristic for the behavior of the TSDQ.

Theorem 3.1. *The proposed encoding scheme is equivalent to the additive noise channels shown in Figure 2, in the sense that*

$$\begin{aligned} R_1' &= \frac{1}{K}I(\mathbf{X};\mathbf{X}+\mathbf{N}_1) \quad R_2' = \frac{1}{K}I(\mathbf{X};\mathbf{X}+\mathbf{N}_2) && (13)\\ R_3 &= \frac{1}{K}I(\mathbf{U};\mathbf{U}+\mathbf{N}_3|\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2) && (14)\\ &= \frac{1}{K}I(\mathbf{X};\mathbf{X}+\mathbf{N}_3|\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2) && (15)\end{aligned}$$

where $\mathbf{N}_i$ *is distributed as* $-\mathbf{Z}_i$, $i = 1,2,3$, *and is independent of the source. The random vector* $\mathbf{U}$ *denotes* $\mathbf{X} - a(\mathbf{X}+\mathbf{N}_1) - b(\mathbf{X}+\mathbf{N}_2)$. *Furthermore, the equivalence holds also with respect to the distribution of the errors after applying the post filters* $\alpha_1, \alpha_2, \alpha_0, a$ *and* b.

Proof Equation (13) is identical to the result stated in (1) and therefore needs no further proof. To prove (15) we need the following Lemma, the proof of which can be found in Appendix A.

Lemma 1. *The joint density of* $(\mathbf{X},\mathbf{X}+\mathbf{N_1},\mathbf{X}+\mathbf{N_2},\mathbf{U},\mathbf{U}+\mathbf{N_3})$ *is equal to the joint density of* $(\mathbf{X}, Q_1 - \mathbf{Z}_1, Q_2 - \mathbf{Z}_2, \mathbf{W}, Q_3 - \mathbf{Z}_3)$.

To prove equation (15), we note the following. Since $\mathbf{Z_i}$ is a deterministic function of $Q_i - \mathbf{Z_i}$, i.e.,

$$\mathbf{Z}_i = Q_i(\mathbf{v_i}) - \mathbf{v_i}, \quad \text{where } \mathbf{v_i} = Q_i - \mathbf{Z}_i, \tag{16}$$

we can re-write (7) as

$$KR_3 = H(Q_3|\mathbf{Z}_3, Q_2 - \mathbf{Z}_2, Q_1 - \mathbf{Z}_1).$$

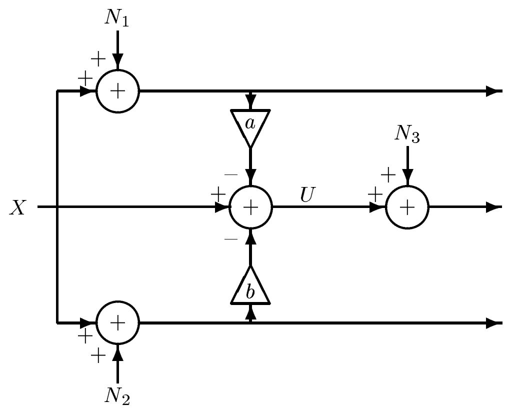

Figure 2: The equivalent channel of the TSDQ.

Finally, since $H(Q_3|\mathbf{Z}_3, Q_2-\mathbf{Z}_2, Q_1-\mathbf{Z}_1, \mathbf{X}) = 0$ we can replace the entropy by mutual information, i.e.

$$\begin{aligned}
KR_3 &= I(Q_3; \mathbf{X}|\mathbf{Z}_3, Q_2 - \mathbf{Z}_2, Q_1 - \mathbf{Z}_1) \\
&= I((Q_3 - \mathbf{Z}_3); \mathbf{X}|\mathbf{Z}_3, Q_2 - \mathbf{Z}_2, Q_1 - \mathbf{Z}_1) \\
&= h(\mathbf{X}|\mathbf{Z}_3, Q_2 - \mathbf{Z}_2, Q_1 - \mathbf{Z}_1) - h(\mathbf{X}|\mathbf{Z}_3, Q_2 - \mathbf{Z}_2, Q_1 - \mathbf{Z}_1, Q_3 - \mathbf{Z}_3) \quad (17) \\
&= h(\mathbf{X}|Q_2 - \mathbf{Z}_2, Q_1 - \mathbf{Z}_1) - h(\mathbf{X}|Q_2 - \mathbf{Z}_2, Q_1 - \mathbf{Z}_1, Q_3 - \mathbf{Z}_3) \quad (18) \\
&= I((Q_3 - \mathbf{Z}_3); \mathbf{X}|Q_2 - \mathbf{Z}_2, Q_1 - \mathbf{Z}_1) \quad (19)
\end{aligned}$$

where in (17) we decomposed the mutual information as a difference between differential entropies, and in (18) we used the fact that $\mathbf{X}$ is independent of $\mathbf{Z}_3$, and that $\mathbf{Z}_3$ is a deterministic function of $Q_3 - \mathbf{Z}_3$ (see (16)). If $\mathbf{X}$ doesn't have a finite differential entropy, we can apply the chain rule to the mutual information instead of equations (17) and (18), and get the same result. These facts and Lemma 1 complete the proof of Theorem 3.1.■

Remark: Since by the orthogonality principle $\mathbf{U}$ is uncorrelated with $\mathbf{X} + \mathbf{N}_1$ and $\mathbf{X} + \mathbf{N}_2$, we conclude from (14) that conditioning only on the dithers in coding Q_3 should not be very costly in terms of rate.

In order to gain more insight into the non explicit expressions in Theorem 3.1, we take advantage of the information theoretic properties of the equivalent additive noise channel, and provide three simple bounds for the rate region. The first bound holds at any resolution and is relatively tight for a Gaussian source. The second bound becomes tight at high resolution conditions for any "smooth" source. The third bound corresponds to the other extreme of low resolution.

From [25] we know that

$$R_i' \leq R^*(d_i) + \frac{1}{2}\log 2\pi e G_{Ki} \qquad i = 1, 2 \tag{20}$$

where $R^*(d_i)$ is the rate distortion function of $X^* \sim N(\mu_x, \sigma_x^2)$ which is $\frac{1}{2}\log(\frac{\sigma_x^2}{d_i})$. As for R_3, we start out from (14) and have

$$
\begin{aligned}
KR_3 &= I(\mathbf{U};\mathbf{U}+\mathbf{N}_3|\mathbf{X}+\mathbf{N}_1,\mathbf{X}+\mathbf{N}_2) \\
&= h(\mathbf{U}+\mathbf{N}_3|\mathbf{X}+\mathbf{N_1},\mathbf{X}+\mathbf{N_2}) \qquad (21)\\
&\quad -h(\mathbf{U}+\mathbf{N}_3|\mathbf{X}+\mathbf{N_1},\mathbf{X}+\mathbf{N_2},\mathbf{X}) \\
&= h(\mathbf{U}+\mathbf{N}_3|\mathbf{X}+\mathbf{N_1},\mathbf{X}+\mathbf{N_2}) - \mathbf{h}(\mathbf{N_3}|\mathbf{X}+\mathbf{N_1},\mathbf{X}+\mathbf{N_2},\mathbf{X}) \\
&= h(\mathbf{U}+\mathbf{N}_3|\mathbf{X}+\mathbf{N_1},\mathbf{X}+\mathbf{N_2}) - \mathbf{h}(\mathbf{N_3}).
\end{aligned}
\qquad (22)
$$

The last equation is true because $\mathbf{N}_3$ is independent of $\mathbf{X},\mathbf{N}_1$ and $\mathbf{N}_2$. By removing the conditioning and recalling that a Gaussian variable has the largest entropy, we can upper-bound R_3 by

$$R_3 \leq \frac{1}{2}\log\left(2\pi e\left(\frac{\sigma_x^2\sigma_2^2\sigma_1^2}{\sigma_x^2\sigma_1^2+\sigma_x^2\sigma_2^2+\sigma_1^2\sigma_2^2}+\sigma_3^2\right)\right)-\frac{1}{K}h(\mathbf{N}_3).$$

Finally by using $\frac{1}{K}h(\mathbf{N}_3)=\frac{1}{2}\log(\frac{\sigma_3^2}{G_{K3}})$, (10) and (12), we get

$$R_1' \leq \frac{1}{2}\log\left(\frac{\sigma_x^2}{d_1}\right)+\frac{1}{2}\log(2\pi e G_{K1}) \qquad (23)$$

$$R_2' \leq \frac{1}{2}\log\left(\frac{\sigma_x^2}{d_2}\right)+\frac{1}{2}\log(2\pi e G_{K2}) \qquad (24)$$

$$
\begin{aligned}
R_1'+R_2'+R_3 &\leq \frac{1}{2}\log\left(\frac{\sigma_x^4}{d_1d_2}\right) \\
&+ \frac{1}{2}\log\left(\frac{\sigma_x^2 d_1 d_2}{d_0(\sigma_x^2(d_1+d_2)-d_1d_2)}\right)+\frac{1}{2}\sum_{i=1}^{3}\log(2\pi e G_{Ki})
\end{aligned}
\qquad (25)
$$

where G_{Ki} is the normalized second moment of lattice i. The tightness of this bound depends on the "closeness" of the source to Gaussianity.

A tighter bound can be found at high resolution conditions where $\sigma_x^2 \gg d_i$, provided that the entropy power of the source $P(X)=\frac{2^{2h(X)}}{2\pi e}$ is positive. Specifically, following the method of [26], we have

$$h(X+N_i)=\frac{1}{2}\log(2\pi e P(X))+o(1)$$

where $o(1)\to 0$ as $\sigma_i \to 0$ for $i=1,2$, which when substituted in (13) gives

$$R_i' = \frac{1}{2}\log\left(\frac{P(X)}{d_1}\right)+\frac{1}{2}\log(2\pi e G_{Ki})+o(1) \quad i=1,2.$$

Hence, at high resolution the bound takes the form of (23), (24) and (25), only σ_x^2 is replaced by $P(X)$, and a $o(1)$ term is added to the right hand sides, where $o(1)\to 0$ as d_1, d_2 and $d_0 \to 0$. Note that in the limit the second term in (25) simplifies to $\frac{1}{2}\log((d_1d_2/d_0(d_1+d_2))$.

On the other extreme, in the limit of very coarse quantization, we have $I(X;X+N_i)\to 0$ as $\sigma_i\to\infty$. In particular, the case where no refinement is transmitted corresponds to $\sigma_3^2=\infty$, implying $R_3=0$, and $d_0=(1/\sigma_x^2+1/\sigma_1^2+1/\sigma_2^2)^{-1}=d_{0max}$, where d_{0max} is defined in (4e). See (10) and (12).

3.3 Comparison with Theoretical Bounds

We are now in a position to compare these bounds to the information theoretic optimum solution for a Gaussian source [15] and for a general source at high resolution [23]. Denote the total rate in the proposed scheme $TR_s \triangleq R_1'+R_2'+R_3$ and the optimal total rate by TR_o. For a Gaussian source under high resolution conditions we obtain from [15]

$$TR_o \geq \tfrac{1}{2}\log\left(\frac{\sigma_x^4}{d_1 d_2}\right) + \frac{1}{2}\log\frac{s_1 s_2}{s_1+s_2-2+2\sqrt{s_1-1}\sqrt{s_2-1}} + o(1) \tag{26}$$

where s_i is used to denote $\frac{d_i}{d_0} \geq 1$, and the $o(1)$ term vanishes as $d_1, d_2, d_0 \to 0$. For a general source with $P(X) > 0$, inequality (26) holds with σ_x^2 replaced by $P(X)$ [23]. Combining with (25), we bound the excess total rate for a general source at high resolution by

$$\begin{aligned} TR_s - TR_o &\leq \frac{1}{2}\log\left(\frac{s_1+s_2-2+2\sqrt{s_1-1}\sqrt{s_2-1}}{s_1+s_2}\right) \\ &+ \sum_{i=1}^{3}\frac{1}{2}\log 2\pi e G_{Ki} + o(1) \\ &\leq \frac{1}{2} + \sum_{i=1}^{3}\frac{1}{2}\log 2\pi e G_{Ki} + o(1) \end{aligned} \tag{27}$$

where $o(1) \to 0$ as $d_1, d_2, d_0 \to 0$. The last inequality becomes tight when $s_1 = s_2 \to \infty$, i.e., for balanced descriptions with $d_1 = d_2 \gg d_0$. The term involving s_1, s_2 vanishes if either s_2/s_1 or $s_1/s_2 \to \infty$, or if $1/s_1 + 1/s_2 = 1$. The former case corresponds to very un-balanced descriptions. The latter case corresponds to having $d_0 = d_{0max}$, implying that Q_3 is idle, $\sigma_3^2 = \infty$, $R_3 = 0$, and the redundancy is only $\sum_{i\in\{1,2\}} \frac{1}{2}\log 2\pi e G_{Ki}$.

For a Gaussian source, we can extend the bounds to lower resolution levels in some cases. For balanced descriptions ($d_1 = d_2$), the bounds in (27) hold at all resolutions. In the case of successive refinement, only d_1 and d_0 are of interest ($d_2 = \sigma_x^2$), quantizer Q_2 is idle ($\sigma_2^2 = \infty$), $R_2 = 0$, and the excess rate becomes only $\sum_{i\in\{1,3\}} \frac{1}{2}\log 2\pi e G_{Ki}$. Finally, in the case where $d_0 = d_{0max}$, we have $R_3 = 0$, and the excess rate is only $\sum_{i\in\{1,2\}} \frac{1}{2}\log 2\pi e G_{Ki}$.

If optimal lattice quantizers are used, then $G_K \to \frac{1}{2\pi e}$ as $K \to \infty$ [24]. In this case, the redundancy of the rate sum with regards to the optimal rate sum for a Gaussian source is asymptotically 0.5 bit per sample for balanced descriptions $d_1 = d_2 \gg d_0$, and it is asymptotically *zero* for successive refinement or for $d_0 = d_{0max}$. Figure 3 illustrates for a Gaussian source and balanced descriptions the optimum rate sum, and the rate sum of the TSDQ for $K = 1$ (where $G_1 = 1/12$) and $K \to \infty$ (where $G_K = 1/2\pi e$).

3.4 Refinement Time Sharing

Although the refinement encoding stage assumes that *both* coarse encodings are available at the decoder, the refinement code may be useful even if only one of them is available. In fact, at moderate side-to-side distortion ratios d_1/d_2 and high side-to-central distortion ratio d_1/d_0, a marginal decoder can significantly improve its reconstruction provided that it gets the entire refinement information c_3, i.e., if $\alpha = 0$ or 1 in (8). Since generally $0 < \alpha < 1$, instead of dividing c_3 between the two descriptions at each encoding session, we may time share it between the two descriptions, thus enhance $\hat{\mathbf{X}}_\mathbf{1}$ a fraction α of the time and $\hat{\mathbf{X}}_\mathbf{2}$ a fraction $1-\alpha$ of the time.

To illustrate this enhancement, consider the balanced high resolution case, where $a = b = 0.5$, $\sigma_1^2 = \sigma_2^2 = d_1 = d_2$, and $\alpha_0 = \alpha_1 = \alpha_2 = 1$; furthermore, assume that $d_0 \ll d_1$ so that $\sigma_3^2 = d_0$. In terms of the equivalent additive noise channel, coarse descriptions C_1 and C_2 amount to $\mathbf{X}+\mathbf{N_1}$ and $\mathbf{X}+\mathbf{N_2}$, respectively, while c_3 amounts to

$$\mathbf{U} + \mathbf{N_3} = \frac{\mathbf{N_1} + \mathbf{N_2}}{2} + \mathbf{N_3}.$$

Given C_1 alone, decoder 1's reconstruction is equivalent to $\hat{\mathbf{X}}_\mathbf{1} = \mathbf{X} + \mathbf{N_1}$, resulting in distortion $d_1 = \sigma_1^2$. However, given C_1 and c_3, the decoder can improve this reconstruction to

$$\hat{\mathbf{X}}_\mathbf{1} = [\mathbf{X} + \mathbf{N_1}] - \beta[\frac{\mathbf{N_1} + \mathbf{N_2}}{2} + \mathbf{N_3}]$$

where the second term is the optimum MSE estimator of N_1 given the information in c_3; under the assumptions above $\beta^{opt} = 1$, and the resulting distortion is $d_1^{enhanced} = d_1/2$, i.e., 3 dB better!

Of course, since decoder 1 has access to c_3 only half the time, the average distortion is $0.5[d_1 + d_1/2] = 0.75 d_1$, i.e., 1 dB better on the average.

The exact expression for the ratio $d_1^{enhanced}/d_1$, when $d_1 = d_2$ and $\alpha = 0.5$ is:

$$\frac{d_1^{enhanced}}{d_1} = 1/2\,\frac{\sigma_x^2{\sigma_1}^4 + 3\,{\sigma_x}^4\sigma_1^2 + 8\,\sigma_3^2{\sigma_x}^4 + 8\,\sigma_3^2\sigma_x^2\sigma_1^2 + 2\,\sigma_3^2{\sigma_1}^4}{(2\,\sigma_x^2 + \sigma_1^2)\,(\sigma_x^2\sigma_1^2 + 2\,\sigma_x^2\sigma_3^2 + \sigma_3^2\sigma_1^2)} \tag{28}$$

Figure 3 illustrates the improvement achieved by this approach, as well as Ozarow's optimum curve for a memoryless Gaussian source. The descriptions are balanced, the source has a unit variance and $d_0 = 10^{-6}$.

3.5 Staggered Quantizers and Dependent Dithers

Staggering is a simple way to enhance the joint performance of two quantizers in parallel. This idea motivated us to explore dependent dithering; in fact, this is where this research originally started.

The dithered scalar staggered quantizer is a uniform scalar quantizer and two dithers that are added to the source and quantized separately. The two dithers are related via $Z_2 = Z_1 + \Delta/2$, where Z_1 is uniform over $(-\Delta/2, \Delta/2)$. The resulting

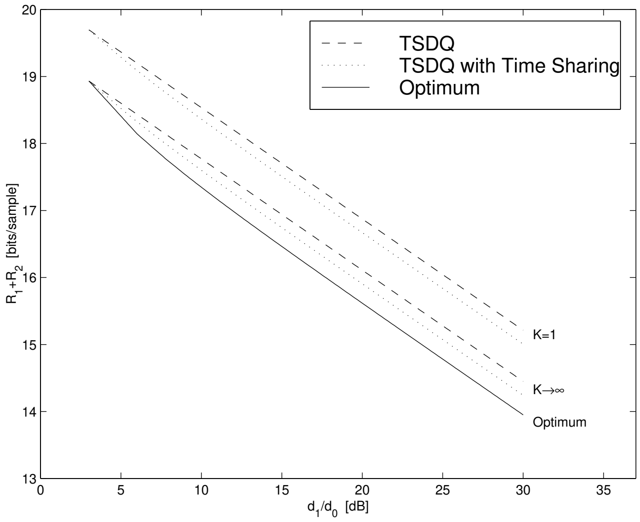

Figure 3: TSDQ with time sharing compared to TSDQ and to the optimum, for a unit variance, memoryless Gaussian source, $d_0 = 10^{-6}$ and $d_1 = d_2$.

correlation coefficient of the quantization errors (not the dithers) is $-1/2$. This allows to get $d_1/d_0 = 6$ dB instead of 3dB in the independent dither case for the same coding rate! To get higher ratios of d_1/d_0, a refinement stage (with an independent dither) is needed.

Unfortunately, our experiments show that at higher dimensions (e.g., with hexagonal lattices in the 2 dimensional case) the staggering gain decreases. This behavior is not surprising in light of the high-dimensional performance predicted by the optimum theoretic solution. Ozarow's solution for the case of no-excess marginal rate implies that, for balanced descriptions, d_0 is only 3dB better than d_1 $(= d_2)$; see Section 2.2. This indicates that as the lattice dimension goes to infinity, the staggering gain must vanish, since otherwise the resulting performance could have been better than the optimum solution! (On the other hand, it is interesting to observe that this somewhat counter-intuitive loss of staggering gain at large dimensions is compensated for almost exactly by the space filling gain of the two lattices.) Thus, we do not expect that correlating Z_1 and Z_2 will enhance the TSDQ performance for general lattice dimensions. Furthermore, it does not seem we can benefit from correlating (Z_1, Z_2) with Z_3, the dither at the refinement stage.

4 Dithering the Vaishampayan Quantizer

The performance of the two-stage encoder does not seem to follow along the expected properties of a multiple description quantizer. Ozarow's solution for a Gaussian source, [15], is given in terms of a test channel. In this test channel, the "side" quantizer errors are negatively correlated, with correlation coefficient ρ as appears in (4d). This negative correlation is the basis for most of the gain of the central decoder. For example, in the balanced case the central decoder gains the first 3dB with uncorrelated errors ($\rho = 0$), while to increase d_1/d_0 further ρ must become negative, and it goes to -1 as $d_1/d_0 \to \infty$.

The two stage encoder does not satisfy this property, since the two side quantization errors are independent. As discussed above, even dependent dithering fails to create dependent errors for large lattice dimensions. This lack of negative correlation may explain the 1/2 bit loss in rate in the balanced description case. Furthermore the TSDQ suffers the space filling loss of three lattice quantizers.

The work of Vaishampayan [19], on the multiple description scalar quantizer (MDSQ), seems promising in this sense. As shown in [20], it suffers the space filling loss of only two scalar quantizers at high resolution conditions. Furthermore, its side quantizer errors are negatively correlated as in Ozarow's test channel.

We describe below a periodic dithered version of Vaishampayan's MDSQ, which incorporates linear post-filters. In this version the quantizer does not depend on the source's distribution, and guarantees pre-specified distortions $d_1 = d_2$ and d_0. Furthermore, using the special properties of the ECDQ, we optimize the post-filters, and develop analytic expressions for the entropy of the quantizers at any resolution. We also compare these expressions to the optimum theoretic solution.

4.1 Design of the Entropy Coded Dithered Periodic MDSQ

The MDSQ is a quantizer for balanced multiple descriptions. It's building blocks are a scalar quantizer that produces the index l, and an index assignment function $a_m(l)$, that produces two indexes i, j from l, where m is a design parameter that determines the ratio $\frac{d_1}{d_0}$. The index assignment is based on a matrix whose (i, j)th element is l. The set of indices that can be produced by a_m lie on the main diagonal of the matrix and on the $2m$ diagonals closest to it. The "central" cell indices are placed in the matrix in cells along lines which are perpendicular to the main diagonal. The direction of the sweep ($\nearrow$ or $\swarrow$) changes every $2m$ such lines, and hence a periodicity is built into the placement of the "central" cell indices. Following is an example where $m = 2$:

1	3	6			
2	5	8	9		
4	7	10	12	14	
	11	13	15	17	21
		16	18	20	23
			19	22	·

The indices i and j are transmitted over the two channels, and a "side" description is produced from each. The central decoder uses the inverse function $a_m^{-1}(\cdot)$ to deduce l from i and j.

At high resolution conditions, when entropy coding is applied, [18], the quantizer becomes a uniform and unbounded scalar quantizer with step size Δ, paralleling the well known Gish and Pierce result in the single description case [10]. The "side quantizers" are defined in this case by the rows and the columns of an infinite matrix.

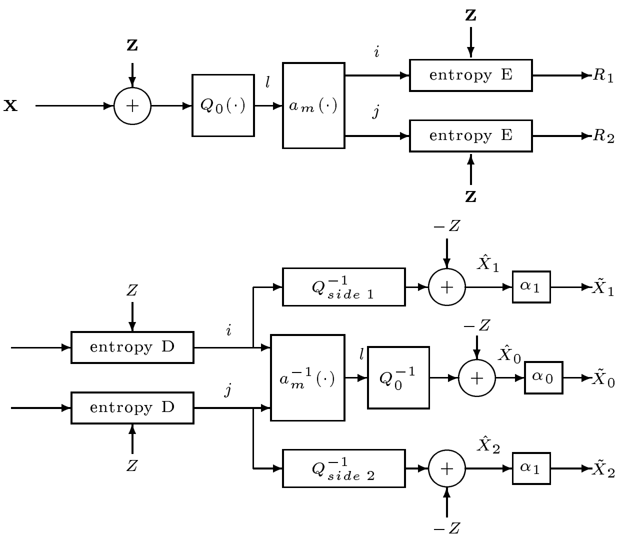

Figure 4: The ECD-PMDSQ Encoder and Decoder.

Each "side" cell is built of $2m+1$ Δ-intervals ("components") that spreads over a range of size $(2m^2+m+1)\Delta$. The spreading pattern belongs to one of $2m$ possible "types". The $2m$ cell types are interlaced, so that all the various components appear in a period of size $(\#types) \times (\#components\ /\ type)\Delta = 2m(2m+1)\Delta$. A detailed formulation of the periodic structure of the periodic MDSQ (PMDSQ) can be found in Appendix B. To illustrate this, we take $m=2$, and tag one period of 20 "central" cells to reflect the type of "side" cell they belong to. Here each side cell is a union of 5 intervals, having one of 4 possible patterns, with diameter of 11Δ (4 and $4'$ denote components of two different type-4 cells):

1	4	$3'$	1	4	2	1	4	2	1	1	2	3	2	3	2	3	$4'$	3	$4'$

where the tagged period is preceded by $\ldots, 4, 3', 4$, and followed by $1', 4', 3, \ldots$. The second side quantizer is constructed of the same periodic structure, the only difference being that it is offset by half a period, $m(2m+1)\Delta$, with respect to the first. To average over all these possible cell types for each source sample x, the appropriate dither Z is uniformly distributed over a full such period, i.e., over $2m(2m+1)$ consequent "central" cells.

The entropy coded dithered PMDSQ (ECD-PMDSQ) is described in Figure 4. $Q_0(\cdot)$ denotes an unbounded uniform scalar quantizer with cell size Δ. The entropy encoders and decoders are conditioned on the dither and the scalar filters (α_0, α_1) will be defined in the next section. We define $Q_{side}(x)$ to be the centroid of side cell i for all $x \in$ "side" cell i, where the "centroid" is the expectation of a random variable uniformly distributed over cell i. Thus the three reconstructions are

$$\tilde{X}_i = \alpha_i(Q_{side\,i}(X+Z) - Z) \quad i = 1, 2 \quad \text{and} \quad \tilde{X}_0 = \alpha_0(Q_0(X+Z) - Z)$$

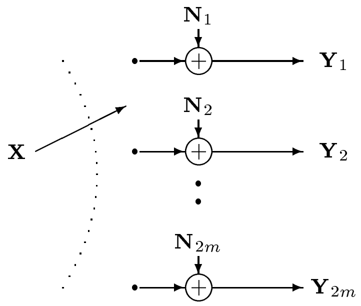

Figure 5: The Equivalent Sum Channel of the ECD-PMDSQ

where $Z \sim \mathcal{U}(0, 2m(2m+1)\Delta)$ denotes the dither, and the coding rate is $R_{ECD-PMDSQ} = H(Q_{side}(X+Z)|Z)$ bit per description per sample.

4.2 The Equivalent Additive Noise Sum Channel of the ECD-PMDSQ and Performance Analysis

The main observation made in this section is that the ECD-PMDSQ is equivalent to the additive noise sum channel in Figure 5. The so called "sum channel" has all $2m$ component channels available for use but only one can be used at any given time. The probability of using a certain component channel is $\frac{1}{2m}$. The input to the sum channel is the source to be quantized, and the additive noise N_t, $t = 1 \ldots 2m$, is uniformly distributed over cell of type t reflected around it's centroid. The equivalence is in the following senses: First, the quantization errors and the reconstructed sources are equal in distribution to the average additive noise and the average output, respectively, of the sum channel. Second, the rate of the ECD-PMDSQ can be written as the average mutual information between the input and output of the sum channel. This property corresponds to the equivalence of the ECDQ to an additive noise channel regarding a single description [26]. Following are two theorems which formulate this equivalence.

Theorem 4.1. *The side quantization error of the dithered PMDSQ, $Q_{side}(X+Z) - (X+Z)$, is independent of the source X and is distributed as N_T, where T is uniformly distributed over $\{1, 2, \cdots, 2m\}$.*

As a simple corollary from Theorem 4.1, $Q_{side}(X+Z) - (X+Z)$ is distributed as $Q_{side}(Z) - Z$.

Proof 4.1. The conditional probability function of the quantization error given the source sample is

$$\Pr\{Q_{side}(X+Z) - (X+Z) \leq \alpha | X\}.$$

We observe that the function $e(t) \triangleq Q_{side}(t) - t$ is periodic in the sense that $e(t) = e(t + 2m(2m+1)l\Delta)$ for every integer l, where Δ is the size of the "central" quantizer cell. Hence, also the indicator function,

$$I_\alpha(t) = \begin{cases} 1 & \text{if} \quad e(t) \le \alpha \\ 0 & otherwise \end{cases}$$

is periodic, with the same periodicity. Now,

$$\begin{aligned} \Pr\{Q_{side}(X+Z) - (X+Z) \le \alpha | X\} &= E_Z\{I_\alpha(X+Z)|X\} \\ &= \frac{1}{2m(2m+1)\Delta} \int_{-m(2m+1)\Delta}^{m(2m+1)\Delta} I_\alpha(X+t)\mathrm{d}t. \end{aligned} \tag{29}$$

The right term of (29) is an integral over a full period of a periodic function, and therefore independent of X. This proves that the quantization error is independent of the source sample. As for the distribution of the error, since it does not depend on the value of X, we substitute $X = 0$ and get

$$\Pr\{Q_{side}(X+Z) - (X+Z) \le \alpha\} = \Pr\{Q_{side}(Z) - Z \le \alpha\}. \tag{30}$$

Since Z is uniformly distributed over the $2m$ cell types, each cell type has an equal probability of $1/2m$. We use $ct(Z)$ to denote the cell type Z belongs to, and thus

$$\begin{aligned} \Pr\{Q_{side}(X+Z) - (X+Z) \le \alpha\} &= \sum_{i=1}^{2m} \Pr(ct(Z) = i) \Pr\{Q_{side}(Z) - Z \le \alpha | ct(Z) = i\} \\ &= \frac{1}{2m} \sum_{i=1}^{2m} \Pr(N_i \le \alpha) \\ &= \Pr\{N_T \le \alpha\}. \end{aligned}$$

■

Theorem 4.1 shows that as far as input/output relations are considered, the dithered PMDSQ is equivalent to the sum channel whose input is X and whose output is $Y_T = X + N_T$. The next theorem shows that the rate of the quantizer may also be stated in terms of the sum channel.

Theorem 4.2. *The rate of the "side" quantizers satisfies*

$$\begin{aligned} R_{ECD-PMDSQ} &\triangleq H(Q_{side}(X+Z)|Z) = I(X; X+N_T|T) \\ &= h(Y_T|T) - \log(2m+1)\Delta. \end{aligned} \tag{31}$$

Note that the term $\log(2m+1)\Delta$ is the entropy of a random variable that is uniformly distributed over $2m+1$ cells of size Δ, which is the case for all the different noise types, i.e., $h(N_T|T) = \log(2m+1)\Delta$.

Proof 4.2. Observe that

$$H(Q_{side}(X+Z)|Z) = H(Q_{side}(X+Z) - Z|Z) = H(\hat{X}_{side}|Z). \tag{32}$$

Since X and Z determine $\hat{X}_{side}$, $H(\hat{X}_{side}|Z,X)=0$, and we have

$$H(\hat{X}_{side}|Z) = H(\hat{X}_{side}|Z) - H(\hat{X}_{side}|Z,X) = I(\hat{X}_{side};X|Z). \tag{33}$$

By the chain rule for mutual information,

$$I(\hat{X}_{side};X|Z) = I(X;\hat{X}_{side},Z) - I(X;Z). \tag{34}$$

However, Z is deterministically given by $\hat{X}_{side}$ and the cell type to which it belongs ($ct(Z)$); also, X is independent of Z and of $ct(Z)$. Hence $I(X;Z)=0$ and $I(X;\hat{X}_{side},Z) = I(X;\hat{X}_{side}|ct(Z))$. Combining with (32) and (33) we thus have

$$H(Q_{side}(X+Z)|Z) = I(X;\hat{X}_{side}|ct(Z)). \tag{35}$$

Since by Theorem 4.1 the statistical relation between X and $\hat{X}_{side}$, given $ct(Z)$, is the same as between X and Y_T given T, we may rewrite (35) as

$$R = I(X;X+N_T|T) = h(Y_T|T) - h(N_T|T) \tag{36}$$

where we used $h(Y_T|X,T) = h(N_T|T)$, a known relation for additive noise channels. ■

The additive nature of the quantization error allows us to easily compute the optimum post filters α_0 and α_1. Specifically, let $\sigma_0^2 = \frac{\Delta^2}{12}$ denote the variance of the "central" quantization noise (which like in ECDQ is uniform over the interval $(-\Delta/2, \Delta/2)$), and let $\sigma_1^2 = \frac{1}{2m}\sum_{t=1}^{2m} \mathrm{Var}(N_t)$ denote the average variance of the side quantization error. The optimum post filters, in the average MSE sense, are $\alpha_0 = \frac{\sigma_x^2}{\sigma_x^2+\sigma_0^2}$ and $\alpha_i = \frac{\sigma_x^2}{\sigma_x^2+\sigma_i^2}$, and the resulting distortions are

$$d_0 = \frac{\sigma_x^2\sigma_0^2}{\sigma_x^2+\sigma_0^2} \quad ; \quad d_1 = d_2 = \frac{\sigma_x^2\sigma_1^2}{\sigma_x^2+\sigma_1^2}. \tag{37}$$

(Some small improvement can be obtained by letting the α's depend on the type t.) From the structure of the PMDSQ, we were able to formulate the second moment of the different cell types, and found that the average second moment as a function of m is

$$\sigma_1^2 = \sigma_2^2 = \frac{[(2m^2+m+1)\Delta]^2}{12}\left[1 + O\left(\frac{1}{m}\right)\right] \tag{38}$$

where $O(1/m) = (20m^4 + 13m^3 - 1.5m^2 - 3m - 1)/(20m^5 + 30m^4 + 35m^3 + 22.5m^2 + 10m + 2.5)$. Thus, each pair (Δ, m) corresponds to a pair (σ_1^2, σ_0^2), which defines a pair of distortions (d_1, d_0).

It is interesting to observe that the equivalent additive noise channel of the ECD-PMDSQ is similar, in a sense, to the additive noise test channel of Ozarow [15]. Like the test channel, the two noises of the ECD-PMDSQ are negatively correlated, and it can be shown that, not only is each side error $Q_{side_i}(X+Z) - X - Z)$, $i=1,2$, statistically independent of the source (Theorem 4.1), but they are *jointly* statistically independent of X. Finally, for a given d_1, the variance of the side quantization error, and the "side" post filters that were found to be optimal for this encoding scheme, are identical to those in the test channel. Another interesting point to make is that replacing the central decoding rule $\alpha_0\{Q_0^{-1}(a_m^{-1}(i,j) - Z)\}$ of the scheme (Fig. 4), by the linear rule induced by Ozarow's test channel for the balanced case, $\alpha_0[0.5Q_{side1}(X+Z) + 0.5Q_{side2}(X+Z) - Z]$, results in almost the same d_0.

Using Theorems 4.1 and 4.2 we study the rate-distortion characteristics of the ECD-PMDSQ. We first derive a simple upper bound for the entropy rate (31). Let X^* denote a Gaussian random variable with the same variance as X, and N^* denotes a Gaussian random variable with variance σ_1^2.

$$\begin{aligned}
R_{ECD-PMDSQ} &= h(X+N_T|T) - \log(2m+1)\Delta \\
&= h(X^*+N^*) - D(X+N_T; X^*+N^*|T) - \log(2m+1)\Delta \quad (39) \\
&\leq h(X^*+N^*) - \log(2m+1)\Delta \quad (40) \\
&= \frac{1}{2}\log\left(2\pi e(\sigma_x^2+\sigma_1^2)\right) - \log(2m+1)\Delta \quad (41) \\
&= \frac{1}{2}\log\left(\frac{\sigma_x^2}{d_1}\right) + \frac{1}{2}\log\left(2\pi e(\sigma_1^2)\right) - \log(2m+1)\Delta \quad (42)
\end{aligned}$$

where (39) follows from an identity proved in Appendix C; the inequality in (40) is achieved by dropping the nonnegative term $D(X+N_T; X^*+N^*|T)$; equation (41) uses the entropy of the sum of two independent Gaussian sources, and the last equation uses (37). The bound becomes tight for a Gaussian source when the variance of the noise is much smaller than the source variance, i.e.

$$D(X+N_T; X^*+N^*|T) \to D(X; X^*) \quad (43)$$

as $\frac{\sigma_1^2}{\sigma_x^2} \to 0$.

4.3 Comparison with Theoretical Bounds

We are now in a position to compare $R_{ECD-PMDSQ}$ with the optimum theoretic solution [15, 23]. For a Gaussian source, we specialize Ozarow's formula [15] to the symmetric case, and express the minimum possible total rate in terms of the pair (σ_0^2, σ_1^2) which corresponds to (d_0, d_1) through (37):

$$TR_o \geq \log\left(\frac{\sigma_x^2}{d_1}\right) + \log\left(\frac{\frac{\sigma_1^2}{\sigma_0^2}}{2\sqrt{\frac{\sigma_1^2}{\sigma_0^2} - 1}}\right). \quad (44)$$

By combining (37),(38),(42) and (44), we obtain that the excess rate for a Gaussian source *at any resolution* is bounded by

$$2R_{ECD-PMDSQ} - TR_o \leq \log\left(\frac{2\pi e}{12}\right) + \log\sqrt{1 - O\left(\frac{1}{m}\right)} \leq \log\left(\frac{2\pi e}{12}\right). \quad (45)$$

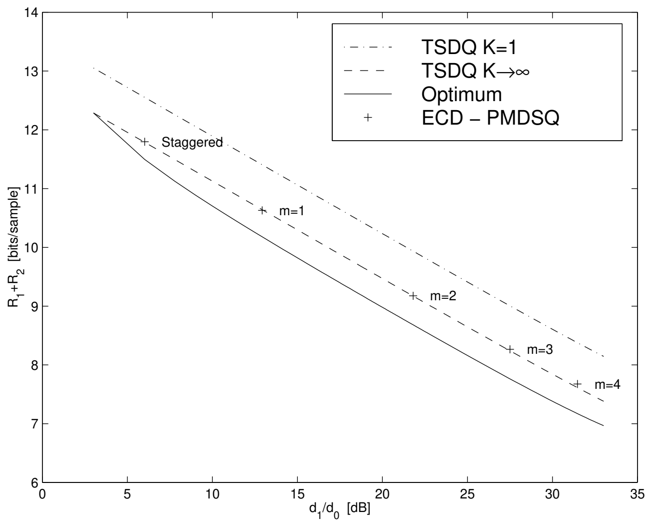

Figure 6: ECD-PMDSQ results compared to TSDQ and to the optimum, for a unit variance, memoryless Gaussian source and $d_0 = 10^{-4}$.

In other words, the redundancy of the ECD-PMDSQ over Ozarow's bound is at most two times the redundancy of a scalar quantizer. The results shown here apply for a Gaussian source at all resolutions and therefore this bound is not asymptotic. Figure 6 illustrates the performance of the ECD-PMDSQ for a Gaussian source in comparison to the optimum and to the TSDQ. The results are for $m = (1, 2, 3, 4)$. Since for $m = 0$ the PMDSQ is simply two identical "side" quantizers, we chose to substitute it with a "staggered quantizer", i.e., a uniform scalar quantizer for the first description, and the same quantizer offset by $\Delta/2$ for the second description.

As in Section 3, to evaluate the ECD-PMDSQ for a non-Gaussian source, we use the "Shannon outer bound" derived in [23]. We begin by defining σ_1^{2*} as $\frac{P(X)d_1}{p_x - d_1}$ and σ_0^{2*} as $\frac{P(X)d_0}{p_x - d_0}$. Thus, the outer bound in (5) on the total rate becomes

$$\begin{aligned} R_1 + R_2 = 2R \;\; &\geq \;\; \log\left(\frac{P(X)}{d_1}\right) + \log\left(\frac{\frac{\sigma_1^{2*}}{\sigma_0^{2*}}}{2\sqrt{\frac{\sigma_1^{2*}}{\sigma_0^{2*}} - 1}}\right) \\ &> \;\; \log\left(\frac{P(X)}{d_1}\right) + \log\left(\frac{\frac{\sigma_1^{2*}}{\sigma_0^{2*}}}{2\sqrt{\frac{\sigma_1^{2*}}{\sigma_0^{2*}}}}\right) \end{aligned} \qquad (46)$$

Notice that $\frac{\sigma_1^{2*}}{\sigma_0^{2*}} = \left(\frac{\sigma_x^2 - d_1}{\sigma_x^2 - d_0}\frac{P(X) - d_0}{P(X) - d_1}\right)\frac{\sigma_1^2}{\sigma_0^2} = B\frac{\sigma_1^2}{\sigma_0^2}$, and that B is greater than 1. As a result, any encoding scheme must satisfy

$$R_1 + R_2 = 2R > \log\left(\frac{P(X)}{d_1}\right) + \log\left(\frac{\frac{\sigma_1^2}{\sigma_0^2}}{2\sqrt{\frac{\sigma_1^2}{\sigma_0^2}}}\right) \qquad (47)$$

Note that by (6), this bound is achievable at high resolution conditions. Combining with (42), and using the relation between (m, Δ) and $(d_0, d_1, \sigma_0^2, \sigma_1^2)$, we obtain the following bound on the redundancy for a general source at any resolution:

$$2R_{ECD-PMDSQ} - TR_o \;\leq\; \log\left(\frac{2\pi e}{12}\right) + 2D(X;X^*) \tag{48}$$

where $D(X;X^*) = \frac{1}{2}\log(\sigma_x^2/P(X))$ is the divergence of the source from Gaussianity. At high resolution conditions, the correction term in (43) $D(X+N_t;X^*+N_t^*) \to D(X;X^*)$, $\forall t$, and therefore the redundancy in (48) can be tightly approximated as

$$2R_{ECD-PMDSQ} - TR_o \;=\; \log\left(\frac{2\pi e}{12}\right) + o(1) \tag{49}$$

where $o(1) \to 0$ as $d_1, d_2 \to 0$.

5 Summary and Discussion

This work has shown the advantages of using entropy coded dithered quantization in multiple description coding of a continuous memoryless source. The extension of our schemes to sources with memory involves incorporation of more general pre- and post-filters, and should follow along the lines of [26] or [4].

The TSDQ is a simple and direct derivative of the single description ECDQ and its equivalence to information theoretic test channels is similar to the single description case [26]. Although the encoders and decoders are constructed of simple and well known building blocks, the performance of this scheme is not optimal. The scheme exceeds the optimum by at most $0.5+\sum_{i=1}^{3}\frac{1}{2}\log 2\pi e G_{Ki}$ bits per sample for high resolution conditions, with some possible improvement using "refinement time sharing" or dependent dithering.

In contrast, the ECD-PMDSQ is based on a more complex building block - the periodic MDSQ. Unlike in [18], the analysis of the ECD-PMDSQ is not asymptotic, and holds for any resolution. We have shown that this scheme exceeds the optimum by only $\log(\frac{2\pi e}{12})$ bits per sample at high resolution and by at most $\log(\frac{2\pi e}{12}) + 2D(X;X^*)$ at lower resolutions. These two results resemble the results of the ECDQ for the single description, in the sense that the redundancy is factored by the number of the descriptions.

It is possible to extend the analysis of the dithered MDSQ to the *vector* case, along the lines of the recent work on vector-lattice MD coding by Vaishampayan, Servetto, and Sloane [17]. A preliminary analysis of the ECD-PMD**V**Q shows that its excess rate is at most $\frac{1}{2}\log 2\pi e\sqrt{G_K G_K^*} + D(X;X^*)$ bit per description, at any resolution, where G_K^* is the normalized second moment of a K dimensional *sphere*. To obtain this expression, we replace the uniform quantizer with step Δ in Section 4 by a K dimensional lattice quantizer with cells of volume V_0 and normalized second moment G_K. We then replace the expression for the average second moment of the side errors in (38) by

$$\sigma_1^2 = G_{side}[\bar{M}V_0]^{2/K}$$

where $\bar{M}$ denotes the number of central cells contained in the convex hull of a side cell (it is $m(2m+1)+1$ in the scalar case), and G_{side} denotes the normalized second moment of this convex hull; furthermore, we substitute in (31) $h(N_T|T) = \frac{1}{K}\log(MV_0)$ for the average entropy of the side error, where M denotes the number of central cells composing a side cell (it is $2m+1$ in the scalar case). Applying the same derivation with these substitutions, we get

$$2R - TR_o = \frac{1}{2}\log 2\pi e\sqrt{G_K G_{side}} + D(X;X^*) + \log(2\bar{M}^{1/K}/M^{2/K}).$$

It follows from the derivation in [16] that there exists lattice-based constructions with $G_{side} = G_K^*$, and for which the term $\log(2\bar{M}^{1/K}/M^{2/K})$ above vanishes.

Appendix A - Proof of Lemma 1

We first prove an auxiliary Lemma:

Lemma 2. *The joint density function of* $(\mathbf{X}, \mathbf{X}+\mathbf{N_1}, \mathbf{X}+\mathbf{N_2})$ *is equal to the joint density function of* $(\mathbf{X}, Q_1 - \mathbf{Z}_1, Q_2 - \mathbf{Z}_2)$

Proof of Lemma 2 From [25] we know that the conditional density of $Q_i - \mathbf{Z}_i - \mathbf{X_1}$ given $\mathbf{X}$ is equal to the density of $\mathbf{N}_i$ for all $\mathbf{X} = \mathbf{x}$. It is also easy to see that the variables form the following two Markov chains

$$\begin{gathered}(\mathbf{X}+\mathbf{N}_1) \leftrightarrow \mathbf{X} \leftrightarrow (\mathbf{X}+\mathbf{N}_2)\\ (Q_1 - \mathbf{Z}_1) \leftrightarrow \mathbf{X} \leftrightarrow (Q_2 - \mathbf{Z}_2)\end{gathered}$$

and therefore (by Bayes' rule) the two joint densities are equal and Lemma 2 is proven. ■

Proof of Lemma 1 To prove this Lemma we will, once more, use Markov chains to show that the two joint density functions are equal. The two Markov chains are

$$\begin{aligned}&(\mathbf{X}+\mathbf{N}_1, \mathbf{X}, \mathbf{X}+\mathbf{N}_2) \leftrightarrow \mathbf{U} \leftrightarrow (\mathbf{U}+\mathbf{N}_3)\\ &(Q_1(\mathbf{X}+\mathbf{Z_1}) - \mathbf{Z}_1, \mathbf{X}, Q_2(\mathbf{X}+\mathbf{Z_2}) - \mathbf{Z}_2)) \leftrightarrow \mathbf{W} \leftrightarrow (Q_3(\mathbf{W}+\mathbf{Z_3}) - \mathbf{Z}_3)\end{aligned} \tag{50}$$

These two chains are Markovian due to the nature of dithered quantization, and the fact that $\mathbf{U}$ is a function of $\mathbf{X}+\mathbf{N}_1, \mathbf{X}$ and $\mathbf{X}+\mathbf{N}_2$ while $\mathbf{W}$ is a function of $\mathbf{X}, Q_1 - \mathbf{Z}_1$ and $Q_2 - \mathbf{Z}_2$. Taking Lemma 2 into account, we see that $\mathbf{U}$ and $\mathbf{W}$ have the same density function, and therefore the right hand sides of the chains are also equivalent in distribution. Thus, the entire chains are equivalent, and Lemma 1 is proven.■

Appendix B - Cell Types of the PMDSQ

The structure of the PMDSQ is the equivalent of the uniform scalar quantizer for the single description. We wish to clarify the terms "side" cell, periodicity and "spread" in terms of the PMDSQ. We limit the discussion to an even m. For $1 \leq n \leq 2m+1$, let

$$\begin{aligned} A_n &= (1, 1+m, 1+2m, \cdots, 1+(n-1)m) \\ B_n &= (1, 2+m, 3+2m, \cdots, n+(n-1)m) \end{aligned}$$

and let U_n denote a vector of n ones. We now wish to show the different types of "side" cells. Each type of "side" cell is a union of "central" cells. There are $2m$ types of "side" cells for each "side" quantizer. The types appear in a periodic manner, and a side cell of type t will reappear every $2m(2m+1)$ "central" cells. We arbitrarily assume that cell type 1's first component is central cell 1. We use a comma to denote concatenation of vectors and thus the $2m$ "side" cells are

$$\begin{array}{llll}
Type\,1 & :(A_{2m} & ,(2m^2+m+2)U_1 \quad -B_1) & +0 \\
Type\,2 & :(A_{2m-2} & ,(2m^2+m+2)U_3 \quad -B_3) & +(2m+1)U_{2m+1} \\
Type\,3 & :(A_{2m-4} & ,(2m^2+m+2)U_5 \quad -B_5) & +2(2m+1)U_{2m+1} \\
\cdots & & & \\
Type\,m & :(A_2 & ,(2m^2+m+2)U_{2m-1}-B_{2m-1}) & +(m-2)(2m+1)U_{2m+1} \\
Type\,m+1 & :(B_{2m} & ,(2m^2+m+2)U_1 \quad -A_1) & +((m-2)(2m+1)+m+1)U_{2m+1} \\
Type\,m+2 & :(B_{2m-2} & ,(2m^2+m+2)U_3 \quad -A_3) & +((m-1)(2m+1)+m+1)U_{2m+1} \\
\cdots & & & \\
Type\,2m & :(B_2 & ,(2m^2+m+2)U_{2m-1}-A_{2m-1}) & +((2m-2)(2m+1)+m+1)U_{2m+1}
\end{array}$$

As shown above, all "side" cells are built of $2m+1$ center cells and spread over $2m^2+m+1$ center cells.

Appendix C - An Identity

Lemma 3. *Let X and $N_1, \ldots N_{2m}$ be given random variables. Let X^* denote a Gaussian random variable with the same variance, σ_x^2, as X. Let N^* denote a Gaussian random variable with the same variance, σ_1^2, as the average variance of $N_1, \ldots N_{2m}$, i.e., σ_1 is the variance of N_T, where T is uniformly distributed over $\{1, \ldots, 2m\}$. Then*

$$h(X + N_T | T) = h(X^* + N^*) - D(X + N_T; X^* + N^* | T)$$

where $D(\cdot; \cdot | T)$ denotes conditional divergence given T.

Proof of Lemma 3

$$
\begin{aligned}
h(X+N_T|T) &= -\frac{1}{2m}\sum_{i=1}^{2m}\int f_{X+N_i}(\alpha)\log(f_{X+N_i}(\alpha)) \\
&= -\frac{1}{2m}\sum_{1=1}^{2m}\int f_{X+N_i}(\alpha)\log\frac{f_{X+N_i}(\alpha)}{f_{X^*+N^*}(\alpha)}\,d\alpha \\
&- \frac{1}{2m}\sum_{i=1}^{2m}\int f_{X+N_i}(\alpha)\log f_{X^*+N^*}(\alpha)\,d\alpha \\
&= -\frac{1}{2m}\sum_{i=1}^{2m}\int f_{X+N_i}(\alpha)\log\frac{1}{\sqrt{2\pi(\sigma_x^2+\sigma_1^2)}}e^{-\frac{\alpha^2}{2(\sigma_x^2+\sigma_1^2)}}\,d\alpha \\
&- D(X+N_T;X^*+N^*|T) \\
&= \frac{1}{2}\log(2\pi e(\sigma_x^2+\sigma_1^2)) - D(X+N_T;X^*+N^*|T) \\
&= h(X^*+N^*) - D(X+N_T;X^*+N^*|T). \qquad (51)
\end{aligned}
$$

6 Acknowledgments

We thank Meir Feder, Vivek Goyal, Jelena Kovacevic and Vinay Vaishampayan for interesting and helpful discussions.

References

[1] R. Ahlswede, *The rate-distortion region for multiple descriptions without excess rate*, IEEE Trans. Information Theory **IT-31** (Nov. 1985), 721–726.

[2] J.C. Batllo and V. A. Vaishampayan, *Asymptotic analysis of multiple-description transform codes*, IEEE Trans. Information Theory **IT-43** (March. 1997), 703–707.

[3] Yael Frank-Dayan, *Lattice-based dithered quantizers for multiple descriptions*, Master's thesis, Tel Aviv University, 1999.

[4] P. Dragotti, S. Servetto and Vetterli *Analysis of Optimal Filter Banks for Multiple Description Coding* , Proc. of Data Compression Conf., Snowbird, Utah, March 2000, pp. 323–332.

[5] W. H. R. Equitz and T. M. Cover, *Successive refinement of information*, IEEE Trans. Information Theory **IT-37** (Nov. 1991), 851–857.

[6] A. El Gamal and T. M. Cover, *Achievable rates for multiple descriptions*, IEEE Trans. Information Theory **IT-28** (Nov. 82), 851–857.

[7] F.W. Fu, R.W. Yeung, R. Zamir, *On the rate-distortion region for multiple descriptions*, IEEE Trans. Information Theory, revised Feb. 2001.

[8] Vivek K. Goyal and Jelena Kovacevic, *Optimal multiple description transform coding of Gaussian vectors*, Proc. of Data Compression Conf., Snowbird, Utah, March 1998., pp. 388–397.

[9] Vivek K. Goyal, Jelena Kovacevic and Martin Vetterli, *Quantized frame expansions as source-channel codes for erasure channels*, Proc. of Data Compression Conf., Snowbird, Utah, March 1999, pp. 326–335.

[10] H. Gish and J. N. Pierce, *Asymptotically efficient quantizing*, IEEE Trans. Inform. Theory, vol. IT-14, pp. 676–683, Sep. 1968.

[11] S.S. Hemami, *Reconstruction of compressed images and video for lossy packet networks*, Ph.D. thesis, Stanford University, Nov. 1994.

[12] N.S. Jayant and S.W. Christensen, *Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure*, IEEE Trans. Communications **COM-29** (Feb. 1981), 101–109.

[13] T. Linder, R. Zamir, and K. Zeger, *The multiple description rate region for high resolution source coding*, Proc. of Data Compression Conf., Snowbird, Utah, March 1998., pp. 149–158.

[14] T. Moriya, *Two-channel conjugate vector quantizer for noisy channel speech coding*, IEEE Journal of Selected Areas in Communications **10** (June 1992), 866–874.

[15] L. H. Ozarow, *On a source coding problem with two channels and three receivers*, Bell Sys. Tech. Jour. **59** (1980), 1909–1922.

[16] S. Servetto, V. Vaishampayan, and N. Sloane, *Multiple description lattice vector quantization*, Proc. Data Compression Conference, Snowbird, March 1999, pp. 13–22.

[17] V. Vaishampayan, N. Sloane, and S. Servetto, *Multiple description vector quantization with lattice codebooks: design and analysis*, IEEE Trans. Information Theory **IT-47** (July 2001), 1718–1734.

[18] V. A. Vaishampayan, *Design of entropy-constrained multiple-description scalar quantizers*, IEEE Trans. Information Theory **IT-40** (Jan. 1994), 245–251.

[19] ______, *Design of multiple description scalar quantizers*, IEEE Trans. Information Theory **IT-39** (May 1993), 821–834.

[20] V. A. Vaishampayan and J.C. Batllo, *Asymptotic analysis of multiple-description quantizers*, IEEE Trans. Information Theory **IT-44** (Jan. 1998), 278–284.

[21] V. A. Vaishampayan and S. John, *Balanced inter-frame multiple description video compression*, in Proc. IEEE ICIP, Oct. 1999, pp. 812–816.

[22] J. K. Wolf, A. D. Wyner, and J. Ziv, *Source coding for multiple descriptions*, Bell System Technical J. **59** (Oct. 1980), 1417–1426.

[23] R. Zamir, *Gaussian codes and Shannon bounds for multiple descriptions*, IEEE Trans. Information Theory (Nov. 1999), 2629–2635.

[24] R. Zamir and M. Feder, *On lattice quantization noise*, IEEE Trans. Information Theory (July 1996), 1152–1159.

[25] ———, *On universal quantization by randomized uniform / lattice quantizer*, IEEE Trans. Information Theory (March 1992), 428–436.

[26] ———, *Information rates of pre/post filtered dithered quantizers*, IEEE Trans. Information Theory (September 1996), 1340–1353.

[27] R. Zamir and R. Yeung, *Multilevel diversity coding via successive refinement*, International Symposium on Information Theory, Ulm, Germany, June 1997, p. 265.

List of Figures

Biographies

Yael Frank-Dayan: a Biography

Yael Frank-Dayan received her BSc. degree (cum laude) in electrical engineering from the Technion, Israel in 1995. In 2000 she received her MSc degree (cum laude) in electrical engineering from the Tel-Aviv University. From 1995 to 2001 she was with Radguard, where she specialized in networking security. She is currently a system architect for Atrica, Israel.

Ram Zamir: a Biography

Ram Zamir was born in Ramat-Gan, Israel in 1961. He received the B.Sc., M.Sc. (summa cum laude) and D.Sc. (with distinction) degrees from Tel-Aviv University, Israel, in 1983, 1991, and 1994, respectively, all in electrical engineering. In the years 1994 - 1996 he spent a post-doctoral period at Cornell University, Ithaca, NY, and at the University of California, Santa Barbara. Since 1996 he has been with the department of Elect. Eng. - Systems at Tel Aviv University. Dr. Zamir has been consulting in the areas of radar and communications, mainly in developing algorithms and in the design of signals, and has been teaching information theory, communications systems and communications circuits at Tel Aviv University.

Dr. Zamir received the Israel Ministry of Communications Award in 1993, and the Wolfson Post-Doctoral Research Award in 1994, and visited the Technical University of Budapest under a joint program of the Israeli and Hungarian academies of science in summer 1995. His research interests include information theory, communication and remote sensing systems, and signal processing.