Dither and Data Compression

Daniël W.E. Schobben, Rob A. Beuker and Werner Oomen

Abstract— **This contribution presents entropy analyseses for dithered and undithered quantized sources. Two methods are discussed that reduce the increase in entropy caused by the dither. The first method supplies the dither to the lossless encoding-decoding scheme. It is argued that this increases the complexity of the encoding-decoding scheme. A method to reduce this complexity is proposed. The second method is the usage of a dead-zone quantizer. A procedure for determining the optimal dead-zone width in mean-square sense is given.**

I. Introduction

When a signal is quantized coarsely, signal dependent quantization errors are introduced which can be perceptually annoying. Roberts [2] first used dither in a simple PCM video system to remove false contouring. A survey of the dithering technique is provided in [3]. The objective of this correspondence is to investigate the effect of dithering on the entropy of the quantized source. Very little has been published in this area, but we wish to point out the work in [7]. We will show that the use of dither can cause an increase in both entropy and mean square error (MSE). Therefore we will introduce methods to reduce both entropy and MSE.

This paper is organized as follows. First, in Section II an introduction to the dithering technique is given. In Section III the distortion and entropy are defined and the source distribution is modeled. It is argued that both entropy and distortion can increase when dither is used. Two ways of reducing the entropy in a dithered scheme will be considered. In Section IV it is shown that the entropy will increase by applying dither prior to quantization if the entropy coder has no knowledge of the dither values. The effect on the entropy when the dither values are known to the lossless encoding-decoding scheme is also discussed. Another way of reducing the increase in entropy using a dead-zone quantizer is discussed in Section V. Furthermore, a method for designing an optimal dead-zone quantizer is shown. Results of experiments with dither in subband coding and transform coding schemes are discussed in Section VI.

II. Dithered Quantizers

A subtractively dithered quantizer, as shown in Fig. 1, is used throughout this paper. Its transfer function is given by

$$\begin{aligned} \tilde{x} &= Q(x+d) - d\,, \\ Q(x) &= \Delta \left\lfloor \frac{x}{\Delta} + \frac{1}{2} \right\rfloor\,, \end{aligned} \tag{1}$$

where $\lfloor \cdot \rfloor$ indicates the floor operator and Δ represents the quantizer step size. Note that the quantizer Q is uniform and infinite. If the dither d is a white noise signal uniformly distributed on $[-\frac{\Delta}{2}, \frac{\Delta}{2})$, the quantization error is also white and uniformly distributed in this interval and is independent of the quantizer input [3]. Note that typically the dither is not transmitted to the decoder, but is regenerated there. If a non-subtractively dithered quantizer were used, in which the dither is not subtracted from the quantizer output, this would result in the same entropy, but in an increased distortion. Derivation of the distortion for a non-subtractively dithered quantizer is straightforward and is not discussed here.

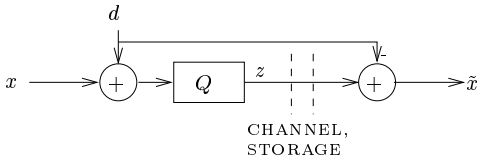

Fig. 1. Subtractively dithered quantizer.

III. Distortion and Entropy

The stochastic variables associated with the signals x, $\tilde{x}$, d, r and z will be denoted X, $\tilde{X}$, D, R and Z throughout. In the sequel, the MSE will be used as the distortion measure and is calculated from

$$\text{MSE} = E[(\tilde{X} - X)^2]\,, \tag{2}$$

where $E[.]$ denotes the expectation operator. In the undithered case, the MSE is smaller than $\frac{\Delta^2}{12}$ for input signals which are small compared to the quantizer step size Δ. For larger input signals the MSE is approximately equal to $\frac{\Delta^2}{12}$. When uniform dither is applied however, the MSE equals $\frac{\Delta^2}{12}$ independent of the input signal's distribution. Note that the MSE is an objective measure of distortion and does not fully reflect the perceived quality.

The entropy represents the lower bound of the bit rate that can be achieved by losslessly coding the quantized data. The entropy, in bit per sample (bps), is defined by

$$H(Z) = -\sum_{i=-\infty}^{\infty} p_z(i) \log_2(p_z(i))\,, \tag{3}$$

where $p_z(i)$ is the probability that the quantizer output equals $i\Delta$. For a uniform quantizer, this probability is given by

$$p_z(i) = \int_{(i-\frac{1}{2})\Delta}^{(i+\frac{1}{2})\Delta} p_r(r) dr\,, \tag{4}$$

with $r = x + d$. The probability density function (pdf) of R, $p_r(r)$, is equal to the convolution of the pdf's of X and D, given by $p_x(x)$ and $p_d(d)$, respectively. In this case, the entropy of the continuous random variable $R = X + D$ is greater than the entropy of X. Typically, the same holds for the entropy of quantizer output [6], i.e.,

$$H(Q(X+D)) \geq H(Q(X))\,. \tag{5}$$

The calculation of both entropy and distortion require a model for the source distribution. We use two pdf's: the uniform pdf and the generalized Gaussian pdf [4] (GG-pdf) and model the source as a random variable. The pdf of a video signal in a simple PCM system roughly resembles the uniform pdf. The generalized Gaussian pdf is used to model the AC-coefficients and differentially coded DC-coefficients in a transform coding system, such as for example DCT and subband transforms [5]. The uniform pfd is given by $p_x(x) = \Pi_w(x)$, with

$$\Pi_w(x) \triangleq \begin{cases} \frac{1}{w}, & -\frac{w}{2} < x \leq \frac{w}{2} \\ 0, & \text{otherwise} \end{cases}\,, \tag{6}$$

D.W.E. Schobben is with the Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

R.A. Beuker and A.W.J. Oomen are with the Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

where w defines the *width* of the pdf. Given the positive constants a and b, defined by

$$a = \frac{b\gamma}{2\Gamma(1/\gamma)}, \qquad b = \frac{1}{\sigma}\sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}}, \tag{7}$$

where $\Gamma(x)$ is the Gamma function [4], the GG-pdf is given by

$$p_x(x;\mu,\sigma^2,\gamma) = ae^{-[b|x-\mu|]^\gamma}, \tag{8}$$

where, μ, σ^2, γ are mean, variance and shape parameter of the distribution, respectively.

IV. Minimizing the Entropy – Method I

In [7] it was shown that the entropy can be reduced by supplying the dither to the lossless encoding-decoding scheme. The effect of this method on the entropy is discussed in this section for uniform and generalized gaussian input distributions.

The input signal is first modeled by a uniform pdf. The entropy is plotted in Fig. 2 as a function of the pdf-width w relative to the quantization stepsize Δ. Note that the entropy only depends on this ratio. The entropy of the output of the quantizer in the undithered and dithered case are indicated with $H(Q(X))$ and $H(Q(X+D))$, respectively.

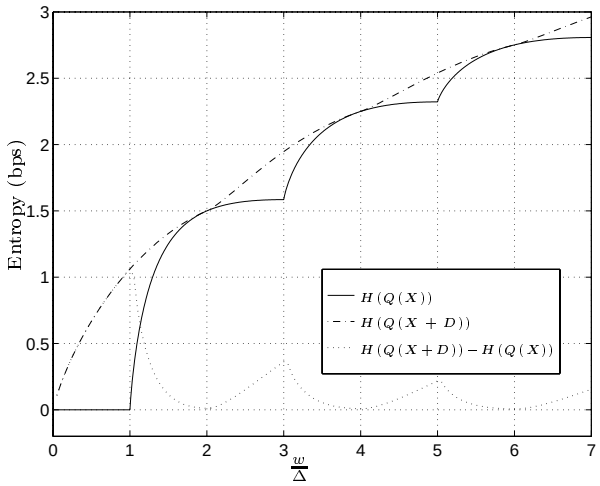

Fig. 2. Increase in entropy due to dithering a uniform source.

The difference between these two curves, the increase of the entropy due to dithering, is also shown. Fig. 2 shows that this increase can amount up to 1.1 bps when the width of the pdf equals Δ.

Using the current dither value in the lossless encoder and decoder can yield a reduction of the entropy. The entropy in case the current dither value is *not* used by the lossless coder equals $H(Q(X+D))$. In case the dither value *is* used the entropy is given by

$$H(Q(X+D)|D) = \int_{-\infty}^{\infty} H(Q(X+\nu))\, p_d(\nu) d\nu, \tag{9}$$

where the entropy $H(Q(X+\nu))$ defines the entropy of the output of the quantizer Q given the value of the dither ν. The entropy $H(Q(X+D)|D)$ is shown in Fig. 3. The increase in entropy due to dithering, given by $H(Q(X+D)|D) - H(Q(X))$, is also shown and amounts to 0.7 bps for a pdf-width of Δ. Note that the increase in entropy due to dither can now be either positive or negative, which implies that for some widths the coding efficiency can even improve when using dither.

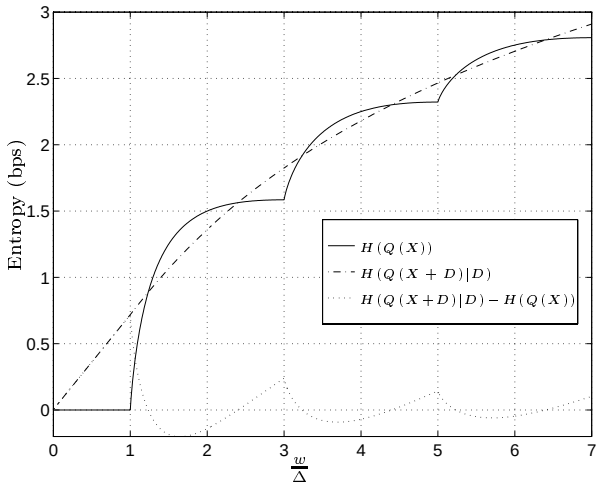

Fig. 3. Increase in entropy due to dithering a uniform source if the dither is supplied to the encoder and decoder.

The problem in a practical implementation when using the dither D in the entropy coder is the complexity increases. Usually, the dither is amplitude discrete, i.e. $d \in \{d_0, ..., d_{N-1}\}$ with N large, say 2^8. The lossless encoder and decoder need to keep a table containing estimates of the probabilities of $Q(X+d)$ for each value of the dither value d, i.e., N tables in total.

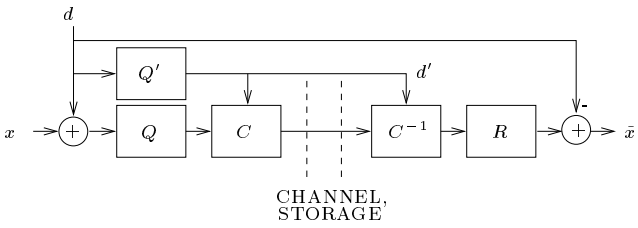

Fig. 4. A dithered source coding system, the dither is partly known to the lossless coding scheme.

A solution is to use a coarsely quantized version of the dither instead of the true value of d, see Fig. 4. The dither is added to the quantizer input, the result is quantized (Q) and entropy coded (C). The entropy coder has limited knowledge of the used dither value, given by $d' = Q'(d)$. The decoder performs the reverse process. The p-level quantized dither $d' = Q'(d)$ is an element of $\{d'_0, ..., d'_{p-1}\}$. The entropy of the quantized signal now equals

$$H(Q(X+D)|Q'(D)) = \sum_{i=0}^{p-1} H(Q(X+D)|Q'(D) = d'_i)\, p_{d'}(d'_i), \tag{10}$$

where

$$p_{d'}(d'_i) = \int_{i\frac{\Delta}{p}-\frac{\Delta}{2}}^{(i+1)\frac{\Delta}{p}-\frac{\Delta}{2}} p_d(\nu) d\nu. \tag{11}$$

The relative entropy loss caused by applying only a 2, 4, or 8 level quantized version of the dither to the coder instead of the full precision dither is shown in Fig. 5 for the uniform source distribution. In the legend of this figure, these levels are denoted as $Q'_1(D), Q'_2(D), Q'_3(D)$ respectively. When the dither is quantized to 2, 4 or 8 levels respectively, only 40%, 20% and 15% of the obtainable coding improvement is lost. Note that $Q'_1(d)$ corresponds to the sign of the dither.

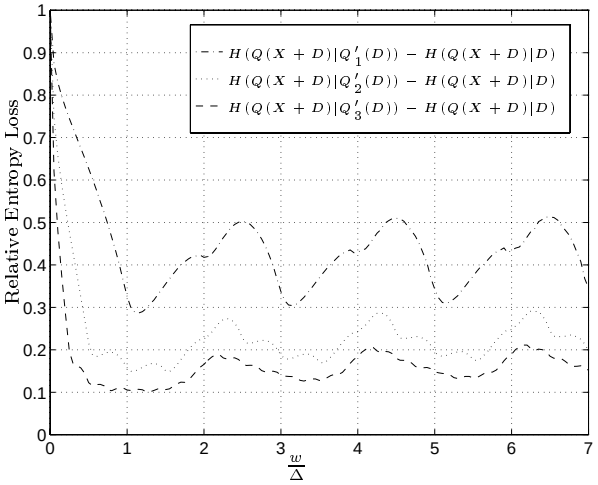

Fig. 5. Loss in entropy due to the quantization of the dither, relative to $H(Q(X+D)) - H(Q(X+D)|D)$ (see text).

The increase in entropy due to dithering is plotted in Fig. 6(a) (solid curve) for a Laplacian pdf ($\gamma = 1$) as a function of $\frac{\sigma}{\Delta}$, where σ is the standard deviation of the distribution. Note that the entropy only depends on the ratio $\frac{\sigma}{\Delta}$. The entropy loss amounts up to 0.52 bit per sample. Again, applying the dither to the lossless encoder and decoder reduces this loss significantly (dashed curve). Even more important than the absolute loss is the relative entropy loss. This is the increase in entropy due to dithering relative to the undithered entropy, as is shown in Fig. 6(b). This figure shows that the entropy increases excessively for $\Delta \gg \sigma$. When $\frac{\sigma}{\Delta} < 0.3$, this increase in entropy is more than 100%. When the dither is supplied to the encoder and the decoder, the relative entropy loss is decreased significantly, but it is still very large for $\frac{\sigma}{\Delta} < 0.3$. This is important when using dither in subband coding or in transform coding schemes, since most of the transformed data is no longer quantized to zero. A method that reduces the entropy significantly for small $\frac{\sigma}{\Delta}$ is discussed in the next section.

V. Minimizing the Entropy – Method II

Another way of reducing the entropy is the usage of a dead-zone quantizer [1] instead of a uniform quantizer. The transfer function of the dead-zone quantizer is depicted in Fig. 7. Note that the resulting quantization error will no longer be fully uncorrelated with the input signal.

In Fig. 8, the entropy and MSE/σ^2 are plotted as a function of $\frac{\sigma}{\Delta}$ for a Laplacian input and for the dead-zone widths $\{\Delta, \frac{3}{2}\Delta, 2\Delta, \frac{5}{2}\Delta, 3\Delta\}$. The relevant plots are indicated with 'zero subtraction'. The distortion increases monotonically with the dead-zone width and the entropy decreases monotonically with the dead-zone width. The entropy is also plotted for the case that the dither is supplied to the lossless coder. Fig. 8(b) shows that applying the dither to the coder only reduces the entropy significantly if a uniform quantizer is used.

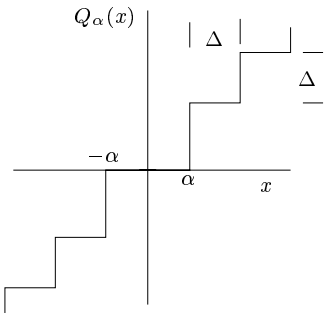

Fig. 7. Dead zone quantizer.

Fig. 8(a) shows that the MSE can become even larger than the signal power σ^2 for $\Delta > 3\sigma$. The distortion can be limited by *not* subtracting the dither from the quantizer output in case the quantizer output equals zero. Signals with a small standard deviation with respect to the quantizer step size are then quantized to zero. The resulting noise power will approximately equal the signal power which is smaller than $\frac{\Delta^2}{12}$. This is indicated with 'no zero subtraction' in Fig. 8(a). The figure shows that the MSE no longer exceeds the signal power for the deadzone quantizers considered.

The rate and distortion plots of Fig. 8 are combined to yield rate-distortion curves of Fig. 9(a). In these rate-distortion plots, the dither is not supplied to the encoder and decoder, and the dither is not subtracted when the quantizer output equals zero. Curves for 31 dead-zones widths in between Δ and 3Δ are shown in Fig. 9(a). In Fig. 9(b), some of these curves are magnified. It follows from this figure that the curves corresponding with different dead-zones cross each other. Therefore, the optimal dead-zone width depends on the entropy to be achieved. The dead-zone width which yields the lowest possible distortion at a given entropy was extracted from Fig. 9(a). This dependency is plotted in Fig. 10 for $\gamma = \frac{1}{2}, 1, \frac{3}{2}$. Note that $\gamma = 1$ corresponds to a Laplacian source.

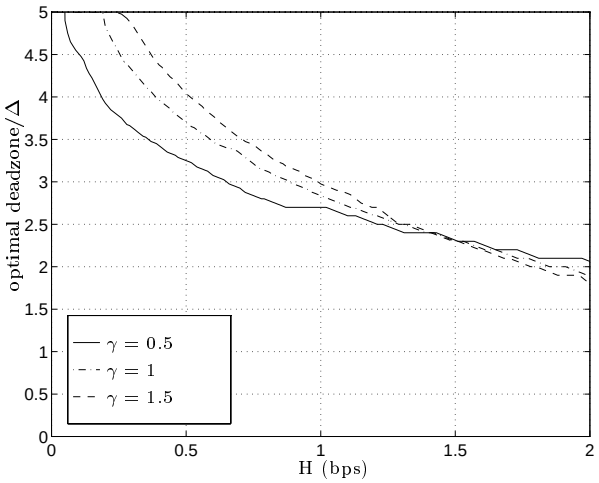

Fig. 10. Optimal dead-zone as a function of the entropy for $\gamma = \frac{1}{2}, 1, \frac{3}{2}$.

The optimal quantizer step size and dead-zone width can now be obtained in the following way. First the standard deviation (σ) and the peakedness (γ) of the source distribution are estimated [5]. Next, the optimal dead-zone width corresponding to the desired bit rate is read from Fig. 10. Then, the ratio $\frac{\sigma}{\Delta}$ is read from the solid curve in Fig. 8(b) corresponding with this dead-zone width and for the desired bit rate. Since the standard deviation is readily estimated, the quantizer step size can now be calculated from this ratio.

VI. Experiments

In [8] experimental results were presented. In the first place, mprovements are reported when dithering the first sub-band in a four band audio sub-band system. In addition, noise shaping was used to shape the resulting flat noise floor conforming to a psychoacoustical hearing curve. Secondly, experiments have been done with the usage of dither in a video DCT system. Dead-zone quantizers were used to reduce the entropy, and the dither was not subtracted for zero quantizer output to reduce the distortion. No significant improvements in perceptual quality are reported however when using dither in this application.

VII. Conclusions

Dithering is a technique which renders the quantization error white, uniformly distributed and independent of the input signal. Care must be taken however when dithering signals with a small standard deviation with respect to the quantizer step size. Both entropy and distortion can increase excessively in this case. This increase in entropy can be reduced by using the dither values in the lossless encoder and decoder. Furthermore a dead-zone quantizer can be used. In this case the resulting error will however no longer be fully uncorrelated with the input signal. The increase in distortion, which occurs for sources with a small standard deviation compared to the quantizer step size, can be reduced by not subtracting the dither when the quantizer output equals zero. A method for the calculation of the optimal quantizer step size and dead-zone width is presented for a dithered quantizer.

Acknowledgement

The authors would like to thank Jean Ritzerfeld and Piet Sommen of the Eindhoven University of Technology for their support.

References

[1] N.S. Jayant and P. Noll. *Digital Coding of Waveforms.* Englewood Cliffs, New Jersey: Prentice-Hall, 1984.

[2] L.G. Roberts. Picture coding using pseudo-random noise. *IRE Trans. Information Theory*, 8:145–154, Feb. 1962.

[3] S.P. Lipshitz et al. Quantization and dither: A theoretical survey. *J. Audio Eng. Soc.*, 40:355–375, May 1992.

[4] M. Abramowith and eds I.A. Stegun. *Handbook of Mathematical Functions.* New York: Dover Publications, Inc., 1965.

[5] K. Sharifi and A. Leon-Garcia. Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video. *IEEE Trans. Circuits Syst. Video Techn.*, 5:52–56, Febr. 1995.

[6] T.M. Cover and J.A. Thomas. *Elements of Information Theory.* New York: John Wiley & sons, 2nd edition, 1991. Use $h(x+\nu) \geq h(x)$, with x independent of ν, combined with Theorem 16.6.3 on page 496 to show that $H(Q(x+\nu)) \geq H(Q(x))$.

[7] R. Zamir and M. Feder. Rate distortion performance in coding band-limited sources by sampling and dithered quantization. *IEEE Trans. on Information Theory*, 41:141–154, Jan. 1995.

[8] D.W.E. Schobben. Dither and data compression. *M.S. thesis, Tech. University of Eindhoven*, Sept. 1995.

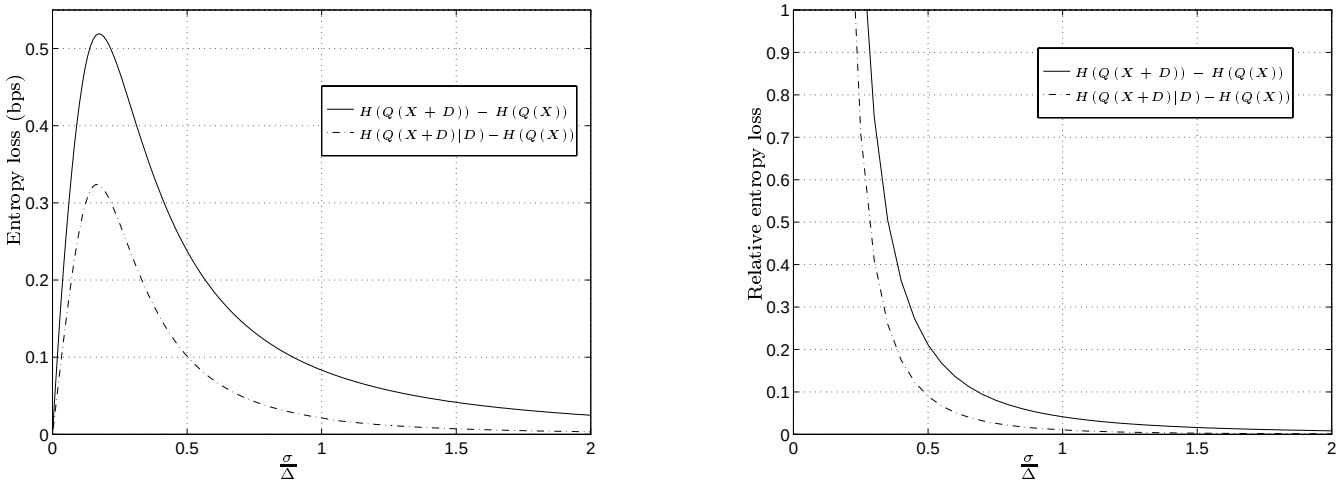

Fig. 6. (a) Entropy loss due to dithering, Laplacian source, (b) same, relative to the entropy when no dither was supplied.

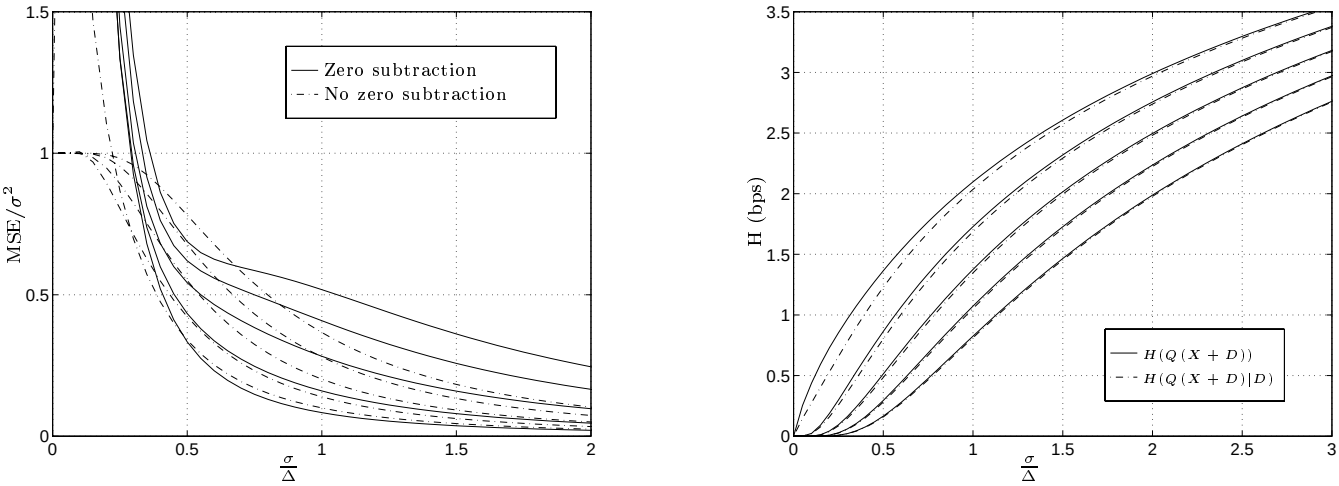

Fig. 8. (a) Distortion and (b) Rate of a Laplacian source for dead-zone widths $\{\Delta, \frac{3}{2}\Delta, 2\Delta, \frac{5}{2}\Delta, 3\Delta\}$.

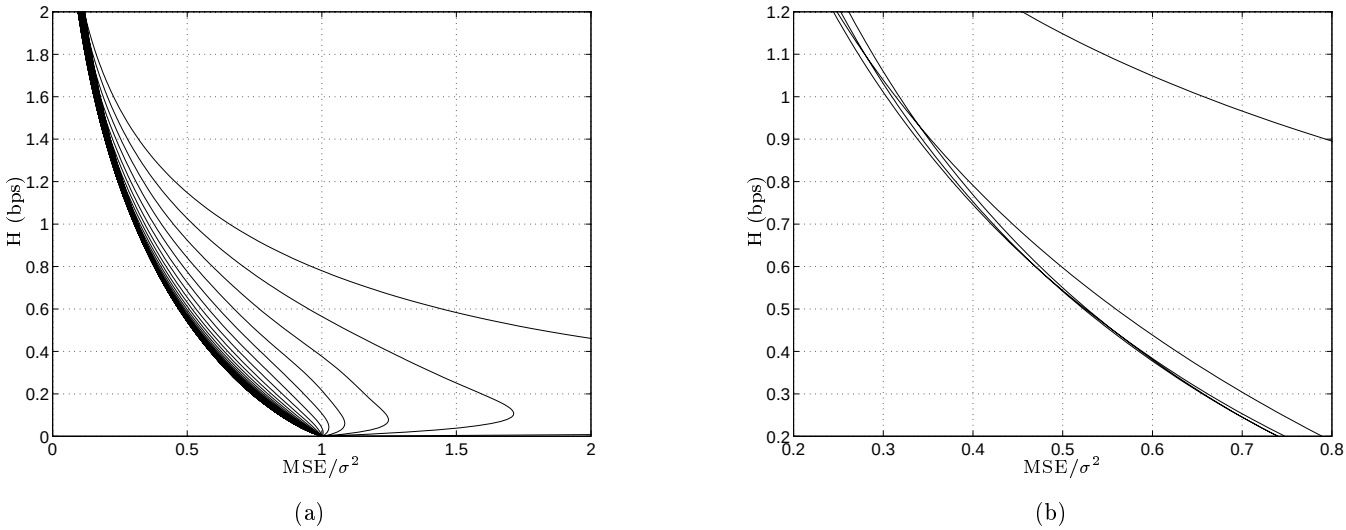

Fig. 9. (a) Rate-Distortion of a Laplacian source for several dead-zone widths and (b) magnification of plot (a).