# Low Resolution Scalar Quantization for Gaussian and Laplacian Sources with Absolute and Squared Error Distortion Measures

Daniel Marco and David L. Neuhoff [*]

January 7, 2006

**Abstract**

This report considers low resolution scalar quantization. Specifically, it considers entropy-constrained scalar quantization for memoryless Gaussian and Laplacian sources with both squared and absolute error distortion measures. The slope of the operational rate-distortion functions of scalar quantization for these sources and distortion measures is found. It is shown that in three of the four cases this slope equals the slope of the corresponding Shannon rate-distortion function, which implies that asymptotic low resolution scalar quantization with entropy coding is an optimal coding technique for these three cases. For the case of a Gaussian source and absolute error distortion measure, however, the slope at rate equal zero of the operational-rate distortion function of scalar quantization is infinite, and hence does not match the slope of the corresponding Shannon rate-distortion function. Consequently, scalar quantization is not an optimal coding technique for Gaussian sources and absolute error distortion measure. The results are obtained via analysis of uniform and binary scalar quantizers, which shows that in low resolution their operational rate-distortion functions, in all four cases, are the same as the corresponding operational rate-distortion functions of scalar quantization in general. Lastly, the slope of the Shannon rate-distortion function (the function itself is not known) at rate equal zero is found for a Laplacian source and squared error distortion measure.

## 1 Introduction

In this report we examine the rate-distortion performance of scalar quantization in the low resolution regime where rate is small. While there are well known, asymptotically accurate, closed

[*] This work was supported by NSF Grant ANI-0112801. It was also supported by the Center for the Mathematics of Information at California Institute of Technology. Portions of this report have been submitted to the IEEE Transactions on Information Theory.

form formulas for the rate-distortion performance of a variety of quantization schemes in the high resolution, i.e. high encoding rate, regime, there is a shortage of similar formulas for the low resolution, i.e. low rate, regime. As a step in this direction, we focus on scalar quantization with entropy coding. For several memoryless sources, namely, exponential, Laplacian and uniform, the low rate performance of such codes was found in or derives directly from previous work giving closed form expressions for the operational rate-distortion function [1, 2].

The main contribution of this report is the derivation of a low rate performance for a memoryless Gaussian source with squared and absolute error distortion measures, for which no closed form expressions exist or seem feasible. Furthermore, the method used to derive the Gaussian works also for the Laplacian source, and so we provide derivations pertaining to this source as well. Notice, however, that the Laplacian case has been already derived in [1] in a different way.

To determine the low resolution performance, we analyze the operational rate-distortion function, $R(D)$, of entropy-constrained scalar quantization in the low rate region. $R(D)$ is defined to be the least output entropy of any scalar quantizer with mean-squared error $D$ or less. As it determines the optimal rate-distortion performance of this kind of quantization, it is important to understand how $R(D)$ depends on the source probability density function (pdf) (we consider memoryless, stationary Gaussian and Laplacian sources, which are completely characterized by their first-order pdfs) and how it compares to the Shannon rate-distortion function. For example, the performance of conventional transform coding, which consists of an orthogonal transform followed by a scalar quantizer for each component of the transformed source vector, depends critically on the allocation of rate to component scalar quantizers, and the optimal rate allocation is determined by the operational rate-distortion functions of the components [3, p. 227].

While $R(D)$ can be determined numerically with various quantizer optimization algorithms [1], [4] - [11], closed form formulas for $R(D)$ are known only for the exponential [1] and uniform [2] sources. A general closed form expression is known only for the high resolution, i.e. high rate, region [12], namely,

$$R(D) = h - \frac{1}{2}\log(12\,D) + o_{D\to 0}\ , \qquad (1)$$

where $h = -\int_{-\infty}^{\infty} f(x)\log_2 f(x)\,dx$ is the differential entropy of the quantized source, $f$ is its pdf, and $o_{D\to x}$ denotes a quantity that goes to zero as $D \to x$.

In the low resolution region, it is well understood that $R(D)$ approaches zero as $D$ approaches $D_{\max}$, where $D_{\max}$ is the minimum distortion attainable with zero rate. Accordingly, one obtains a first-order approximation to $R(D)$ in this region by finding the slope of $R(D)$ at $D = D_{\max}$, namely,

$$R(D) = s\Big(1 - \frac{D}{D_{\max}}\Big)\left[1 + o_{D \to D_{\max}}\right] \,, \tag{2}$$

where $s$ is a slope determining factor, namely, it is the magnitude of the slope with respect to normalized distortion, and where the assumption throughout this report is that in $o_{D \to D_{\max}}$, $D$ goes to $D_{\max}$ from below.

The parametric formula of Sullivan [1] for the exponential source and of Gyorgy and Linder [2] for the uniform source imply $s = 0$ and $s = \infty$, respectively, for both squared and absolute error distortion measures. Likewise, $s = 0$ for the Laplacian source with both distortion measures [1]. Whereas these calculations are enabled by the special tractability of exponential and uniform pdfs, the principal result of this report uses, primarily, tail behavior to find the slope determining factors for a Gaussian source with both squared and absolute error distortion measures, which are $\frac{\log_2 e}{2}$ and $\infty$, respectively.

As a result, for the aforementioned sources, whose low resolution slope determining factors are summarized in Table 1, we now have simple, accurate approximations to the performance of optimal entropy-constrained scalar quantization in both the high and low resolution regions, as illustrated in Figure 1.

| | exponential | Laplacian | uniform | Gaussian |
|---|---|---|---|---|
| squared error | 0 | 0 | $\infty$ | $\frac{\log_2 e}{2}$ |
| absolute error | 1 | $\log_2 e$ | $\infty$ | $\infty$ |

Table 1: Slope determining factor $s$ of the operational rate-distortion function $R(D)$ at $D = D_{\max}$.

It is interesting to compare the low resolution behavior of $R(D)$ for a given source pdf to that of the Shannon rate-distortion function, denoted $\mathcal{R}(D)$, of a stationary memoryless source with the same pdf. Since $\mathcal{R}(D)$ represents the best performance attainable by any quantization technique, it must be that $\mathcal{R}(D) \leq R(D)$. It follows from this and the fact that, like $R(D)$, $\mathcal{R}(D) \to 0$ as $D \to D_{\max}$, that the magnitude of the slope of $\mathcal{R}(D)$ at $D = D_{\max}$ is no larger than that of $R(D)$.

We observe from Table 1 that since the slopes of $R(D)$ at $D = D_{\max}$ equal 0 for exponential and Laplacian sources with squared error, they must equal the slopes of the corresponding Shannon rate-distortion functions (because the latter's magnitudes could be no larger). Furthermore, for a Gaussian source with squared error, and for Laplacian and exponential sources with absolute error,

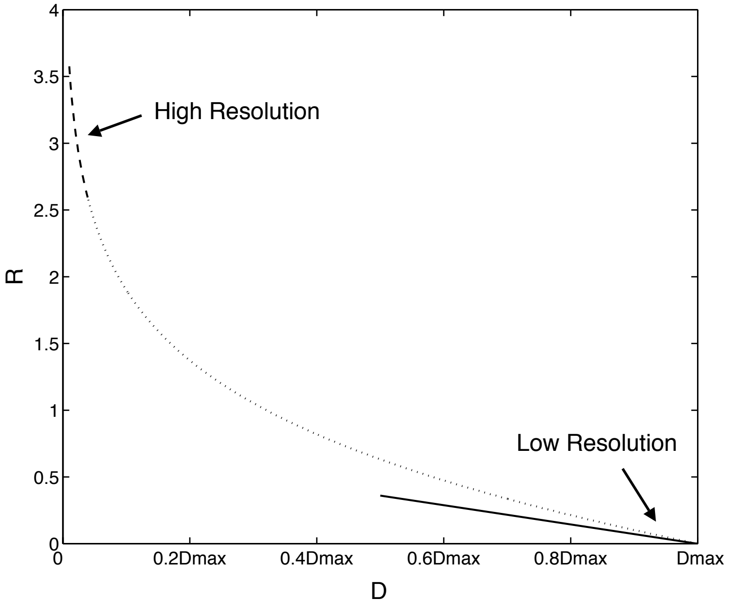

Figure 1: The dotted line is a qualitative representation of the operational rate-distortion function of scalar quantization for some source. The dashed line indicates the section of the curve that is well described by (1). The solid line, which shows the tangent of the curve at $D = D_{\max}$, indicates the low resolution performance given by (2).

the Shannon rate-distortion functions are known [13, 14][1], and their slopes match the corresponding slopes of $R(D)$. Thus, in low resolution, scalar quantization for these sources and distortion measures is asymptotically optimal, i.e. as good as any quantization technique — vector or otherwise. For a uniform source with both distortion measures, and for a Gaussian source with absolute error, the slopes of $R(D)$ at $D = D_{\max}$ are infinite, whereas the slopes of the corresponding Shannon rate-distortion functions must be finite (because they are convex). Thus, for these sources and distortion measures, low resolution scalar quantization is far from optimal.

We conclude this introduction with a few additional comments. To derive the low resolution slope for a Gaussian and Laplacian sources, we focus on uniform threshold quantizers with infinitely many cells, optimal reconstruction levels, and increasingly large step sizes $\Delta$. While it is easy to see that under ordinary conditions, distortion $D \approx D_{\max}$ and quantizer output entropy $H \approx 0$ when $\Delta$ is large, the slope at which $H$ approaches 0 as $D \to D_{\max}$ is not obvious. Nevertheless, we find

[1] Reference [14] makes a small error in applying its Theorem 2 to compute the rate-distortion function, with respect to absolute error, of an exponential source. Specifically, for $f(x) = \alpha e^{-\alpha x}$, $\alpha > 0$, a correct application of this theorem yields $\mathcal{R}(D) = \alpha D - \ln(e^{\alpha D} - 1) - \ln 2$, rather than the formula given in (24) of [14].

accurate approximations to $D$ and $H$ from which the low resolution slope can be straightforwardly determined.

Whereas the high resolution formula (1) is based on the fact that the source density can be approximated by a constant on most sufficiently small cells, the low resolution formula (2) is based on the fact that when the cells are large, the tail of the source probability density decays sufficiently quickly that only a few cells contribute materially to distortion and rate. We will show precisely which cells dominate the distortion and the entropy.

We also analyze binary quantization and show that it has low resolution performance characterized by the same slope. Thus, it, too, is asymptotically as good in the low resolution region as any quantization technique for Gaussian and Laplacian sources with both distortion measures.

As Laplacian and Gaussian are the two most commonly cited models for transform components (usually called coefficients) [15, pp. 215-218], [16, p. 564], the fact that scalar quantization is asymptotically as good for them as any type of quantization in the low resolution region has interesting ramifications for transform coding. In particular, in situations where transform coding is most effective, a sizable fraction of the coefficients must be coded at low rate. For such coefficients, simple scalar quantization is essentially as effective as any more sophisticated quantization technique. In contrast, to encode the coefficients that must be encoded with high resolution, scalar quantization requires approximately one quarter bit per sample more than optimal vector quantization.

We note that scalar quantization with fixed-rate coding does not attain the rate-distortion performance described in (2). This is because with fixed-rate coding, the smallest nonzero rate is at least 1, which implies that for any $D < D_{\max}$, the least rate of any fixed-rate scalar quantizer with mean-squared error $D$ or less is at least 1. Consequently, the discussion throughout this report is relevant to variable-rate coding, i.e. scalar quantization with entropy coding.

Lastly, we comment that this report includes derivations and results contained in two submissions to the IEEE Transaction on Information Theory [17, 18]. In [17] the slope in the case of a Gaussian source and squared-error distortion measure is derived, and in [18] the slope in the case of a Gaussian source and absolute error is found. In addition to these results this report derives the slope for a Laplacian source with both distortion measures. The slopes for the Laplacian case were not included in the submitted papers because they were previously derived by Sullivan in [1]. However, Sullivan found these slopes using methods enabled by the tractability of exponential densities, whereas we derive them here using, essentially, the method used for the Gaussian and squared-error case in [17]. We do this in order to demonstrate the generality of the method, which, we believe, could be used to derive the slope for other density/distortion measure combinations. In comparison to [17], which considered only a Gaussian source and squared-error, to generalize the results to absolute error and Laplacian sources we had to use Lemmas 10, 11 and A1, and Theorem 14, which were not needed in [17]. Finally, [18], which focused on the Gaussian and absolute error case, finds the slope to be infinite. This enabled a derivation tailored to this specific case, that is simpler than that given here, which applies to several cases. For example, there is no need to focus on uniform or binary quantizers. Instead, it directly shows that the low rate slope for optimal quantizers of any form is infinite.

The remainder of the report is organized as follows. Section 2 provides some background and introduces notation. Section 3 derives expressions for the asymptotic entropy of uniform quantizers for both Laplacian and Gaussian sources. In Section 4 expressions for the asymptotic distortion of uniform quantizers is provided for both sources and both distortion measures. Section 5 provides asymptotic operational rate-distortion analysis for uniform quantizers and general quantizers. In Section 6 binary quantization is considered. Section 7 offers concluding remarks. Finally, the Appendix contains proofs of certain lemmas.

## 2 Background and Notation

An infinite-level uniform threshold scalar quantizer (UTQ) with step size $\Delta$ and offset $0 \leq \alpha < 1$ is a scalar quantizer with partition having cells $S_k = [(k-\alpha)\Delta, (k+1-\alpha)\Delta)$, $k \in \mathbb{Z}$, along with reconstruction levels $r_k \in S_k$, $k \in \mathbb{Z}$. Its quantization rule is $q(x) = r_k$, when $x \in S_k$. The offset $\alpha$ indicates the fraction of cell $S_0$ that lies to the left of the origin. For example, when $\alpha = 1/2$, cell $S_0$ is centered at the origin, whereas when $\alpha = 0$, cell $S_0$ begins at the origin. Let $\bar{\alpha} \triangleq 1 - \alpha$.

We assume throughout that the source to be quantized is stationary, memoryless Gaussian or Laplacian with mean zero and variance $\sigma^2$, denoted $\mathcal{N}(0,\sigma^2)$ or $\mathcal{L}(0,\sigma^2)$, respectively (ordinarily we do not mention stationarity or memorylessness). We will superscript quantities by $G$ or $L$, and subscript quantities by 1 or 2, to reflect the source and distortion measure used. Such superscripts and subscripts will be omitted when they are clear from context, or when the statement made holds for both sources or for both distortion measures. The (output) entropy of this quantizer is

$$H(\alpha, \Delta, \sigma^2) = -\sum_{k=-\infty}^{\infty} P_k \log P_k \ ,$$

where $P_k$ denotes the probability of the $k^{th}$ cell, and all logarithms in this report have base 2. For a Gaussian source

$$P_k = Q\Big((k-\alpha)\frac{\Delta}{\sigma}\Big) - Q\Big((k+1-\alpha)\frac{\Delta}{\sigma}\Big) \ , \tag{3}$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt$. We let $G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ denote the Gaussian density with zero mean and unit variance. For a Laplacian source

$$P_k = \begin{cases} \frac{L\left((k-\alpha)\frac{\Delta}{\sigma}\right) - L\left((k+1-\alpha)\frac{\Delta}{\sigma}\right)}{\sqrt{2}}, & k > 0 \\ 1 - \frac{L\left(\alpha\frac{\Delta}{\sigma}\right) + L\left(\bar{\alpha}\frac{\Delta}{\sigma}\right)}{\sqrt{2}}, & k = 0 \\ \frac{L\left((k+1-\alpha)\frac{\Delta}{\sigma}\right) - L\left((k-\alpha)\frac{\Delta}{\sigma}\right)}{\sqrt{2}}, & k < 0 \end{cases} \ , \tag{4}$$

where $L(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}$ denotes the Laplacian density with mean zero and unit variance.

The mean-squared and absolute error of this quantizer is

$$d_s(\alpha, \Delta, \sigma^2) = \int_{-\infty}^{\infty} |x - q(x)|^s f(x)\, dx \ ,$$

where $s \in \{1, 2\}$, and $f$ is the density of the source. In the case of squared error, we take the reconstruction levels to be the centroids of their respective cells; i.e., $r_k = \int_{S_k} x \frac{f(x)}{P_k}\, dx$. When considering absolute error, we take the reconstruction levels to be the medians of their respective cells; i.e., $r_k$ is such that

$$\int_{(k-\alpha)\Delta}^{r_k} f(x)\, dx = \int_{r_k}^{(k+1-\alpha)\Delta} f(x)\, dx \ . \tag{5}$$

As is well known, these choices minimize mean-squared error and absolute error, respectively, for a given partition.

The operational rate-distortion function of infinite-level uniform threshold quantization is the function

$$R_{s,U,\sigma^2}(D) = \inf_{0 \le \alpha < 1, \Delta > 0 : d_s(\alpha,\Delta,\sigma^2) \le D} H(\alpha, \Delta, \sigma^2) \ , \tag{6}$$

where $s \in \{1, 2\}$ represents the distortion measure. This function specifies the least entropy of any such quantizer with distortion $D$ or less. We denote by $\mathcal{R}^G_{2,\sigma^2}(D), \mathcal{R}^L_{2,\sigma^2}(D), \mathcal{R}^G_{1,\sigma^2}(D)$, and $\mathcal{R}^L_{1,\sigma^2}(D)$, the Shannon rate-distortion functions of the appropriate source and distortion measure, where the sources have variance $\sigma^2$. Two of these rate-distortion functions are known. Specifically, $\mathcal{R}^G_{2,\sigma^2}(D) = \frac{1}{2}\log\frac{\sigma^2}{D}$ and $\mathcal{R}^L_{1,\sigma^2}(D) = \log\frac{\sqrt{2}D}{\sigma}$. We let $D_{\max}$ denote the minimum distortion attainable when the rate is zero. Specifically, for a source with variance $\sigma^2$ we have $D^G_{\max,2} = D^L_{\max,2} = \sigma^2$, $D^G_{\max,1} = \sqrt{\frac{2}{\pi}}\sigma$, and $D^L_{\max,1} = \frac{\sigma}{\sqrt{2}}$. If the source is normalized, i.e. it has unit variance, then we will write $D^*_{\max}$ to reflect this.

We set $\lambda$ to be the ratio $\Delta/\sigma$ and refer to it throughout as the *normalized step size.* We notice that $P_k$ depends only on $\alpha$ and $\lambda$, and for emphasis, we will sometimes denote it $P_k(\alpha, \lambda)$. Consequently, $H(\alpha, \Delta, \sigma^2) = H(\alpha, \Delta/\sigma, 1)$, depends only on $\alpha$ and $\lambda$ as well. Therefore, we will frequently use the notation $H(\alpha, \lambda)$. Similarly, $d_2(\alpha, \Delta, \sigma^2) = \sigma^2 d_2(\alpha, \Delta/\sigma, 1) = \sigma^2 d_2(\alpha, \lambda, 1)$, and $d_1(\alpha, \Delta, \sigma^2) = \sigma d_1(\alpha, \Delta/\sigma, 1) = \sigma d_1(\alpha, \lambda, 1)$. It follows from these remarks that $R_{2,U,\sigma^2}(D) = R_{2,U,1}\left(\frac{D}{\sigma^2}\right)$ and $R_{1,U,\sigma^2}(D) = R_{1,U,1}\left(\frac{D}{\sigma}\right)$ .

To find the slope of $R_{U,\sigma^2}(D)$ at $D = D_{\max}$ we need to consider what happens when $H(\alpha, \lambda) \to 0$ and $\sigma^2 d_2(\alpha, \lambda, 1) \to \sigma^2$ or $\sigma d_1(\alpha, \lambda, 1) \to D_{\max,1}$. We observe that in order for entropy to go to zero it is necessary and sufficient that $\alpha\lambda \to \infty$ and $\bar{\alpha}\lambda \to \infty$. Moreover, because of this, for sufficiently large values of $D$, it suffices to restrict $\alpha$ to be greater than 0 in the definition of $R_{s,U,\sigma^2}(D)$ in (6).

Before proceeding, we introduce notation and facts to be used later. Let the entropy function be defined as $\mathcal{H}(\ldots, z_{-1}, z_0, z_1, \ldots) = -\sum_{k=-\infty}^{\infty} z_k \log z_k$, where $0 < z_k \leq 1$ for all $k$, are a finite or countably infinite set of numbers that need not sum to one. Let $o_x$ denote a quantity that converges to zero as $x \to \infty$. More generally, let $o_{x\to x_o, y\to y_o}$ denote a quantity that converges to zero as $x \to x_o$ and $y \to y_o$, where it will be clear from context whether $x \nearrow x_o$, $x \searrow x_o$ or simply $x \to x_o$, and similarly for the variable $y$. If this quantity depends on parameters other than $x$ and $y$, its convergence to zero is uniform in such parameters. To keep notation short, we write $o_x$ instead of $o_{x\to\infty}$, when $x_o = \infty$, and we let $o_{x,y}$ denote $o_{x\to\infty, y\to\infty}$.

The following facts provide elementary bounds and approximations to the $Q$ function, and thus are relevant for Gaussian densities only.

**Fact G1:** $Q(x) \leq \sqrt{\frac{\pi}{2}}\, G(x)$, $x \geq 0$.

**Fact G2:** $Q(x) < \frac{1}{x}\, G(x)$, $x > 0$.

**Fact G3:** $Q(x) > \frac{1}{x}(1 - \frac{1}{x^2})\, G(x)$, $x > 0$.

**Fact G4:** $Q(x) > \begin{cases} \frac{1}{2x}\, G(x), & x \geq \sqrt{2} \\ Q(\sqrt{2}), & x < \sqrt{2} \end{cases}$ .

**Fact G5:** $Q(x) = \frac{1}{x}\, G(x)\left[1 + o_x\right]$, $x > 0$.

**Fact G6:** $Q((x+1)\lambda) = Q(x\lambda)\, o_\lambda$, $x \geq 0$; i.e. $\frac{Q((x+1)\lambda)}{Q(x\lambda)} \to 0$ as $\lambda \to \infty$, uniformly for $x \geq 0$.

**Fact G7:** For all sufficiently large $\lambda$, $Q\big((x+1)\lambda\big) < \frac{1}{2} Q(x\lambda)$ for all $x \geq 0$.

**Fact G8:** For all sufficiently large $\lambda$, $Q(x\lambda) - Q((x+1)\lambda) > \begin{cases} \frac{1}{4x\lambda}\, G(x\lambda), & x\lambda \geq \sqrt{2} \\ \frac{Q(\sqrt{2})}{2}, & 0 \leq x\lambda < \sqrt{2} \end{cases}$ .

Facts G1, G2 and G3 are demonstrated in [19, pp. 82-83]. Fact G4 truncates the lower bound of Fact G3. Fact G5 follows from Facts G2 and G3. Fact G6 is derived by upper bounding $\frac{Q((x+1)\lambda)}{Q(x\lambda)}$ using Facts G1 and G4 when $x\lambda < \sqrt{2}$, and using Facts G2 and G4 when $x\lambda \geq \sqrt{2}$. Fact G7 follows from Fact G6, and Fact G8 follows from Facts G4 and G7.

The next fact, which is relevant for Laplacian densities only, derives directly from properties of exponentials.

**Fact L1:** $L((x+1)\lambda) = L(x\lambda)\, o_\lambda$, $x \geq 0$; i.e. $\frac{L((x+1)\lambda)}{L(x\lambda)} \to 0$ as $\lambda \to \infty$, uniformly for $x \geq 0$.

Finally, we list facts that are relevant for both Gaussian and Laplacian densities, and where $f$ denotes either one of these densities, normalized to have zero mean and unit variance.

**Fact GL1:** $C(x) \overset{\Delta}{=} \int_x^\infty t\, f(t)\, dt\,;\;\; C^G(x) = G(x)\,;\;\; C^L(x) = \frac{xL(x)}{\sqrt{2}}\,[1 + o_x]$.

**Fact GL2:** $V(x) \overset{\Delta}{=} \int_x^\infty t^2\, f(t)\, dt\,; V^G(x) = x\, G(x) + Q(x) = x\, G(x)\,[1+o_x]\,; V^L(x) = \frac{x^2L(x)}{\sqrt{2}}\,[1+o_x]$.

**Fact GL3:** $C((x+1)\lambda) = C(x\lambda)\, o_\lambda$, $x \geq 0$; i.e. $\frac{C((x+1)\lambda)}{C(x\lambda)} \to 0$ as $\lambda \to \infty$, uniformly for $x \geq 0$.

**Fact GL4:** $V((x+1)\lambda) = V(x\lambda)\, o_\lambda$, $x \geq 0$; i.e. $\frac{V((x+1)\lambda)}{V(x\lambda)} \to 0$ as $\lambda \to \infty$, uniformly for $x \geq 0$.

Fact GL1, the first equality of Fact GL2 that considers $V^G$, and the part of Fact GL2 that considers $V^L$ derive from elementary integration. The second equality of Fact GL2 that considers $V^G$ follows from Fact G5. Fact GL3 (for both sources) and Fact GL4 for Laplacian source follow from Fact GL1 and simple manipulation of exponentials. Finally, Fact GL4 for a Gaussian source is derived using Facts GL2, G4 and G2. Specifically, for $x\lambda < \sqrt{2}$, Fact G4 is used to lower bound $V^G(x\lambda)$, and the fact that $x\lambda < \sqrt{2}$ is used to upper bound $V^G((x+1)\lambda)$. When $x\lambda \geq \sqrt{2}$, Fact G2 is used to upper bound $V^G((x+1)\lambda)$.

## 3 Asymptotic Entropy

We note that the output entropy of the UTQ does not depend on the distortion measure. Thus, the expressions provided in this section hold for both distortion measures.

We begin with several lemmas (most of which are proven in the Appendix) that lead to the main result of this section, a low resolution approximation for entropy for both Gaussian and Laplacian sources.

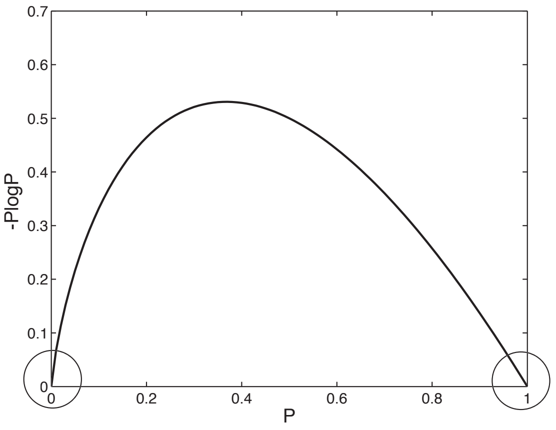

Figure 2: The entropy function, $-p \log p$.

**Lemma 1** *When a UTQ with offset $\alpha$, $0 < \alpha < 1$, and sufficiently large $\alpha\lambda$ and $\bar{\alpha}\lambda$ is applied to a $\mathcal{N}(0, \sigma^2)$ or a $\mathcal{L}(0, \sigma^2)$ source,*

$$\begin{array}{lll} A. & P_{k+1}(\alpha, \lambda) < P_k(\alpha, \lambda) P_1(\alpha, \lambda) & \textit{for all } k \geq 1 \ , \\ B. & P_{k-1}(\alpha, \lambda) < P_k(\alpha, \lambda) P_{-1}(\alpha, \lambda) & \textit{for all } k \leq -1 \ . \end{array}$$

**Lemma 2**

$$\lim_{p \to 0} \frac{\mathcal{H}(1 - p + p\, o_{p \to 0})}{\mathcal{H}(p)} = 0 \ .$$

We comment that this lemma is due to the fact that the entropy function $\mathcal{H}(p) = -p \log p$ has infinite slope at $p = 0$ and finite slope at $p = 1$, as illustrated in Figure 2. A formal proof is provided in the appendix.

The next lemma shows that in low resolution, quantizer entropy is dominated by the cells adjacent to the center cell.

**Lemma 3** *For a UTQ with offset $\alpha$, $0 < \alpha < 1$, and normalized step size $\lambda$ applied to a $\mathcal{N}(0, \sigma^2)$ or a $\mathcal{L}(0, \sigma^2)$ source,*

$$\mathcal{H}(\ldots, P_{-1}(\alpha, \lambda), P_0(\alpha, \lambda), P_1(\alpha, \lambda), \ldots) = \mathcal{H}\big(P_{-1}(\alpha, \lambda), P_1(\alpha, \lambda)\big) \left[1 + o_{\alpha\lambda, \bar{\alpha}\lambda}\right] \ ,$$

*Proof:* For brevity, we omit the parameters $\alpha$ and $\lambda$ from $P_k(\alpha, \lambda)$. The proof is composed of two main steps. In Step 1, we show that $\mathcal{H}(\ldots, P_{-1}, P_0, P_1, \ldots)$ can be asymptotically approximated by the three middle terms; that is, $\mathcal{H}(\ldots, P_{-1}, P_0, P_1, \ldots) = \mathcal{H}\big(P_{-1}, P_0, P_1\big) \left[1 + o_{\alpha\lambda, \bar{\alpha}\lambda}\right]$. In Step 2, it is shown that these three terms can be asymptotically approximated by only two terms; that is, $\mathcal{H}\big(P_{-1}, P_0, P_1\big) = \mathcal{H}\big(P_{-1}, P_1\big)\big[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\big]$, where we note that $P_0 \to 1$ as $\alpha\lambda \to \infty$ and $\bar{\alpha}\lambda \to \infty$.

Step 1: We first show that for all sufficiently large $\alpha\lambda$ and $\bar{\alpha}\lambda$,

$$1 < \frac{\mathcal{H}(\dots, P_{-1}, P_0, P_1, \dots)}{\mathcal{H}(P_{-1}, P_0, P_1)} < 1 + 6P_1 + 6P_{-1} \ . \tag{7}$$

The left inequality is trivial. We upper bound the middle term in the following way:

$$\begin{aligned} \frac{\sum_{k=-\infty}^{\infty} -P_k \log P_k}{\sum_{k=-1}^{1} -P_k \log P_k} &= 1 + \frac{\sum_{k=-\infty}^{-2} -P_k \log P_k + \sum_{k=2}^{\infty} -P_k \log P_k}{\sum_{k=-1}^{1} -P_k \log P_k(s)} \\ &< 1 + \frac{\sum_{k=-\infty}^{-2} -P_k \log P_k}{-P_{-1} \log P_{-1}} + \frac{\sum_{k=2}^{\infty} -P_k \log P_k}{-P_1 \log P_1} \ . \end{aligned} \tag{8}$$

Consider the terms in the last summation. When $\alpha\lambda$ and $\bar{\alpha}\lambda$ are sufficiently large, $-P_k \log P_k < -P_1^k \log P_1^k$ for all $k \geq 2$. To see this we observe that when $\alpha\lambda$ and $\bar{\alpha}\lambda$ are sufficiently large, Lemma 1 implies $P_k < P_{k-1}P_1$ for all $k \geq 2$, which in turn implies $P_k < P_1^k < P_1$, for all $k \geq 2$. Next, for a Gaussian source (3) and Fact G1 imply that $P_1 = Q(\bar{\alpha}\lambda) - Q((1+\bar{\alpha})\lambda) < Q(\bar{\alpha}\lambda) = o_{\bar{\alpha}\lambda}$, and for a Laplacian source (4) and Fact L1 imply $P_1 = \frac{L(\bar{\alpha}\lambda)\,[1+o_\lambda]}{\sqrt{2}} = o_{\bar{\alpha}\lambda}$. Therefore, for both sources, $P_1 < \frac{1}{e}$ when $\bar{\alpha}\lambda$ is sufficiently large. Since $-p \log p$ increases for $p < \frac{1}{e}$, $-P_k \log P_k < -P_1^k \log P_1^k$ for all $k \geq 2$, when $\alpha\lambda$ and $\bar{\alpha}\lambda$ are sufficiently large. Substituting this into the last summation of (8), we have that when $\alpha\lambda$ and $\bar{\alpha}\lambda$ are sufficiently large,

$$\frac{\sum_{k=2}^{\infty} -P_k \log P_k}{-P_1 \log P_1} < \frac{\sum_{k=2}^{\infty} -P_1^k \log P_1^k}{-P_1 \log P_1} = \sum_{k=2}^{\infty} k P_1^{k-1} = \frac{2P_1}{(1-P_1)^2} - \frac{P_1^2}{(1-P_1)^2} < \frac{2P_1}{(1-P_1)^2} < 6P_1 \ ,$$

where the last inequality derives from the fact that $P_1 < \frac{1}{e}$. In much the same way it follows that $P_{-1} = o_{\alpha\lambda}$, and that when $\alpha\lambda$ and $\bar{\alpha}\lambda$ are sufficiently large, $\frac{\sum_{k=-\infty}^{-2} -P_k \log P_k}{-P_{-1} \log P_{-1}} < 6P_{-1}$. This shows (7). Substituting $P_{-1} = o_{\alpha\lambda}$ and $P_1 = o_{\bar{\alpha}\lambda}$ into (7), we obtain

$$\mathcal{H}(\dots, P_{-1}, P_0, P_1, \dots) = \mathcal{H}\big(P_{-1}, P_0, P_1\big)\ \big[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\big] \ ,$$

which completes Step 1.

Step 2: We will show that $\mathcal{H}(P_0) = \mathcal{H}(P_{-1}, P_1) o_{\alpha\lambda,\bar{\alpha}\lambda}$ from which it will follow that

$$\mathcal{H}\big(P_{-1}, P_0, P_1\big) = \mathcal{H}\big(P_{-1}, P_1\big)\ \big[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\big] \ .$$

Define $\widetilde{P} = \sum_{k=-\infty}^{-2} P_k + \sum_{k=2}^{\infty} P_k$. Using the fact that for all sufficiently large $\alpha\lambda$ and $\bar{\alpha}\lambda$, $P_k < P_{k-1}P_1$ for all $k \geq 2$, we upper bound the second sum as

$$\sum_{k=2}^{\infty} P_k < \sum_{k=2}^{\infty} P_1^k = \frac{P_1^2}{1 - P_1} < 2P_1^2 \ ,$$

where the last inequality is due to $P_1 < \frac{1}{2}$ for all sufficiently large $\bar{\alpha}\lambda$. The first sum in the definition of $\widetilde{P}$ can be upper bounded in much the same way. Thus when $\alpha\lambda$ and $\bar{\alpha}\lambda$ are both sufficiently large, $\widetilde{P} < 2(P_{-1}^2 + P_1^2)$. Therefore, $P_0 > 1 - P_{-1} - P_1 - 2(P_{-1}^2 + P_1^2) > 1 - P_{-1} - P_1 - 2(P_{-1} + P_1)^2$. Since $P_{-1} + P_1 = o_{\alpha\lambda} + o_{\bar{\alpha}\lambda}$, it follows that when $\alpha\lambda$ and $\bar{\alpha}\lambda$ are sufficiently large, $P_0 > 1 - P_{-1} - P_1 - 2(P_{-1} + P_1)^2 > \frac{1}{e}$, which since $\mathcal{H}(p)$ decreases monotonically for $p > \frac{1}{e}$, implies that $\mathcal{H}(P_0) < \mathcal{H}\big(1 - P_{-1} - P_1 - 2(P_{-1} + P_1)^2\big)$. Consequently,

$$\begin{aligned}\frac{\mathcal{H}(P_0)}{\mathcal{H}\big(P_{-1}, P_1\big)} &< \frac{\mathcal{H}(1 - P_{-1} - P_1 - 2(P_{-1} + P_1)^2)}{\mathcal{H}(P_{-1}, P_1)} < \frac{\mathcal{H}\big(1 - P_{-1} - P_1 - 2(P_{-1} + P_1)^2\big)}{\mathcal{H}(P_{-1} + P_1)} \\ &= \frac{\mathcal{H}(1 - p - 2p^2)}{\mathcal{H}(p)} ,\end{aligned}$$

where $p \stackrel{\Delta}{=} P_{-1} + P_1$, and where the second inequality follows from the easy to prove fact that for any $a, b \in \mathbb{R}^+$, $\mathcal{H}(a + b) < \mathcal{H}(a, b)$. We observe that as $\alpha\lambda$ and $\bar{\alpha}\lambda$ tend to infinity, $p$ goes to zero. Therefore, by Lemma 2 it follows that $\frac{\mathcal{H}(1-p-2p^2)}{\mathcal{H}(p)} \to 0$ as $p \to 0$. This shows that $\mathcal{H}(P_0) = \mathcal{H}(P_{-1}, P_1)\, o_{\alpha\lambda,\bar{\alpha}\lambda}$, which completes Step 2 and the proof of the lemma. □

**Lemma 4** *Let $a(s)$ and $b(s)$ be positive functions on $\mathbb{R}$ such that $a(s) = b(s)[1 + o_s]$, and for some $\varepsilon > 0$, $|b(s) - 1| > \varepsilon$ for all $s$. Then*

$$\mathcal{H}(a(s)) = \mathcal{H}(b(s))\, \big[1 + o_s\big] \ .$$

**Lemma 5**

$$\mathcal{H}\big(Q(x)\big) = \frac{\log e}{2} x\, G(x)\, \big[1 + o_x\big] \ .$$

**Lemma 6**

$$\mathcal{H}\Big(\frac{L(x)}{\sqrt{2}}\Big) \ = \ xL(x)(\log e)\big[1 + o_x\big] \ .$$

The following theorem gives the low resolution approximation to the entropy of uniform quantization for a Gaussian or Laplacian source.

**Theorem 7** *For a UTQ with offset $\alpha$, $0 < \alpha < 1$, and normalized step size $\lambda$ applied to a $\mathcal{N}(0, \sigma^2)$*

$$H(\alpha, \lambda) \ = \ \frac{\log e}{2}\Big(\alpha\lambda\, G(\alpha\lambda) + \bar{\alpha}\lambda\, G(\bar{\alpha}\lambda)\Big)\, \big[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\big] \ , \qquad (9)$$

*and applied to a $\mathcal{L}(0, \sigma^2)$*

$$H(\alpha, \lambda) \ = \ (\log e)\Big(\alpha\lambda\, L(\alpha\lambda) + \bar{\alpha}\lambda\, L(\bar{\alpha}\lambda)\Big)\, \big[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\big] \ , \qquad (10)$$

*where $H(\alpha, \lambda) = \mathcal{H}(\ldots, P_{-1}(\alpha, \lambda), P_0(\alpha, \lambda), P_1(\alpha, \lambda), \ldots)$ is the quantizer entropy.*

If one fixes $\alpha$, this theorem shows the rate at which entropy converges to 0 as $\lambda \to \infty$. However, the convergence is not uniform in $\alpha$, and this theorem shows how entropy depends on $\alpha$ as well as $\lambda$. In particular, it gives an accurate approximation to quantizer entropy when both $\alpha\lambda$ and $\bar{\alpha}\lambda$ are large. Notice that $\alpha = 0$ is not allowed since $H(0, \lambda) = 1 + o_\lambda$, namely, the output entropy does not go to zero as $\lambda \to \infty$.

*Proof:* For brevity, we omit the parameters $\alpha$ and $\lambda$ from $P_k(\alpha, \lambda)$. Lemma 3 shows that

$$H(\alpha, \lambda) = \mathcal{H}(\ldots, P_{-1}, P_0, P_1, \ldots) = \mathcal{H}\big(P_{-1}, P_1\big) \left[1 + o_{\alpha\lambda, \bar{\alpha}\lambda}\right] . \tag{11}$$

For a Gaussian source (3) and Fact G6 imply that $P_{-1} = Q(\alpha\lambda) - Q\big((1+\alpha)\lambda\big) = Q(\alpha\lambda)[1 + o_\lambda]$, and thus in particular $P_{-1} = Q(\alpha\lambda)[1 + o_{\alpha\lambda}]$, since $0 < \alpha < 1$. Since $|Q(\alpha\lambda) - 1| \geq \frac{1}{2}$ for all $\alpha\lambda$, it follows from Lemma 4 that $\mathcal{H}(P_{-1}) = \mathcal{H}(Q(\alpha\lambda)) \left[1 + o_{\alpha\lambda}\right]$. Next, applying Lemma 5, we obtain $\mathcal{H}(Q(\alpha\lambda)) = \left(\frac{1}{2}\log e\right) \alpha\lambda\, G(\alpha\lambda) \left[1 + o_{\alpha\lambda}\right]$. Combining these yields $\mathcal{H}(P_{-1}) = \left(\frac{1}{2}\log e\right) \alpha\lambda\, G(\alpha\lambda) \left[1 + o_{\alpha\lambda}\right]$. In a similar way $\mathcal{H}(P_1) = \left(\frac{1}{2}\log e\right) \bar{\alpha}\lambda\, G(\bar{\alpha}\lambda) \left[1 + o_{\bar{\alpha}\lambda}\right]$. Combining the expressions for $\mathcal{H}(P_{-1})$ and $\mathcal{H}(P_1)$ together with (11) complete the proof of the Gaussian case.

Next, consider the Laplacian case. From (4) and Fact L1 we have that $P_{-1} = \frac{L(\alpha\lambda) - L((1+\alpha)\lambda)}{\sqrt{2}} = \frac{L(\alpha\lambda)\,[1+o_\lambda]}{\sqrt{2}}$, and thus in particular $P_{-1} = \frac{L(\alpha\lambda)}{\sqrt{2}}[1 + o_{\alpha\lambda}]$, since $0 < \alpha < 1$. Since $\left|\frac{L(\alpha\lambda)}{\sqrt{2}} - 1\right| \geq \frac{1}{2}$ for all $\alpha\lambda$, it follows from Lemma 4 that $\mathcal{H}(P_{-1}) = \mathcal{H}(L(\frac{\alpha\lambda)}{\sqrt{2}}) \left[1 + o_{\alpha\lambda}\right]$. Next, applying Lemma 6, we obtain $\mathcal{H}(\frac{L(\alpha\lambda)}{\sqrt{2}}) = (\log e)\, \alpha\lambda L(\alpha\lambda) \left[1 + o_{\alpha\lambda}\right]$. Combining these yields $\mathcal{H}(P_{-1}) = (\log e)\, \alpha\lambda\, L(\alpha\lambda) \left[1 + o_{\alpha\lambda}\right]$. In a similar way $\mathcal{H}(P_1) = (\log e)\, \bar{\alpha}\lambda\, L(\bar{\alpha}\lambda) \left[1 + o_{\bar{\alpha}\lambda}\right]$. Combining the expressions for $\mathcal{H}(P_{-1})$ and $\mathcal{H}(P_1)$ together with (11) shows the Laplacian case and completes the proof of the theorem. □

We now comment on the cell or cells that dominate entropy when it is small. The entropy $H(\alpha, \lambda)$ will be small if and only if $P_0 \approx 1$ and $P_k \approx 0$, $k \neq 0$, which makes $-P_k \log P_k \approx 0$ for all $k$, and which happens if and only if $\alpha\lambda$ and $\bar{\alpha}\lambda$ are both large. Lemma 3 shows that $H(\alpha, \lambda)$ is dominated by the cells, $S_{-1}$ and $S_1$, immediately adjacent to the center cell. This is not coincidental; rather, as mentioned earlier, it follows from the fact, illustrated in Figure 2, that the entropy function, $\mathcal{H}(p) = -p \log p$, has infinite slope at $p = 0$ and finite slope at $p = 1$. Thus, when entropy is nearly zero, it is dominated by the largest of the nearly zero probabilities, which are $P_{-1}$ and/or $P_1$. Indeed the two terms within the large parentheses in (9) and (10) correspond to $\mathcal{H}(P_{-1})$ and $\mathcal{H}(P_1)$, respectively. If $\alpha\lambda << \bar{\alpha}\lambda$, e.g. if $\alpha < \frac{1}{2}$ and $\lambda$ is very large, then $P_{-1} >> P_1$, and it is cell $S_{-1}$ and the first term within the parentheses that dominate the entropy. Conversely, if $\bar{\alpha}\lambda << \alpha\lambda$, then $P_1 >> P_{-1}$, and it is cell $S_1$ and the second term within the parentheses that dominate. Finally, if $\alpha\lambda \approx \bar{\alpha}\lambda$, then the two dominating cells contribute roughly the same to the entropy.

# 4 Asymptotic Distortion

The main idea in deriving an asymptotic expression for distortion of a UTQ is the same for both sources and distortion measures. Specifically, Our goal in all cases is to find an asymptotic expression, as $d(\alpha, \Delta, \sigma^2) \to D_{\max}$, for the difference between $D_{\max}$ and $d(\alpha, \Delta, \sigma^2)$ normalized by $D_{\max}$, i.e., an asymptotic expression for $\frac{D_{\max} - d(\alpha,\Delta,\sigma^2)}{D_{\max}}$. For brevity we omit the arguments of $d(\alpha, \Delta, \sigma^2)$. To this end, we define the following: $d_{k,s} = \int_{S_k} |x - r_k|^s f(x)\, dx$ is the contribution to distortion of the $k^{th}$ cell, and $\sigma_{k,s} = \int_{S_k} |x|^s f(x)\, dx$ is the contribution to the variance or to $D_{\max,1}$ of the $k^{th}$ cell, where $f$ is the pdf of the source and $s \in \{1, 2\}$. We observe that $D_{\max,s} = \sum_k \sigma_{k,s}$ and that $d = \sum_k d_{k,s}$. Next, we write

$$D_{\max} - d = D_{\max} - d_0 - d_{-1} - d_1 - \sum_{|k| \geq 2} d_k \tag{12}$$

and evaluate the terms above.

Since there are differences in the derivations of the two distortion measures, we divide this section into two subsections, each of which considers one distortion measure and both sources. We begin with the squared error distortion measure, which is the simpler of the two.

## 4.1 Squared Error Distortion Measure

**Theorem 8** *For a UTQ with offset $\alpha$, $0 < \alpha < 1$, normalized step size $\lambda$, centroid reconstruction levels, and a $\mathcal{N}(0, \sigma^2)$ source*

$$\frac{\sigma^2 - d(\alpha, \Delta, \sigma^2)}{\sigma^2} = \Big(\alpha\lambda\, G(\alpha\lambda) + \bar{\alpha}\lambda\, G(\bar{\alpha}\lambda)\Big) \left[1 + o_{\alpha\lambda, \bar{\alpha}\lambda}\right] ,$$

*and for a $\mathcal{L}(0, \sigma^2)$ source*

$$\frac{\sigma^2 - d(\alpha, \Delta, \sigma^2)}{\sigma^2} = \frac{1}{\sqrt{2}} \Big((\alpha\lambda)^2\, L(\alpha\lambda) + (\bar{\alpha}\lambda)^2\, L(\bar{\alpha}\lambda)\Big) \left[1 + o_{\alpha\lambda, \bar{\alpha}\lambda}\right] .$$

*Proof:* We will deviate slightly in this proof from the previously stated notation convention and let $\sigma_k^2$ denote $\sigma_{k,2}$. We begin by noticing that $\sigma_k^2 = \sigma^2\big(V((k - \alpha)\lambda) - V((k + 1 - \alpha)\lambda)\big)$, where $V(x)$ is defined in Fact GL2, and that $d_k = \sigma_k^2 - r_k^2 P_k$, where we recall that $r_k = \int_{S_k} x \frac{f(x)}{P_k}\, dx$. We now evaluate the terms in (12) in reverse order. First, since $D_k \leq \sigma_k^2$,

$$\begin{aligned}
\sum_{|k|\geq 2} d_k \ \leq \ \sum_{|k|\geq 2} \sigma_k^2 \ &= \ \int_{-\infty}^{-(\alpha+1)\Delta} x^2 f(x)\,dx + \int_{(2-\alpha)\Delta}^{\infty} x^2 f(x)\,dx \\
&= \ \sigma^2 V((\alpha+1)\lambda) + \sigma^2 V((2-\alpha)\lambda) \\
&\overset{(a)}{=} \ \sigma^2 V(\alpha\lambda)\, o_\lambda + \sigma^2 V(\bar{\alpha}\lambda)\, o_\lambda \ , \\
&\overset{(b)}{=} \ \begin{cases} \sigma^2 \alpha\lambda\, G(\alpha\lambda)\, o_{\alpha\lambda} + \sigma^2 \bar{\alpha}\lambda\, G(\bar{\alpha}\lambda)\, o_{\bar{\alpha}\lambda}, & \text{for } \mathcal{N}(0,\sigma^2) \\ \sigma^2 (\alpha\lambda)^2\, L(\alpha\lambda)\, o_{\alpha\lambda} + \sigma^2 (\bar{\alpha}\lambda)^2\, L(\bar{\alpha}\lambda)\, o_{\bar{\alpha}\lambda}, & \text{for } \mathcal{L}(0,\sigma^2) \end{cases} ,
\end{aligned} \tag{13}$$

where $(a)$ follows from Fact GL4, and $(b)$ is obtained using Fact GL2.

Next, in order to evaluate $D_1$, we will need an expression for $P_1$. Specifically, we have

$$\begin{aligned}
A. \quad P_1^G \ &= \ Q((1-\alpha)\lambda) - Q((2-\alpha)\lambda) \ \overset{(a)}{=} \ Q((1-\alpha)\lambda)\left[1 + o_\lambda\right] \ \overset{(b)}{=} \ \frac{1}{\bar{\alpha}\lambda} G(\bar{\alpha}\lambda)\left[1 + o_{\bar{\alpha}\lambda}\right] \ , \\
B. \quad P_1^L \ &= \ \frac{L((1-\alpha)\lambda) - L((2-\alpha)\lambda)}{\sqrt{2}} \ \overset{(c)}{=} \ \frac{1}{\sqrt{2}} L(\bar{\alpha}\lambda)\left[1 + o_\lambda\right] \ ,
\end{aligned} \tag{14}$$

where $(a)$ follows from Fact G6, $(b)$ follows from Fact G5, and $(c)$ follows from Fact L1. We now have

$$\begin{aligned}
d_1 \ &= \ \sigma_1^2 - r_1^2 P_1 \ \overset{(a)}{=} \ \sigma^2 (V((1-\alpha)\lambda) - V((2-\alpha)\lambda)) - \left(\frac{\sigma C((1-\alpha)\lambda) - \sigma C((2-\alpha)\lambda)}{P_1}\right)^2 P_1 \\
&\overset{(b)}{=} \ \sigma^2 V((1-\alpha)\lambda)\left[1 + o_\lambda\right] - \frac{\sigma^2 \Big(C((1-\alpha)\lambda)\left[1 + o_\lambda\right]\Big)^2}{P_1} \\
&\overset{(c)}{=} \ \begin{cases} \sigma^2 \bar{\alpha}\lambda\, G(\bar{\alpha}\lambda)\,[1 + o_{\bar{\alpha}\lambda}] - \frac{\sigma^2 G^2(\bar{\alpha}\lambda)\left[1+o_\lambda\right]}{\frac{1}{\bar{\alpha}\lambda} G(\bar{\alpha}\lambda)\left[1+o_{\bar{\alpha}\lambda}\right]}, & \text{for } \mathcal{N}(0,\sigma^2) \\ \sigma^2 \frac{1}{\sqrt{2}} (\bar{\alpha}\lambda)^2\, L(\bar{\alpha}\lambda)\,[1 + o_{\bar{\alpha}\lambda}] - \frac{\sigma^2 \frac{1}{2} (\bar{\alpha}\lambda)^2 L^2(\bar{\alpha}\lambda)\left[1+o_{\bar{\alpha}\lambda}\right]}{\frac{1}{\sqrt{2}} L(\bar{\alpha}\lambda)\left[1+o_\lambda\right]}, & \text{for } \mathcal{L}(0,\sigma^2) \end{cases} ,
\end{aligned}$$

where $(a)$ is due to the definition of $C(x)$ given in Fact GL1, $(b)$ follows from Facts GL3 and GL4, and $(c)$ derives from (14) and Facts GL1 and GL2. We now obtain that

$$d_1 \ = \ \begin{cases} \sigma^2 \bar{\alpha}\lambda\, G(\bar{\alpha}\lambda)\, o_{\bar{\alpha}\lambda}, & \text{for } \mathcal{N}(0,\sigma^2) \\ \sigma^2 (\bar{\alpha}\lambda)^2\, L(\bar{\alpha}\lambda)\, o_{\bar{\alpha}\lambda}, & \text{for } \mathcal{L}(0,\sigma^2) \end{cases} . \tag{15}$$

By symmetry it follows that

$$d_{-1} \ = \ \begin{cases} \sigma^2 \alpha\lambda\, G(\alpha\lambda)\, o_{\alpha\lambda}, & \text{for } \mathcal{N}(0,\sigma^2) \\ \sigma^2 (\alpha\lambda)^2\, L(\alpha\lambda)\, o_{\alpha\lambda}, & \text{for } \mathcal{L}(0,\sigma^2) \end{cases} . \tag{16}$$

Finally,

$$\sigma^2 - d_0 = (\sigma^2 - \sigma_0^2) + (\sigma_0^2 - d_0) \ , \tag{17}$$

where as in (13) above

$$\begin{aligned}
\sigma^2 - \sigma_0^2 \ &= \ \sum_{k\neq 0} \sigma_k^2 \ = \ \sigma^2 V(\alpha\lambda) + \sigma^2 V\big((1-\alpha)\lambda\big) \\
&= \begin{cases} \sigma^2 \alpha\lambda \, G(\alpha\lambda) \left[1 + o_{\alpha\lambda}\right] + \sigma^2 \bar{\alpha}\lambda \, G(\bar{\alpha}\lambda) \left[1 + o_{\bar{\alpha}\lambda}\right], & \text{for } \mathcal{N}(0,\sigma^2) \\ \frac{\sigma^2}{\sqrt{2}} (\alpha\lambda)^2 \, L(\alpha\lambda) \left[1 + o_{\alpha\lambda}\right] + \frac{\sigma^2}{\sqrt{2}} (\bar{\alpha}\lambda)^2 \, L(\bar{\alpha}\lambda) \left[1 + o_{\bar{\alpha}\lambda}\right], & \text{for} \mathcal{L}(0,\sigma^2) \end{cases} ,
\end{aligned} \tag{18}$$

where the last equality uses Fact GL2. To evaluate $\sigma_0^2 - D_0$ we first note that $P_0^G = 1 - Q(\alpha\lambda) - Q(\bar{\alpha}\lambda) = 1 + o_{\alpha\lambda} + o_{\bar{\alpha}\lambda}$, and similarly, $P_0^L = 1 - \frac{L(\alpha\lambda)}{\sqrt{2}} - \frac{L(\bar{\alpha}\lambda)}{\sqrt{2}} = 1 + o_{\alpha\lambda} + o_{\bar{\alpha}\lambda}$. Proceeding as in (15)

$$\begin{aligned}
\sigma_0^2 - D_0 \ &= \ r_0^2 P_0 \ = \ \left( \frac{\sigma C(\alpha\lambda) - \sigma C((1-\alpha)\lambda)}{P_0} \right)^2 P_0 \ = \ \frac{\sigma^2 \big(C(\alpha\lambda) - C(\bar{\alpha}\lambda)\big)^2}{1 + o_{\alpha\lambda} + o_{\bar{\alpha}\lambda}} \\
&= \ \frac{\sigma^2 \Big( C(\alpha\lambda)\big(C(\alpha\lambda) - C(\bar{\alpha}\lambda)\big) + C(\bar{\alpha}\lambda)\big(G(\bar{\alpha}\lambda) - C(\alpha\lambda)\big) \Big)}{1 + o_{\alpha\lambda} + o_{\bar{\alpha}\lambda}} \\
&\overset{(a)}{=} \ \frac{\sigma^2 \big( \alpha\lambda C(\alpha\lambda) o_{\alpha\lambda,\bar{\alpha}\lambda} + \bar{\alpha}\lambda C(\bar{\alpha}\lambda) o_{\alpha\lambda,\bar{\alpha}\lambda} \big)}{1 + o_{\alpha\lambda} + o_{\bar{\alpha}\lambda}} \\
&\overset{(b)}{=} \ \begin{cases} \sigma^2 \Big( \alpha\lambda \, G(\alpha\lambda) + \bar{\alpha}\lambda \, G(\bar{\alpha}\lambda) \Big) \, o_{\alpha\lambda,\bar{\alpha}\lambda}, & \text{for } \mathcal{N}(0,\sigma^2) \\ \sigma^2 \Big( (\alpha\lambda)^2 \, L(\alpha\lambda) + (\bar{\alpha}\lambda)^2 \, L(\bar{\alpha}\lambda) \Big) \, o_{\alpha\lambda,\bar{\alpha}\lambda}, & \text{for } \mathcal{L}(0,\sigma^2) \end{cases} ,
\end{aligned} \tag{19}$$

where $(a)$ follows from having $C(\alpha\lambda) - C(\bar{\alpha}\lambda) = \alpha\lambda \frac{C(\alpha\lambda) - C(\bar{\alpha}\lambda)}{\alpha\lambda} = \alpha\lambda o_{\alpha\lambda,\bar{\alpha}\lambda}$ and similarly $C(\bar{\alpha}\lambda) - C(\alpha\lambda) = \bar{\alpha}\lambda o_{\alpha\lambda,\bar{\alpha}\lambda}$, and $(b)$ follows from Fact GL1. Substituting (18) and (19) into (17) yields

$$\sigma^2 - d_0 \ = \ \begin{cases} \sigma^2 \Big( \alpha\lambda \, G(\alpha\lambda) + \bar{\alpha}\lambda \, G(\bar{\alpha}\lambda) \Big) \left[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\right], & \text{for } \mathcal{N}(0,\sigma^2) \\ \frac{\sigma^2}{\sqrt{2}} \Big( (\alpha\lambda)^2 \, L(\alpha\lambda) + (\bar{\alpha}\lambda)^2 \, L(\bar{\alpha}\lambda) \Big) \left[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\right], & \text{for } \mathcal{L}(0,\sigma^2) \end{cases} . \tag{20}$$

Substituting (13), (15), (16) and (20) into (12) yields

$$\sigma^2 - d \ = \ \begin{cases} \sigma^2 \Big( \alpha\lambda \, G(\alpha\lambda) + \bar{\alpha}\lambda \, G(\bar{\alpha}\lambda) \Big) \left[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\right], & \text{for } \mathcal{N}(0,\sigma^2) \\ \frac{\sigma^2}{\sqrt{2}} \Big( (\alpha\lambda)^2 \, L(\alpha\lambda) + (\bar{\alpha}\lambda)^2 \, L(\bar{\alpha}\lambda) \Big) \left[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\right], & \text{for } \mathcal{L}(0,\sigma^2) \end{cases} .$$

Dividing the above by $\sigma^2$ gives the desired result. □

### 4.2 Absolute Error Distortion Measure

**Theorem 9** *For a UTQ with offset $\alpha$, $0 < \alpha < 1$, normalized step size $\lambda$, median reconstruction levels, and a $\mathcal{N}(0, \sigma^2)$ source*

$$\frac{D_{\max} - d(\alpha, \Delta, \sigma^2)}{D_{\max}} \;=\; \sqrt{\frac{\pi}{2}}\Big(G(\alpha\lambda) + G(\bar{\alpha}\lambda)\Big)\left[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\right] \;,$$

*and for a $\mathcal{L}(0, \sigma^2)$ source*

$$\frac{D_{\max} - d(\alpha, \Delta, \sigma^2)}{D_{\max}} \;=\; \Big(\alpha\lambda\, L(\alpha\lambda) + \bar{\alpha}\lambda\, L(\bar{\alpha}\lambda)\Big)\left[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\right] \;.$$

*Proof:* We denote $\sigma_{k,1}$ by $\sigma_k$ for short. It is not hard to see that

$$\sigma_k \;=\; \begin{cases} \sigma\big[C((k-\alpha)\lambda) - C((k+1-\alpha)\lambda)\big], & k > 0 \\ \sigma\big[\frac{D_{\max}}{\sigma} - C(\alpha\lambda) - C(\bar{\alpha}\lambda)\big], & k = 0 \\ \sigma\big[C((k+1-\alpha)\lambda) - C((k-\alpha)\lambda)\big], & k < 0 \end{cases} \cdot \tag{21}$$

Furthermore, we have

$$\begin{aligned} d_k \;&=\; \int_{(k-\alpha)\Delta}^{(k+1-\alpha)\Delta} |x - r_k| f(x)\,dx \;=\; \int_{(k-\alpha)\Delta}^{r_k} (x - r_k) f(x)\,dx + \int_{r_k}^{(k+1-\alpha)\Delta} (r_k - x) f(x)\,dx \\ &=\; r_k\Big[\int_{(k-\alpha)\Delta}^{r_k} f(x)\,dx + \int_{r_k}^{(k+1-\alpha)\Delta} f(x)\,dx\Big] - \int_{(k-\alpha)\Delta}^{r_k} x f(x)\,dx + \int_{r_k}^{(k+1-\alpha)\Delta} x f(x)\,dx \\ &=\; \begin{cases} \sigma_k - 2\sigma\big[C((k-\alpha)\lambda) - C(\frac{r_k}{\sigma})\big], & k > 0 \\ \sigma_k - 2\sigma\big[C(0) - C(\frac{r_0}{\sigma})\big], & k = 0 \\ \sigma_k - 2\sigma\big[C((k+1-\alpha)\lambda) - C(\frac{r_k}{\sigma})\big], & k < 0 \end{cases} , \end{aligned} \tag{22}$$

where $f$ is the density of the source, and where the last equality follows from the fact that $r_k$ is the median of its cell. We now evaluate the terms in (12) in reverse order. First, from (22) we have that $d_k \le \sigma_k$, thus using (21) we obtain

$$\sum_{|k|\ge 2} d_k \;\le\; \sum_{|k|\ge 2} \sigma_k \;=\; \sigma C((\alpha+1)\lambda) + \sigma C((2-\alpha)\lambda) = \sigma C(\alpha\lambda)\, o_\lambda + \sigma C(\bar{\alpha}\lambda)\, o_\lambda \;, \tag{23}$$

where the last equality follows from Fact GL3. Next, we consider $d_1$. From (21) and (22) we have

$$\begin{aligned} d_1 \;&=\; \sigma C(\bar{\alpha}\lambda) - \sigma C((1+\bar{\alpha})\lambda) - 2\sigma\big[c(\bar{\alpha}\lambda) - C(\frac{r_1}{\sigma})\big] \;\overset{(a)}{=}\; 2\sigma C(\frac{r_1}{\sigma}) - \sigma C(\bar{\alpha}\lambda)[1 + o_\lambda] \\ &\overset{(b)}{=}\; \sigma C(\bar{\alpha}\lambda) o_{\bar{\alpha}\lambda} \;, \end{aligned} \tag{24}$$

where $(a)$ uses Fact GL3, and $(b)$ follows from Lemma 10 below, which evaluates $C(\frac{r_1}{\sigma})$ and whose proof is left to the appendix.

**Lemma 10** *For a UTQ with offset $\alpha$, $0 < \alpha < 1$, normalized step size $\lambda$, median reconstruction levels, and a $\mathcal{N}(0, \sigma^2)$ or $\mathcal{L}(0, \sigma^2)$ source,*

$$C(\frac{r_1}{\sigma}) \;=\; \frac{1}{2} C(\bar{\alpha}\lambda)[1 + o_{\bar{\alpha}\lambda}] \;.$$

By symmetry it follows that

$$d_{-1} = \sigma C(\alpha\lambda) o_{\alpha\lambda} \;. \tag{25}$$

Finally,

$$D_{\max} - d_0 = (D_{\max} - \sigma_0) + (\sigma_0 - d_0) \;, \tag{26}$$

where as in (23) above

$$D_{\max} - \sigma_0 \;=\; \sum_{k \neq 0} \sigma_k \;=\; \sigma\Big(C(\alpha\lambda) + C(\bar{\alpha}\lambda)\Big) \;, \tag{27}$$

and using (22)

$$\sigma_0 - d_0 \;=\; 2\sigma\big[C(0) - C(\frac{r_0}{\sigma})\big] \;=\; 2\sigma \int_0^{\frac{r_0}{\sigma}} x f(x)\, dx \;\leq\; 2\sigma f(0) \big(\frac{r_0}{\sigma}\big)^2 \;=\; \sigma\Big(C(\alpha\lambda) + C(\bar{\alpha}\lambda)\Big) o_{\alpha\lambda,\bar{\alpha}\lambda}, \tag{28}$$

where $f(x) = G(x)$ or $f(x) = L(x)$, depending on the source, and where the last equality follows from Lemma 11 below, which evaluates $\big(\frac{r_0}{\sigma}\big)^2$ and whose proof is left to the appendix.

**Lemma 11** *For a UTQ with offset $\alpha$, $0 < \alpha < 1$, normalized step size $\lambda$, median reconstruction levels, and a $\mathcal{N}(0, \sigma^2)$ or $\mathcal{L}(0, \sigma^2)$ source,*

$$\big(\frac{r_0}{\sigma}\big)^2 \;=\; \Big(C(\alpha\lambda) + C(\bar{\alpha}\lambda)\Big) o_{\alpha\lambda,\bar{\alpha}\lambda} \;.$$

Substituting (27) and (28) into (26) yields

$$D_{\max} - d_0 \;=\; \sigma\Big(C(\alpha\lambda) + \sigma C(\bar{\alpha}\lambda)\Big)[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}] \;. \tag{29}$$

Substituting (23), (24), (25) and (29) into (12) and using Fact GL1 yields

$$D_{\max} - d \;=\; \begin{cases} \sigma\Big(G(\alpha\lambda) + G(\bar{\alpha}\lambda)\Big)\, \big[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\big], & \text{for } \mathcal{N}(0, \sigma^2) \\ \frac{\sigma}{\sqrt{2}}\Big((\alpha\lambda)\, L(\alpha\lambda) + (\bar{\alpha}\lambda)\, L(\bar{\alpha}\lambda)\Big)\, \big[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}\big], & \text{for } \mathcal{L}(0, \sigma^2) \end{cases} \;.$$

Dividing the above by $D_{\max}$ and recalling that $D^G_{\max} = \sqrt{\frac{2}{\pi}}\sigma$ and $D^L_{\max} = \frac{\sigma}{\sqrt{2}}$, gives the desired result. □

**Remark 1:** Like Theorem 7, Theorems 8 and 9 give accurate approximations when both $\alpha\lambda$ and $\bar{\alpha}\lambda$ are large.

**Remark 2:** The rate at which distortion converges to $D_{\max}$ is dominated by the center cell for both distortion measures. Specifically, when $d \approx D_{\max}$, both $\alpha\lambda$ and $\bar{\alpha}\lambda$ are large. From (20) and (29), we see that $d_0 \approx D_{\max}$, and from (13), (15), (16) and (23), (24), (25) we see that $d_k \approx 0$ for $k \neq 0$. We are interested, however, in finding the cells that dominate the rate at which distortion converges to $D_{\max}$. Since $d_0 \to D_{\max}$ and $d_k \to 0$, $k \neq 0$, it makes most sense to compare $D_{\max} - d_0$ and the sum of the $d_k$'s, $k \neq 0$. Comparing (20) to (13), (15), and (16), and (29) to (23), (24), and (25), reveals that $\sum_{k \neq 0} d_k$ is asymptotically negligible relative to $D_{\max} - d_0$. We conclude that when $d \approx D_{\max}$, $D_{\max} - d_0$ is the dominant component of $D_{\max} - d$.

# 5 Asymptotic Rate-Distortion

Directly applying Theorems 7, 8 and 9 yields the following lemma, which is used in showing Theorem 13 below.

**Lemma 12** *For a UTQ with offset $\alpha$, $0 < \alpha < 1$, normalized step size $\lambda$, optimal reconstruction levels, and a $\mathcal{N}(0,\sigma^2)$ or $\mathcal{L}(0,\sigma^2)$ source,*

$$\lim_{\alpha\lambda, \bar{\alpha}\lambda \to \infty} \frac{H(\alpha,\lambda)}{D_{\max} - d(\alpha,\Delta,\sigma^2)} = \begin{cases} \frac{\log e}{2\sigma^2}, & \textit{for } \mathcal{N}(0,\sigma^2) \textit{ and squared error} \\ 0, & \textit{for } \mathcal{L}(0,\sigma^2) \textit{ source squared error} \\ \infty, & \textit{for } \mathcal{N}(0,\sigma^2) \textit{ and absolute error} \\ \frac{\log e}{D_{\max,1}}, & \textit{for } \mathcal{L}(0,\sigma^2) \textit{ and absolute error} \end{cases}.$$

The following is one of the principal results in this report.

**Theorem 13** *The operational rate-distortion functions of infinite-level uniform threshold scalar quantization for Gaussian and Laplacian sources with variance $\sigma^2$ satisfy*

$$\lim_{D \to D_{\max}} \frac{R_{U,\sigma^2}(D)}{D_{\max} - D} = \begin{cases} \frac{\log e}{2\sigma^2}, & \textit{for } \mathcal{N}(0,\sigma^2) \textit{ and squared error} \\ 0, & \textit{for } \mathcal{L}(0,\sigma^2) \textit{ source squared error} \\ \infty, & \textit{for } \mathcal{N}(0,\sigma^2) \textit{ and absolute error} \\ \frac{\sqrt{2}\log e}{\sigma}, & \textit{for } \mathcal{L}(0,\sigma^2) \textit{ and absolute error} \end{cases}.$$

*Proof:* Since $R_{U,\sigma^2}(D) = R_{U,1}\big(\frac{D}{D_{\max}}\big)$ it suffices to show

$$\lim_{D \to D^*_{\max}} \frac{R_{U,1}(D)}{D^*_{\max} - D} = \begin{cases} \frac{\log e}{2}, & \text{for } \mathcal{N}(0,1) \text{ and squared error} \\ 0, & \text{for } \mathcal{L}(0,1) \text{ and squared error} \\ \infty, & \text{for } \mathcal{N}(0,1) \text{ and absolute error} \\ \sqrt{2}\log e, & \text{for } \mathcal{L}(0,1) \text{ and absolute error} \end{cases} . \tag{30}$$

As shorthand we let $\Gamma$ denote the quantity on the right hand side of (30), which, of course, depends on the source and distortion measure.

Next, we rewrite the operational rate-distortion function as

$$R_{U,1}(D) = \inf_{0<\alpha<1} R_{U,1,\alpha}(D) \ ,$$

where $R_{U,1,\alpha}(D) \stackrel{\Delta}{=} \inf_{\Delta>0:d(\alpha,\Delta,1)\leq D} H(\alpha,\Delta)$ is the operational rate-distortion function of UTQ with fixed offset $\alpha$, and as mentioned before, $\alpha = 0$ can be omitted from the constraint since $D$ is sufficiently large. As a preliminary to showing (30) we will show $R_{U,1,\alpha}(D)$ satisfies (30) for any fixed $\alpha \in (0,1)$. For both sources and distortion measures we have

$$\limsup_{D \to D^*_{\max}} \frac{R_{U,1,\alpha}(D)}{D^*_{\max} - D} \stackrel{(a)}{=} \limsup_{\lambda \to \infty} \frac{R_{U,1,\alpha}\big(d(\alpha,\lambda,1)\big)}{D^*_{\max} - d(\alpha,\lambda,1)} \stackrel{(b)}{\leq} \limsup_{\lambda \to \infty} \frac{H(\alpha,\lambda)}{D^*_{\max} - d(\alpha,\lambda,1)} \ , \tag{31}$$

where $(a)$ derives from the fact that $d(\alpha,\lambda,1)$ goes continuously to $D^*_{\max}$ as $\lambda \to \infty$, and $(b)$ follows from the definition of $R_{U,1,\alpha}\big(d(\alpha,\lambda,1)\big)$. Next,

$$\liminf_{D \to D^*_{\max}} \frac{R_{U,1,\alpha}(D)}{D^*_{\max} - D} \geq \liminf_{\lambda \to \infty} \frac{H(\alpha,\lambda)}{D^*_{\max} - d(\alpha,\lambda,1)} \ , \tag{32}$$

where the inequality is shown as follows. By the definition of $R_{U,1,\alpha}(D)$, for any $D \in (0, D^*_{\max})$, there exists $\lambda(D)$ such that

$$H(\alpha,\lambda(D)) \leq R_{U,1,\alpha}(D) + \varepsilon(D) \quad \text{and} \quad d(\alpha,\lambda(D),1) \leq D \ , \tag{33}$$

where $\varepsilon(D) > 0$ and $\lim_{D \to D^*_{\max}} \frac{\varepsilon(D)}{D^*_{\max} - D} = 0$. (The choices of $\varepsilon(D)$ and $\lambda(D)$ are not unique, but any fixed choices will do.) As shown below, $R_{U,1,\alpha}(D) \to 0$ as $D \to D^*_{\max}$. Thus, by (33) $H(\alpha,\lambda(D)) \to 0$ as $D \to D^*_{\max}$, which implies that $\lambda(D) \to \infty$ as $D \to D^*_{\max}$, since $H(\alpha,\lambda) \to 0$ if and only if $\lambda \to \infty$. This and (33) yield

$$\liminf_{D \to D^*_{\max}} \frac{R_{U,1,\alpha}(D)}{D^*_{\max} - D} \geq \liminf_{D \to D^*_{\max}} \frac{H(\alpha,\lambda(D)) - \varepsilon(D)}{D^*_{\max} - d(\alpha,\lambda(D),1)} \geq \liminf_{\lambda \to \infty} \frac{H(\alpha,\lambda)}{D^*_{\max} - d(\alpha,\lambda,1)} \ .$$

It remains to show that indeed $R_{U,1,\alpha}(D) \to 0$ as $D \to D^*_{\max}$. Since $d(\alpha,\lambda,1)$ goes continuously from 0 to $D^*_{\max}$ as $\lambda$ goes from 0 to $\infty$, the mean value theorem implies that for any $D \in (0, D^*_{\max})$, there exists $\tilde{\lambda}(D)$ such that $d(\alpha, \tilde{\lambda}(D), 1) = D$ (the choice of $\tilde{\lambda}(D)$ may or may not be unique, but any fixed choice will do). Since $d(\alpha,\lambda,1) \to D^*_{\max}$ if and only if $\lambda \to \infty$, it follows that $\tilde{\lambda}(D) \to \infty$ as $D \to D^*_{\max}$. Therefore,

$$\limsup_{D \to D^*_{\max}} R_{U,1,\alpha}(D) \;\le\; \limsup_{D \to D^*_{\max}} \inf_{\lambda : d(\alpha,\lambda,1) \le D} H(\alpha,\lambda) \;\le\; \limsup_{D \to D^*_{\max}} H(\alpha, \tilde{\lambda}(D)) \;\le\; \limsup_{\lambda \to \infty} H(\alpha,\lambda) \;=\; 0\,. \tag{34}$$

It now follows from (31), (32), and Lemma 12 that for any $\alpha \in (0,1)$

$$\lim_{D \to D^*_{\max}} \frac{R_{U,1,\alpha}(D)}{D^*_{\max} - D} \;=\; \Gamma(G,L)\;. \tag{35}$$

Finally, to obtain the result of the theorem we proceed as follows:

$$\limsup_{D \to D^*_{\max}} \frac{R_{U,1}(D)}{D^*_{\max} - D} \;\overset{(a)}{=}\; \limsup_{D \to D^*_{\max}} \frac{\inf_\alpha R_{U,1,\alpha}(D)}{D^*_{\max} - D} \;\overset{(b)}{\le}\; \inf_\alpha \limsup_{D \to D^*_{\max}} \frac{R_{U,1,\alpha}(D)}{D^*_{\max} - D} \;\overset{(c)}{=}\; \Gamma(G,L)\;. \tag{36}$$

where $(a)$ follows from the definition of $R_{U,1}(D)$, $(b)$ is elementary, and $(c)$ is obtained from (35).

Next, we follow similar steps to those used in showing (32). Specifically, for any $D \in (0, D^*_{\max})$ there exists $\alpha(D)$ and $\lambda(D)$ such that

$$H(\alpha(D), \lambda(D)) \;\le\; R_{U,1}(D) + \varepsilon(D) \quad \text{and} \quad d(\alpha(D), \lambda(D), 1) \;\le\; D\;, \tag{37}$$

where $\varepsilon(D) > 0$ and $\lim_{D \to D^*_{\max}} \frac{\varepsilon(D)}{D^*_{\max} - D} = 0$, and as before, $\varepsilon(D)$, $\alpha(D)$ and $\lambda(D)$ are not unique, but any fixed choices will do. From (34) we have that $R_{U,1,\alpha}(D) \to 0$ as $D \to D_{\max}$, which implies $R_{U,1}(D) \to 0$ as $D \to D_{\max}$. Furthermore, we have that $\varepsilon(D) \to 0$ as $D \to D^*_{\max}$. Therefore, $H(\alpha(D), \lambda(D)) \to 0$ as $D \to D^*_{\max}$, which implies that $\alpha(D)\lambda(D) \to \infty$ and $(1 - \alpha(D))\lambda(D) \to \infty$ as $D \to D^*_{\max}$. Combining this with (37) we obtain

$$\liminf_{D \to D^*_{\max}} \frac{R_{U,1}(D)}{D^*_{\max} - D} \;\ge\; \liminf_{D \to D^*_{\max}} \frac{H(\alpha(D), \lambda(D)) - \varepsilon(D)}{D^*_{\max} - d(\alpha(D), \lambda(D), 1)} \;\ge\; \liminf_{\alpha\lambda, \bar{\alpha}\lambda \to \infty} \frac{H(\alpha,\lambda)}{D^*_{\max} - d(\alpha,\lambda,1)} \;=\; \Gamma(G,L)\,, \tag{38}$$

where the last equality follows from Lemma 12. Equations (36) and (38) show (30) and the theorem.

We make two notes. First (32) and (38) could have been shown for the cases of a Gaussian source and squared error distortion measure, and a Laplacian source and absolute error distortion measure using Shannon's rate-distortion function as a lower bound. However, the approach taken above demonstrates that $\lim_{D \to D_{\max}} \frac{R_{U,1,\alpha}(D)}{D_{\max} - D} = \lim_{\lambda \to \infty} \frac{H(\alpha,\lambda)}{D_{\max} - d(\alpha,\lambda,1)}$ without using either the source density or Shannon's rate-distortion function. It requires only that the latter limit exist.

Furthermore, it works also for the other two cases for which the Shannon rate-distortion function is not known.

Secondly, one could have, in fact, skipped (32) altogether and show (38) directly, however, (32) is needed in showing (35), which provides the operational rate-distortion function of uniform threshold quantization with offset $\alpha \neq 0$. Additionally, from (35) and the relation between $R_{U,\sigma^2,\alpha}(D)$ and $R_{U,1,\alpha}(D)$, one concludes that in low resolution, for any $\alpha \neq 0$, the operational rate-distortion function $R_{U,\sigma^2,\alpha}(D)$ of uniform threshold quantization with offset $\alpha$ is the same as the operational rate-distortion function of uniform quantization in general, for both sources and distortion measures considered. □

We comment that from the dominance results presented previously, the slope of the operational rate distortion functions of uniform threshold quantization is approximately equal to $\frac{H(P_{-1})+H(P_1)}{D_{\max}-D_0}$, i.e. the distortion term is dominated by the center cell and the entropy is dominated by the two adjacent cells. This holds for all four cases.

We now consider how the operational rate-distortion function of uniform scalar quantization compares to that of arbitrary unconstrained scalar quantization.

**Theorem 14** *At $D = D_{\max}$, the slope of the operational rate distortion function of uniform scalar quantization equals that of arbitrary unconstrained scalar quantization, for Gaussian and Laplacian sources with both squared and absolute error distortion measure.*

*Proof:* For three of the four possible cases we show that the slope at $D = D_{\max}$ of the operational rate-distortion function of uniform scalar quantizers matches the slope of the corresponding Shannon rate-distortion function, which implies the theorem for these cases. Specifically, for a Gaussian source with squared error distortion measure, and a Laplacian source with absolute error distortion measure, we have $\mathcal{R}^G_{2,\sigma^2}(D) = \frac{1}{2}\log\frac{\sigma^2}{D} = \frac{\log e}{2}\left[1-\frac{D}{\sigma^2}\right]\left[1+o_{_{D\to\sigma^2}}\right]$ and $\mathcal{R}^L_{1,\sigma^2}(D) = \log\frac{\sqrt{2}D}{\sigma} = (\log e)\left[1-\frac{D}{D^L_{\max,1}}\right]\left[1+o_{_{D\to D^L_{\max,1}}}\right]$, respectively. Thus, for these sources and distortion measures the slopes at $D = D_{\max}$ of the operational rate-distortion functions of uniform scalar quantizers, as shown in Theorem 13, equal the slopes of the corresponding Shannon rate-distortion functions. For the case of a Laplacian source and squared error distortion measure, the slope of $R^L_{2,U,\sigma^2}(D)$ at $D = D_{\max}$ equals 0, as shown in Theorem 13, and thus must equal the slope of the corresponding Shannon rate-distortion function (because the magnitudes of the latter could be no larger).

It remains to show the theorem for the case of a Gaussian source and absolute error distortion measure. It suffices to show that scalar quantizers with contiguous cells can do no better than uniform scalar quantizers, as follows from [20]. By definition of $R_{\sigma^2}(D)$, for any $D \in (0, D_{\max})$ there exists a quantizer $q_D$ such that

$$H(q_D) \;\leq\; R_{\sigma^2}(D) + \varepsilon(D) \quad \text{and} \quad d(q_D) \leq D \;, \tag{39}$$

where $\varepsilon(D) > 0$ and $\lim_{D \to D_{\max}} \frac{\varepsilon(D)}{D_{\max} - D} = 0$. (As before, the choices of $q_D$ and $\varepsilon(D)$ are not unique, but any fixed choices will do.) Let $S_{o,D} = (-A_D, B_D)$, denote the cell of $q_D$ containing the origin (it is immaterial if the cell is open or closed on either side). As $D \to D_{\max}$, $A_D, B_D \to \infty$. Note that either $A_D$ or $B_D$ (but not both simultaneously) might be infinite. Let $D_{o,D}$ be the contribution to distortion of cell $S_{o,D}$, where the reconstruction level of $S_{o,D}$ lies at the median of the cell. Repeating similar steps to those in (26) – (28), we obtain

$$D_{\max} - D_{o,D} \;=\; \sigma\Big(G(\frac{A_D}{\sigma}) + G(\frac{B_D}{\sigma})\Big)\,[1 + o_{A_D,B_D}] \;. \tag{40}$$

Next, applying Lemma 5 we obtain

$$\mathcal{H}\big(Q(\frac{A_D}{\sigma})\big) + \mathcal{H}\big(Q(\frac{B_D}{\sigma})\big) \;=\; \frac{\log e}{2}\Big(\frac{A_D}{\sigma}G(\frac{A_D}{\sigma}) + \frac{B_D}{\sigma}G(\frac{B_D}{\sigma})\Big)\,[1 + o_{A_D,B_D}] \;. \tag{41}$$

Finally, we have that

$$\liminf_{D \to D_{\max}} \frac{R_{\sigma^2}(D)}{D_{\max} - D} \overset{(a)}{\geq} \liminf_{D \to D_{\max}} \frac{H(q_D) - \varepsilon(D)}{D_{\max} - d(q_D)} \overset{(b)}{\geq} \liminf_{D \to D_{\max}} \frac{\mathcal{H}\big(Q(\frac{A_D}{\sigma})\big) + \mathcal{H}\big(Q(\frac{B_D}{\sigma})\big)}{D_{\max} - D_{o,D}} \overset{(c)}{=} \infty \;,$$

where $(a)$ follows from (39), $(b)$ is due to an elementary property of entropy and from having $D_{o,D} \leq d(q_D)$, and $(c)$ derives from (40) and (41). Thus, $\lim_{D \to D_{\max}} \frac{R_{\sigma^2}(D)}{D_{\max} - D} = \liminf_{D \to D_{\max}} \frac{R_{\sigma^2}(D)}{D_{\max} - D} = \infty$, as needed to show. □

We now make a few observations. First, the expressions for the Shannon rate-distortion functions provided in the proof of Theorem 14 together with the theorem statement, imply that in the low resolution region, for three of the four cases considered (i.e. all cases except for a Gaussian source and absolute error distortion measure), scalar quantization is asymptotically as good as any quantization technique — scalar, block, or otherwise. Since Theorem 14 shows that uniform quantization is as good as any kind of scalar quantization for Gaussian and Laplacian sources and both distortion measures, in particular it is asymptotically as good as any quantization technique for the stated three out of the four cases. Secondly, for the case of a Gaussian source and absolute error distortion measure, the slope of $R^G_{1,\sigma^2}(D)$ at $D = D^G_{\max,1}$ is infinite, whereas the slope of $\mathcal{R}^G_1(D)$

is finite (because it is convex). Therefore, in low rate for this case, scalar quantization is nowhere near as good as the best high-dimensional vector quantizers. Furthermore, in this case it can be shown directly (with no use of uniform quantizers) that the operational rate-distortion function of arbitrary unconstrained scalar quantization has infinite slope at $D = D_{\max}$ [18], and so of course it is infinite for uniform and binary quantizers as well. Finally, Theorem 14 implies that the slope of the Shannon rate-distortion function for a Laplacian source and squared error distortion measure (the function itself is not known) is zero at rate equal zero (this was also shown in [1]).

# 6 Binary Quantizers

A binary (two-level) scalar quantizer is characterized by three numbers: a threshold $t$ and two reconstruction levels $r_0 < t$ and $r_1 \geq t$. Let $S_0(t) = (-\infty, t)$ and $S_1(t) = [t, \infty)$ be the two quantization cells, and let the quantization rule be $q(x) = r_k$ when $x \in S_k$, $k = 0, 1$. As in the case of uniform quantizers, the reconstruction levels $r_0$ and $r_1$ are taken to be the cell centroids or cell medians, depending on whether squared or absolute error distortion measure is considered, respectively. We let $P_k$ or $P_k(t, \sigma^2)$ denote the probability of the source value lying in $S_k$, $k = 0, 1$.

Let $H(t, \sigma^2) = \mathcal{H}(P_0(t, \sigma^2), P_1(t, \sigma^2))$ denote the entropy of the quantizer output with threshold $t$ for either the Gaussian or Laplacian source. Let the mean-squared or absolute error of this quantizer be $d_s(t, \sigma^2) \triangleq \int_{-\infty}^{\infty} |x - q(x)|^s f(x)\, dx$, where $s \in \{1, 2\}$, and $f$ is the density of the source. The operational rate-distortion function of binary quantization is $R_{B,\sigma^2}(D) = \inf_{t: d(t,\sigma^2) \leq D} H(t, \sigma^2)$, which specifies the least entropy of any such quantizer with distortion $D$ or less.

It is easy to see that $P_k(t, \sigma^2) = P_k(\frac{t}{\sigma}, 1)$, which implies that $H(t, \sigma^2) = H(\frac{t}{\sigma}, 1)$. Thus, it is convenient to parameterize $H$ by $\lambda = \frac{t}{\sigma}$, i.e. $H(\lambda)$. In this section when we refer to the *normalized step size* $\lambda$, we will be referring to $\frac{t}{\sigma}$. Furthermore, we have $d_2(t, \sigma^2) = \sigma^2 d_2(\frac{t}{\sigma}, 1)$, $d_1(t, \sigma^2) = \sigma d_1(\frac{t}{\sigma}, 1)$, $R_{2,B,\sigma^2}(D) = R_{2,B,1}(\frac{D}{\sigma^2})$ and $R_{1,B,\sigma^2}(D) = R_{1,B,1}(\frac{D}{\sigma})$. Due to the symmetry of the sources, it suffices to restrict attention to $t > 0$.

As in the case of uniform quantizers, we find asymptotic low resolution approximations to entropy and distortion, and then combine these to determine the asymptotic low resolution expression for $R_{B,\sigma^2}$(D). We also determine which cells dominate entropy and distortion Since the derivations in the binary case are similar in spirit to those in the uniform case, we will only state the results and provide no proofs, so as to spare the reader repetitive details.

**Theorem 15** *For a binary scalar quantizer with normalized step size $\lambda$ applied to a $\mathcal{N}(0,\sigma^2)$ source*

$$H(\lambda) \;=\; \frac{\log e}{2}\lambda\, G(\lambda)\,\big[1+o_\lambda\big],$$

*and applied to a $\mathcal{L}(0,\sigma^2)$ source*

$$H(\lambda) \;=\; (\log e)\big(\lambda\, L(\lambda)\big)\,\big[1+o_\lambda\big]\;.$$

**Theorem 16** *For a binary scalar quantizer with normalized step size $\lambda$ and centroid reconstruction levels applied to a $\mathcal{N}(0,\sigma^2)$ source*

$$\frac{\sigma^2 - d(t,\sigma^2)}{\sigma^2} \;=\; \lambda\, G(\lambda)\,\big[1+o_\lambda\big]\;,$$

*and applied to a $\mathcal{L}(0,\sigma^2)$ source*

$$\frac{\sigma^2 - d(t,\sigma^2)}{\sigma^2} \;=\; \frac{\lambda^2\, L(\lambda)}{\sqrt{2}}\,\big[1+o_{\alpha\lambda,\bar{\alpha}\lambda}\big]\;.$$

**Theorem 17** *For a binary scalar quantizer with normalized step size $\lambda$ and median reconstruction levels applied to a $\mathcal{N}(0,\sigma^2)$ source*

$$\frac{D_{\max} - d(t,\sigma^2)}{D_{\max}} \;=\; \frac{\pi}{\sqrt{2}}G(\lambda)\,\big[1+o_\lambda\big]\;,$$

*and applied to a $\mathcal{L}(0,\sigma^2)$ source*

$$\frac{D_{\max} - d(t,\sigma^2)}{D_{\max}} \;=\; \frac{\lambda\, L(\lambda)}{\sqrt{2}}\,\big[1+o_{\alpha\lambda,\bar{\alpha}\lambda}\big]\;.$$

**Theorem 18** *The operational rate-distortion functions of binary scalar quantization for Gaussian and Laplacian sources with variance $\sigma^2$ satisfy*

$$\lim_{D\to D_{\max}} \frac{R_{B,\sigma^2}(D)}{D_{\max}-D} \;=\; \begin{cases} \frac{\log e}{2\sigma^2}, & \textit{for } \mathcal{N}(0,\sigma^2) \textit{ and squared error} \\ 0, & \textit{for } \mathcal{L}(0,\sigma^2) \textit{ source squared error} \\ \infty, & \textit{for } \mathcal{N}(0,\sigma^2) \textit{ and absolute error} \\ \frac{\sqrt{2}\log e}{\sigma}, & \textit{for } \mathcal{L}(0,\sigma^2) \textit{ and absolute error} \end{cases}\;.$$

Notice that the expressions given in this theorem for binary quantization are precisely the same as those given in Theorem 13 for infinite-level uniform threshold quantization, three of which match the Shannon rate-distortion function in the low resolution region. We conclude that binary quantization is another type of quantization that is asymptotically optimal in the low resolution region for Gaussian sources and squared error distortion measure, and for Laplacian sources and squared and absolute error distortion measures.

We now comment on the cells that dominate the entropy and distortion. As before, when entropy is small, it is dominated by the cell that has largest probability not close to one, which is $S_2$. And just as with uniform quantizers, when distortion is close to $D_{\max}$, $D_{\max} - d$ is dominated by the cell whose probability is nearly one, namely, $S_1$. That is, $\frac{D_{\max}-d_1}{D_{\max}-d} \approx 1$.

# 7 Conclusions

This report considered the asymptotic performance of scalar quantizers in the low resolution domain, which is determined by the slope of the operational rate-distortion function of such quantizers at rate equal zero. For the cases of exponential, Laplacian and uniform sources and difference distortion measures, this slope has been provided in or can be determined from [1, 2]. The focus of this report has been on Gaussian and Laplacian sources with squared and absolute error distortion measures. Although the case of a Laplacian source has been provided in [1], the method employed here proved useful in a wider context, as it works for both Gaussian and Laplacian sources.

We considered infinite-level uniform threshold and binary scalar quantizers with asymptotic low rate, namely as cell sizes tend to infinity (for the uniform case) and as quantizer threshold tends to infinity (for the binary case). We derived simple formulas for the rate of convergence of entropy to zero and of distortion to $D_{\max}$.

The convergence of entropy and distortion as $\lambda \to \infty$ for uniform quantization is not uniform in the offset $\alpha$. The derived formulas show how entropy and distortion depend on $\alpha$ as well as $\lambda$. Specifically, they provide accurate approximations when both $\alpha\lambda$ and $\bar{\alpha}\lambda$ are large.

Using these convergence formulas, the slopes with which the operational rate-distortion functions of infinite-level uniform threshold and binary scalar quantization, for both sources and distortion measures, approach zero as $D \to D_{\max}$ have been determined. Furthermore, it has been shown that the operational rate-distortion functions of scalar quantization in general have the same slopes. Thus, in low resolution, optimal uniform and binary quantizers have the same performance as optimal scalar quantizers of any form. Additionally, these slopes are the same as the slopes of the Shannon rate-distortion functions in the cases of a Gaussian source with squared error distortion measure and a Laplacian source with both distortion measures. Therefore, in low resolution, scalar quantization is an asymptotically optimal coding technique for these cases. However, for a Gaussian source with absolute error distortion measure, scalar quantization is nowhere near optimal.

Finally, we determined that the slope of the Shannon rate-distortion function for a Laplacian source and squared error distortion measure (the function itself is not known) is zero at rate equal zero (this was also shown in [1]).

# APPENDIX

**Proof of Lemma 1:**

We will show Part $A$; Part $B$ follows by symmetry. To simplify notation, we omit the parameters $\alpha$ and $\lambda$ from $P_k(\alpha,\lambda)$. We first consider the result for a Laplacian source, for which for $k \geq 1$,

$$\frac{P_{k+1}}{P_k} = \frac{L\big((k+1-\alpha)\lambda\big) - L\big((k+2-\alpha)\lambda\big)}{L\big((k-\alpha)\lambda\big) - L\big((k+1-\alpha)\lambda\big)} = e^{-\sqrt{2}\lambda} \ .$$

Next, for all sufficiently large $\lambda$

$$P_1 = \frac{L\big((1-\alpha)\lambda\big) - L\big((2-\alpha)\lambda\big)}{\sqrt{2}} = \frac{1}{2} e^{-\sqrt{2}(1-\alpha)\lambda}\,[1 - e^{-\sqrt{2}\lambda}] > \frac{1}{4} e^{-\sqrt{2}(1-\alpha)\lambda} \ ,$$

where the inequality follows from having $1 - e^{-\sqrt{2}\lambda} > \frac{1}{2}$ for all sufficiently large $\lambda$.

Finally, to show that $\frac{P_k+1}{P_k} < P_1$ for all sufficiently large $\alpha\lambda$ for all $k \geq 1$, it suffices to show that $e^{-\sqrt{2}\lambda} < \frac{1}{4} e^{-\sqrt{2}(1-\alpha)\lambda}$, or equivalently, that $e^{\sqrt{2}\alpha\lambda} > 4$ for all sufficiently large $\alpha\lambda$, which clearly is the case. This shows part $A$ of the lemma for a Laplacian source.

Next, we show part $A$ for a Gaussian source. Consider $k \geq 1$. First, Fact G8, with $(k-\alpha)$ playing the role of $x$, shows that for all sufficiently large $\lambda$, the following lower bound to $P_k$ holds for all $k \geq 1$:

$$P_k = Q\big((k-\alpha)\lambda\big) - Q\big((k+1-\alpha)\lambda\big) > \begin{cases} \frac{1}{4}\frac{1}{(k-\alpha)\lambda}\, G((k-\alpha)\lambda), & (k-\alpha)\lambda \geq \sqrt{2} \qquad (a) \\ \frac{Q(\sqrt{2})}{2}, & (k-\alpha)\lambda < \sqrt{2} \qquad (b) \end{cases} . \qquad \text{(A1)}$$

Next, we upper bound $P_{k+1}$ using Fact G2:

$$P_{k+1} = Q\big((k+1-\alpha)\lambda\big) - Q\big((k+2-\alpha)\lambda\big) < \frac{1}{(k+1-\alpha)\lambda}\, G((k+1-\alpha)\lambda) \ .$$

Combining the lower bound to $P_k$ with the upper bound to $P_{k+1}$, we obtain

$$\frac{P_{k+1}}{P_k} < \begin{cases} 4e^{-\frac{(2(k-\alpha)+1)\lambda^2}{2}}, & (k-\alpha)\lambda \geq \sqrt{2} \qquad (a) \\ \frac{2\,G((k+1-\alpha)\lambda)}{Q(\sqrt{2})(k+1-\alpha)\lambda}, & (k-\alpha)\lambda < \sqrt{2} \qquad (b) \end{cases} . \qquad \text{(A2)}$$

It now suffices to show that for all sufficiently large $\lambda$, the above upper bound to $\frac{P_{k+1}}{P_k}$ is smaller than the lower bound to $P_1$ obtained from (A1). We do so by considering two cases.

*Case 1:* $(k-\alpha)\lambda < \sqrt{2}$ — In this case, $(1-\alpha)\lambda < \sqrt{2}$. Thus, by (A1b), $P_1 > \frac{Q(\sqrt{2})}{2}$. Next, by (A2b), $\frac{P_{k+1}}{P_k} < \frac{2\,G((k+1-\alpha)\lambda)}{Q(\sqrt{2})(k+1-\alpha)\lambda} < \frac{2\,G(\lambda)}{Q(\sqrt{2})\lambda}$, where the last inequality uses $k+1-\alpha > 1$. Since $P_1 > \frac{Q(\sqrt{2})}{2}$ and $\frac{P_{k+1}}{P_k} < \frac{2\,G(\lambda)}{Q(\sqrt{2})\lambda} \to 0$ as $\lambda \to \infty$, we conclude that for all sufficiently large $\lambda$, $\frac{P_{k+1}}{P_k} < P_1$, for all $k, \alpha$ such that $(k-\alpha)\lambda < \sqrt{2}$.

*Case 2:* $(k-\alpha)\lambda \geq \sqrt{2}$ — We consider two subcases. First, suppose $(1-\alpha)\lambda < \sqrt{2}$. Then by (A1b), $P_1 > \frac{Q(\sqrt{2})}{2}$. Next, by (A2a), $\frac{P_{k+1}}{P_k} < 4e^{-\frac{(2(k-\alpha)+1)\lambda^2}{2}} < 4e^{-\frac{\lambda^2}{2}}$. We conclude that for all sufficiently large $\lambda$, $\frac{P_{k+1}}{P_k} < P_1$, for all all $k, \alpha$ such that $(k-\alpha)\lambda \geq \sqrt{2}$ and $(1-\alpha)\lambda < \sqrt{2}$. Next, suppose $(1-\alpha)\lambda \geq \sqrt{2}$. Then by (A1a), $P_1 > \frac{1}{4}\frac{1}{(1-\alpha)\lambda}\,G((1-\alpha)\lambda) > \frac{1}{4}\frac{1}{\lambda}\,G(\lambda)$, using $1-\alpha < 1$. By (A2a), $\frac{P_{k+1}}{P_k} < 4e^{-\frac{(2(k-\alpha)+1)\lambda^2}{2}} < 4e^{-\frac{\lambda^2}{2}}e^{-\sqrt{2}\lambda}$, using $(k-\alpha)\lambda \geq \sqrt{2}$. Since $e^{-\sqrt{2}\lambda} \to 0$ faster than $\frac{1}{\lambda} \to 0$, we have that for all sufficiently large $\lambda$, $\frac{P_{k+1}}{P_k} < P_1$, for all $k, \alpha$ such that $(k-\alpha)\lambda \geq \sqrt{2}$ and $(1-\alpha)\lambda \geq \sqrt{2}$. This completes the proof of Part $A$ for a Gaussian source and of the lemma. □

**Proof of Lemma 2:**

We need to show that $\lim_{p\to 0} \frac{-(1-p+p\,o_{p\to 0})\ln(1-p+p\,o_{p\to 0})}{-p\ln p} = 0$. The fact that $\lim_{x\to 0}\frac{\ln(1-x)}{-x} = 1$, or equivalently, that $\frac{\ln(1-x)}{-x} = 1 + o_{x\to 0}$, implies

$$\begin{aligned}
&\frac{-(1-p+p\,o_{p\to 0})\ln(1-p+p\,o_{p\to 0})}{-p\ln p} \\
&\quad= \left[\frac{-(1-p+p\,o_{p\to 0})\ln(1-p+p\,o_{p\to 0})}{(1-p+p\,o_{p\to 0})(p+p\,o_{p\to 0})}\right]\cdot\left[\frac{(1-p+p\,o_{p\to 0})(p+p\,o_{p\to 0})}{-p\ln p}\right] \\
&\quad= \left[1+o_{p\to 0}\right]\cdot\left[\left(1-p+p\,o_{p\to 0}\right)\frac{1+o_{p\to 0}}{-\ln p}\right] \longrightarrow 0 \ \text{ as } p\to 0\ ,
\end{aligned}$$

which proves the lemma. □

**Proof of Lemma 4:**

It is sufficient to show $\lim_{s\to s_o}\frac{\mathcal{H}(a(s))}{\mathcal{H}(b(s))} = 1$. We have the following string of equalities.

$$\frac{\mathcal{H}(a(s))}{\mathcal{H}(b(s))} = \frac{-a(s)\log a(s)}{-b(s)\log b(s)} = \frac{a(s)}{b(s)}\frac{\log\left[\frac{a(s)}{b(s)}b(s)\right]}{\log b(s)} = \frac{a(s)}{b(s)}\left[1+\frac{\log\frac{a(s)}{b(s)}}{\log b(s)}\right] = \frac{a(s)}{b(s)} + \frac{\frac{a(s)}{b(s)}\log\frac{a(s)}{b(s)}}{\log b(s)}\ .$$

Since $|b(s)-1| > \varepsilon$ for all $s$, it follows that either $\log b(s) > \log(1+\varepsilon)$ or $\log b(s) < \log(1-\varepsilon)$ for all $s$. Therefore, $\log b(s)$ is bounded away from zero. Combining this with the fact that $\frac{a(s)}{b(s)} \to 1$ as $s\to\infty$, and that $\frac{a(s)}{b(s)}\log\frac{a(s)}{b(s)} \to 0$ as $s\to\infty$, the result follows. □

**Proof of Lemma 5:**

The lemma is obtained by using Fact G5 in the first equality below

$$\begin{aligned}\mathcal{H}\big(Q(x)\big) &= \mathcal{H}\big(\frac{1}{x}\,G(x)\,\big[1+o_x\big]\big) = -\,\frac{1}{x}\,G(x)\,\big[1+o_x\big]\log\Big(\frac{1}{x}\,G(x)\,\big[1+o_x\big]\Big)\\ &= \frac{1}{x}\,G(x)\,\big[1+o_x\big]\Big[\log\sqrt{2\pi}x+\frac{x^2}{2}\log e-\log\big[1+o_x\big]\Big] = \frac{\log e}{2}x\,G(x)\,\big[1+o_x\big]\ .\end{aligned}$$

□

**Proof of Lemma 6:**

The lemma statement is relevant when $x \to \infty$, thus we may assume $x > 0$, and obtain.

$$\begin{aligned}\mathcal{H}\Big(\frac{L(x)}{\sqrt{2}}\Big) &= \mathcal{H}\Big(\frac{1}{2}e^{-\sqrt{2}x}\Big) = -\,\frac{1}{2}e^{-\sqrt{2}x}\big[\log\frac{1}{2}+\log e^{-\sqrt{2}x}\big] = \frac{1}{2}e^{-\sqrt{2}x}\big[\sqrt{2}x(\log e)+1\big]\\ &= xL(x)(\log e)\big[1+o_x\big]\end{aligned}$$

□

**Lemma A1** *If* $Q(b) = \frac{Q(a)[1+o_a]}{2}$, *then* $b = a[1+o_a]$.

*Proof:* What needs to be shown is that $\lim_{a\to\infty}\frac{b}{a} = 1$, where one thinks of $b$ as being a function of $a$, i.e. $b = b(a)$. We show this by contradiction. Specifically, suppose the limit is not 1 (in particular it may not exist). Then there exists $\varepsilon > 0$ such that $\liminf_{a\to\infty}\frac{b}{a} \geq 1+\varepsilon$ or $\limsup_{a\to\infty}\frac{b}{a} \leq 1-\varepsilon$. Before considering both cases, note that

$$\frac{Q(a)[1+o_a]}{2Q(b)} = \frac{\frac{1}{a}\frac{1}{\sqrt{2\pi}}e^{-\frac{a^2}{2}}[1+o_a]}{\frac{2}{b}\frac{1}{\sqrt{2\pi}}e^{-\frac{b^2}{2}}[1+o_b]} = \frac{b}{2a}e^{\frac{a^2(\frac{b^2}{a^2}-1)}{2}}[1+o_a] \longrightarrow 1 \quad \text{as} \quad a \longrightarrow \infty\ , \tag{A3}$$

where the first equality derives from Fact G5, and the second equality uses the easily seen fact that $a \to \infty$ if and only $b \to \infty$. Using the above equation we now have that either

$$\liminf_{a\to\infty}\frac{Q(a)[1+o_a]}{2Q(b)} \geq \frac{(1+\varepsilon)}{2}\liminf_{a\to\infty}e^{\frac{a^2\varepsilon(\varepsilon+2)}{2}} = \infty\ ,$$

or

$$\limsup_{a\to\infty}\frac{Q(a)[1+o_a]}{2Q(b)} \leq \frac{(1-\varepsilon)}{2}\limsup_{a\to\infty}e^{-\frac{a^2\varepsilon(2-\varepsilon)}{2}} = 0\ ,$$

where the last equality is due to having $0 < \varepsilon \leq 1$, where $\varepsilon \leq 1$ derives from the fact that as $a \to \infty$ also $b \to \infty$, from which it follows that $0 \leq \limsup_{a\to\infty}\frac{b}{a}$. Combining this with the given fact that $\limsup_{a\to\infty}\frac{b}{a} \leq 1-\varepsilon$ implies that $\varepsilon \leq 1$. The last two equations contradict (A3). Therefore, the initial assumption that $\lim_{a\to\infty}\frac{b}{a} \neq 1$ or that the limit does not exist is incorrect, as needed to be shown. □

**Proof of Lemma 10:**

We need to show that $C(\frac{r_1}{\sigma}) = \frac{1}{2}C(\bar{\alpha}\lambda)[1+o_{\bar{\alpha}\lambda}]$ for both Gaussian and Laplacian sources. We begin with the Laplacian case. Using simple algebraic steps one can easily obtain from (5) an expression for $r_1$. Specifically,

$$r_1 \;=\; -\frac{\sigma}{\sqrt{2}}\ln\big[\frac{e^{-\sqrt{2}\bar{\alpha}\lambda}(1+e^{-\sqrt{2}\lambda})}{2}\big] \;=\; \sigma\bar{\alpha}\lambda + \frac{\sigma\ln 2}{\sqrt{2}} - \frac{\sigma\ln(1+e^{-\sqrt{2}\lambda})}{\sqrt{2}} \;=\; \sigma\bar{\alpha}\lambda\big[1+\frac{\ln 2}{\sqrt{2}}\frac{1}{\bar{\alpha}\lambda}+\frac{o_\lambda}{\bar{\alpha}\lambda}\big]$$

Using this we now obtain

$$\begin{aligned} C(\frac{r_1}{\sigma}) &\overset{(a)}{=} \frac{r_1}{\sqrt{2}\sigma}L(\frac{r_1}{\sigma})[1+o_{\frac{r_1}{\sigma}}] \overset{(b)}{=} \frac{\bar{\alpha}\lambda}{\sqrt{2}}[1+o_{\bar{\alpha}\lambda}]L\big(\bar{\alpha}\lambda\big[1+\frac{\ln 2}{\sqrt{2}}\frac{1}{\bar{\alpha}\lambda}+\frac{o_\lambda}{\bar{\alpha}\lambda}\big]\big)[1+o_{\bar{\alpha}\lambda}] \\ &= \frac{\bar{\alpha}\lambda}{\sqrt{2}}\frac{1}{\sqrt{2}}e^{-\sqrt{2}\bar{\alpha}\lambda}e^{-\ln 2}e^{-\sqrt{2}o_\lambda}[1+o_{\bar{\alpha}\lambda}] \;=\; \frac{\bar{\alpha}\lambda}{\sqrt{2}}L(\bar{\alpha}\lambda)\frac{1}{2}[1+o_\lambda][1+o_{\bar{\alpha}\lambda}] \\ &= \frac{1}{2}C(\bar{\alpha}\lambda)[1+o_{\bar{\alpha}\lambda}]\ , \end{aligned}$$

where $(a)$ follows from Fact GL1, and $(b)$ derives from the fact that $\frac{r_1}{\sigma}\to\infty$ as $\bar{\alpha}\lambda\to\infty$. This shows the Laplacian case.

Next, we consider the Gaussian case. From (5) and Fact G6 one can easily obtain that

$$Q(\frac{r_1}{\sigma}) \;=\; \frac{Q(\bar{\alpha}\lambda)-Q((1+\bar{\alpha})\lambda)}{2} \;=\; \frac{Q(\bar{\alpha}\lambda)[1+o_\lambda]}{2}\ .$$

Using this we get

$$\begin{aligned} C(\frac{r_1}{\sigma}) &\overset{(a)}{=} G(\frac{r_1}{\sigma}) \overset{(b)}{=} \frac{r_1}{\sigma}Q(\frac{r_1}{\sigma})[1+o_{\frac{r_1}{\sigma}}] \overset{(c)}{=} \frac{r_1}{\sigma}\frac{Q(\bar{\alpha}\lambda)}{2}[1+o_\lambda][1+o_{\bar{\alpha}\lambda}] \\ &\overset{(d)}{=} \frac{r_1/\sigma}{\bar{\alpha}\lambda}\frac{1}{2}G(\bar{\alpha}\lambda)[1+o_{\bar{\alpha}\lambda}] \overset{(e)}{=} \frac{1}{2}C(\bar{\alpha}\lambda)[1+o_{\bar{\alpha}\lambda}]\ , \end{aligned}$$

where $(a)$ follows from Fact GL1, $(b)$ and $(d)$ follow from Fact G5, $(c)$ derives from the equation above and the fact that $\frac{r_1}{\sigma}\to\infty$ as $\bar{\alpha}\lambda\to\infty$, and $(e)$ uses Fact GL1 and Lemma A1, which implies that $\frac{r_1}{\sigma}=\bar{\alpha}\lambda[1+o_{\bar{\alpha}\lambda}]$ . This shows the Gaussian case, and concludes the proof of the lemma. □

**Proof of Lemma 11:**

We need to show that $\big(\frac{r_0}{\sigma}\big)^2 \;=\; \big(C(\alpha\lambda)+C(\bar{\alpha}\lambda)\big)\,o_{\alpha\lambda,\bar{\alpha}\lambda}$. We begin with the Laplacian case. Using simple algebraic steps one can easily obtain from (5) that

$$r_0 \;=\; \begin{cases} -\frac{\sigma}{\sqrt{2}}\ln\big[1+\frac{e^{-\sqrt{2}\bar{\alpha}\lambda}-e^{-\sqrt{2}\alpha\lambda}}{2}\big], & 0<\alpha\leq\frac{1}{2} \\ \frac{\sigma}{\sqrt{2}}\ln\big[1+\frac{e^{-\sqrt{2}\alpha\lambda}-e^{-\sqrt{2}\bar{\alpha}\lambda}}{2}\big], & \frac{1}{2}<\alpha\ <1 \end{cases}\ .$$

We recall the expansion $\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\ldots=x(1-\frac{x}{2}+\frac{x^2}{3}-\ldots)=x(1+o_{f(x),g(x)})$, where $f(x)$ and $g(x)$ are arbitrary functions such that when they both tend to infinity $x\to 0$. Using this expansion and the expression above for $r_0$, we obtain for $0 < \alpha \leq \frac{1}{2}$

$$\begin{aligned} r_0 \;&=\; -\frac{\sigma}{\sqrt{2}}\Big(\frac{e^{-\sqrt{2}\bar{\alpha}\lambda} - e^{-\sqrt{2}\alpha\lambda}}{2}\Big)[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}] \;=\; \frac{\sigma}{\sqrt{2}}\Big(\frac{\alpha\lambda L(\alpha\lambda)}{\sqrt{2}}\frac{1}{\alpha\lambda} - \frac{\bar{\alpha}\lambda L(\bar{\alpha}\lambda)}{\sqrt{2}}\frac{1}{\bar{\alpha}\lambda}\Big)[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}] \\ &\overset{(a)}{=}\; \frac{\sigma}{\sqrt{2}}C(\alpha\lambda)[1 + o_{\alpha\lambda}]\, o_{\alpha\lambda,\bar{\alpha}\lambda} - C(\bar{\alpha}\lambda)[1 + o_{\bar{\alpha}\lambda}]\, o_{\alpha\lambda,\bar{\alpha}\lambda} \;=\; \sigma\big(C(\alpha\lambda) + C(\bar{\alpha}\lambda)\big)\, o_{\alpha\lambda,\bar{\alpha}\lambda}\;, \end{aligned} \tag{A4}$$

where $(a)$ follows from Fact GL1. Similarly, it can be shown that the same expression for $r_0$ holds for $\frac{1}{2} < \alpha < 1$. It now follows that

$$\big(\frac{r_0}{\sigma}\big)^2 \;=\; \big(C(\alpha\lambda) + C(\bar{\alpha}\lambda)\big)^2 o_{\alpha\lambda,\bar{\alpha}\lambda} \;=\; \big(C(\alpha\lambda) + C(\bar{\alpha}\lambda)\big)\, o_{\alpha\lambda,\bar{\alpha}\lambda}\;,$$

where the second equality is due the fact that $C(\alpha\lambda) + C(\bar{\alpha}\lambda) \to 0$ as both $\alpha\lambda$ and $\bar{\alpha}\lambda$ tend to infinity. This shows the Laplacian case.

We now consider the Gaussian case. From (5) we obtain that

$$Q(\frac{r_0}{\sigma}) \;=\; \frac{Q(-\alpha\lambda) + Q(\bar{\alpha}\lambda)}{2} \;=\; \frac{1 - Q(\alpha\lambda) + Q(\bar{\alpha}\lambda)}{2}\;. \tag{A5}$$

Next, let $a \in \mathbb{R}$ be arbitrary. If $a \geq 0$, then

$$Q(a) \;=\; \int_a^\infty G(x)\,dx \;=\; \frac{1}{2} - \int_0^a G(x)\,dx \;\leq\; \frac{1}{2} - aG(a)\;,$$

from which it follows that $0 \leq a \leq \frac{\frac{1}{2} - Q(a)}{G(a)}$. Similarly, if $a < 0$, then

$$Q(a) \;=\; 1 - Q(|a|) \;=\; 1 - \Big(\frac{1}{2} - \int_0^{|a|} G(x)\,dx\Big) = \frac{1}{2} + \int_0^{|a|} G(x)\,dx \geq \frac{1}{2} + aG(a) = \frac{1}{2} - aG(a)\;,$$

from which it follows that $\frac{\frac{1}{2} - Q(a)}{G(a)} \leq a < 0$. The last two equations show that $a^2 \leq \Big(\frac{\frac{1}{2} - Q(a)}{G(a)}\Big)^2$. This is now used as follows:

$$\begin{aligned} \big(\frac{r_0}{\sigma}\big)^2 \;&\leq\; \Big(\frac{\frac{1}{2} - Q(\frac{r_0}{\sigma})}{G(\frac{r_0}{\sigma})}\Big)^2 \;\overset{(a)}{=}\; \Big(\frac{\frac{1}{2} - \frac{1}{2} + \frac{Q(\alpha\lambda)}{2} - \frac{Q(\bar{\alpha}\lambda)}{2}}{G(\frac{r_0}{\sigma})}\Big)^2 \;\overset{(b)}{=}\; \Big(\frac{Q(\alpha\lambda) - Q(\bar{\alpha}\lambda)}{\frac{4}{\sqrt{2\pi}}[1 + o_{\alpha\lambda,\bar{\alpha}\lambda}]}\Big)^2 \\ &\overset{(c)}{=}\; \big(Q(\alpha\lambda) - Q(\bar{\alpha}\lambda)\big)\, o_{\alpha\lambda,\bar{\alpha}\lambda} \;\overset{(d)}{=}\; \Big(\frac{G(\alpha\lambda)}{\alpha\lambda}[1 + o_{\alpha\lambda}] - \frac{G(\bar{\alpha}\lambda)}{\bar{\alpha}\lambda}[1 + o_{\bar{\alpha}\lambda}]\Big)\, o_{\alpha\lambda,\bar{\alpha}\lambda} \\ &\overset{(e)}{=}\; \big(C(\alpha\lambda)[1 + o_{\alpha\lambda}]\, o_{\alpha\lambda} - C(\bar{\alpha}\lambda)[1 + o_{\bar{\alpha}\lambda}]\, o_{\bar{\alpha}\lambda}\big)\, o_{\alpha\lambda,\bar{\alpha}\lambda} \;=\; \big(C(\alpha\lambda) + C(\bar{\alpha}\lambda)\big)\, o_{\alpha\lambda,\bar{\alpha}\lambda}\;, \end{aligned} \tag{A6}$$

where $(a)$ follows from (A5), $(b)$ is obtained from the fact that $\frac{r_0}{\sigma} \to 0$ as both $\alpha\lambda$ and $\bar{\alpha}\lambda$ tend to infinity, $(c)$ follows from having $Q(\alpha\lambda) - Q(\bar{\alpha}\lambda) \to 0$ as both $\alpha\lambda$ and $\bar{\alpha}\lambda$ tend to infinity, $(d)$ derives from Fact G5, and $(e)$ follows from Fact GL1. This shows the Gaussian case, and completes the proof of the lemma. □

## References

[1] G. J. Sullivan, "Efficient scalar quantization of exponential and laplacian random variables," *IEEE Trans. Info. Theory*, vol. 42, pp. 1365–1374, Sep. 1996.

[2] A. Gyorgy and T. Linder, "Optimal entropy-constrained scalar quantization of a uniform source," *IEEE Trans. Info. Theory*, vol. 46, pp. 2704–2711, Nov. 2000.

[3] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer, Boston, 1992.

[4] T. Berger, "Optimum quantizers and permutation codes," *IEEE Trans. Info. Theory*, vol. 18, pp. 759–765, Nov. 1972.

[5] A. N. Netravali and R. Saigal, "Optimal quantizer design using a fixed-point algorithm," *Bell Syst. Tech. J.*, vol. 55, pp. 1423–1435, Nov. 1976.

[6] P. Noll and R. Zelinski, "Bounds on quantizer performance in the low bit-rate region," *IEEE Trans. Comm.*, vol. 26, pp. 300–305, Feb. 1978.

[7] N. Farvardin and J. W. Modestino, "Optimum quantizer performance for a class of non-gaussian memoryless sources," *IEEE Trans. Info. Theory*, vol. 30, pp. 485–497, May 1984.

[8] J. C. Kieffer, T. M. Jahns, and V. A. Obuljen, "New results on optimal entropy-constrained quantization," *IEEE Trans. Info. Theory*, vol. 34, pp. 1250–1258, Sep. 1988.

[9] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, pp. 31–42, Jan. 1989.

[10] J. Buhmann and H. Kuhnel, "Vector quantization with complexity costs," *IEEE Trans. Info. Theory*, vol. 39, pp. 1133–1145, Jul. 1993.

[11] K. Rose and D. Miller, "A deterministic annealing algorithm for entropy-constrained vector quantizer design," *Asilomar Conf. on Signals, Systems and Computers*, vol. 2, pp. 1651–1655, Nov. 1993.

[12] H. Gish and J. N. Pierce, "Asymptotically efficient quantization," *IEEE Trans. Info. Theory*, vol. 14, pp. 676–683, Sep. 1968.

[13] T. Berger, *Rate Distortion Theory*, Prentice-Hall, Englewood Cliffs, 1971.

[14] K. Yao and H. H. Tan, “Absolute error rate-distortion functions for sources with constrained magnitudes,” *IEEE Trans. Info. Theory*, vol. 24, pp. 499–503, July 1978.

[15] A. N. Netravali and B. G. Haskell, *Digital Pictures: Representation, Compression and Standards*, Plenum, New York, second edition, 1988.

[16] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice-Hall, Englewood Cliffs, 1988.

[17] D. Marco and D. L. Neuhoff, “Low resolution scalar quantization for Gaussian sources and squared error,” to appear in *IEEE Trans. Info. Theory.*

[18] D. Marco and D. L. Neuhoff, “Low resolution scalar quantization for Gaussian sources and absolute error,” submitted to the *IEEE Trans. Info. Theory.*

[19] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, John-Wiley & Sons Inc., New York, 1967.

[20] A. Gyorgy and T. Linder, “On the Structure of optimal entropy-constrained scalar quantizers,” *IEEE Trans. Info. Theory*, vol. 48, pp. 416–427, Feb. 2002.

[21] D. Marco, “Asymptotic quantization and applications to sensor networks,” *Ph.D. Thesis, EECS Department, University of Michigan*, 2004.