Efficient Scalar Quantization of Exponential and Laplacian Random Variables

Gary J. Sullivan, *Member, IEEE*

***Abstract*— This paper presents solutions to the entropy-constrained scalar quantizer (ECSQ) design problem for two sources commonly encountered in image and speech compression applications: sources having the exponential and Laplacian probability density functions. We use the memoryless property of the exponential distribution to develop a new noniterative algorithm for obtaining the optimal quantizer design. We show how to obtain the optimal ECSQ either with or without an additional constraint on the number of levels in the quantizer. In contrast to prior methods, which require multidimensional iterative solution of a large number of nonlinear equations, the new method needs only a single sequence of solutions to one-dimensional nonlinear equations (in some Laplacian cases, one additional two-dimensional solution is needed). As a result, the new method is orders of magnitude faster than prior ones. We show that as the constraint on the number of levels in the quantizer is relaxed, the optimal ECSQ becomes a uniform threshold quantizer (UTQ) for exponential, but not for Laplacian sources. We then further examine the performance of the UTQ and optimal ECSQ, and also investigate some interesting alternatives to the UTQ, including a uniform-reconstruction quantizer (URQ) and a constant dead-zone ratio quantizer (CDZRQ).**

***Index Terms*— Quantization, exponential random variables, Laplacian random variables, Lagrange multiplier optimization, entropy constraint.**

I. Introduction

CONSIDERABLE attention has been focused on the design of optimal quantizers for sources encountered in image, speech, and other compression applications. However, noniterative methods for designing optimal quantizers for typical sources have been essentially nonexistent. Here we introduce a method of solving such a problem for two typical sources: exponential and Laplacian random variables. Specifically, we consider scalar quantization, in which each input random variable is separately mapped to its output approximation, as in conventional analog-to-digital conversion. However, we concern ourselves not just with optimizing the quantizer performance for a given number of possible output values, but also with a limit on the information rate of the quantizer as measured by its output entropy, resulting in an entropy-constrained scalar quantizer (ECSQ).

A fast quantizer design method for the Laplacian source was described by Nitadori [1], later independently derived by Lanfer [2], and discussed in [3], after other efforts using iterative multidimensional optimization [4]–[7]. Nitadori used the memoryless property of exponential decay to obtain optimal quantizers from only a single sequence of solutions to one-dimensional nonlinear equations. However, he did not consider entropy-constrained quantization, and only considered the mean-square error distortion measure for optimization. He also only considered quantizers having an even number of levels, which implies having a decision threshold located exactly at zero (so-called "mid-rise" quantizers). Lanfer's similar work also mentioned the ease with which the method could be applied to quantizers having an odd number of levels (so-called "mid-tread" quantizers). We generalize their approach to obtain solutions incorporating entropy constraints and general distortion measures, using the same memoryless property to derive a fast method of optimal ECSQ design for a general distortion measure. We also provide a new formulation of Nitadori's solution, describing the solution using the Lambert W function.

Berger [8] considered entropy-constrained quantizers, and noted that necessary conditions for ECSQ optimality were fulfilled by certain quantizers having an infinite number of levels and equal step widths, termed uniform-threshold quantizers (UTQ). Such quantizers are, in fact, optimal for sources having an exponential pdf, and the memoryless property is the key to Berger's observation as well. However, Berger noted that these quantizers were *not* optimal for the Laplacian source, as they fulfilled only necessary, but not sufficient conditions. We show that even Berger's best infinite-level quantizers can be improved upon by careful consideration of the center region of the Laplacian pdf. Berger also did not provide any straightforward solution for quantizers having only a finite number of levels, which we do provide here.

Other methods of designing optimized quantizers have typically involved iterative refinement techniques for multidimensional optimization [4]–[7], [9]–[11]. The only other publication describing optimal ECSQ design without requiring multidimensional optimization applied only to uniformly distributed random variables [10]. Thus this work describes the first reasonably fast optimal ECSQ designs for sources that are commonly encountered in practical quantization problems. A preliminary description of the work which we report here appeared in [12].

We begin by examining the general problem of entropy-constrained scalar quantization in Section II. We focus on ECSQ for the exponential random variable in Section III, and present an algorithm and theorem which determine the solution to the ECSQ design problem for any well-behaved difference-

Mansucript received May 5, 1995; revised February 6, 1996. The material in this paper was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'94), Adelaide, South Australia, April 1994.

The author is with PictureTel Corporation, Danvers, MA 01923 USA.

Publisher Item Identifier S 0018-9448(96)05597-6.

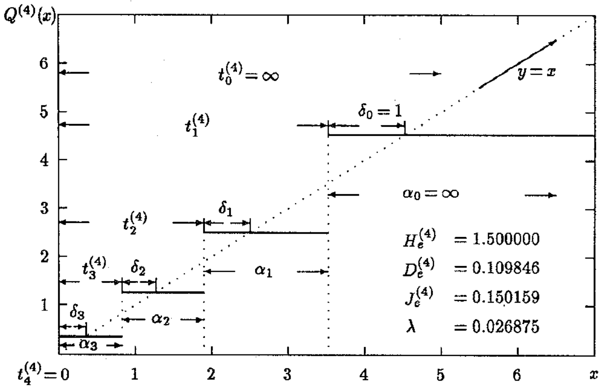

Fig. 1. A four-level MSE-optimal exponential ECSQ.

based distortion measure. In Section IV, we then show how the optimal ECSQ for a Laplacian-distributed random variable is found by building on our prior analysis of the exponentially distributed random variable. Also in Section IV, we examine the performance of the UTQ and optimal ECSQ, and also investigate interesting alternatives to the UTQ, including a uniform-reconstruction quantizer (URQ) and a constant dead-zone ratio quantizer (CDZRQ). Finally, the conclusions of our study are presented in Section V.

II. Scalar Quantization

Consider an input random variable X having a smooth pdf $f(x)$ which is greater than zero over $(0, \infty)$ and zero elsewhere. (We first consider only a one-sided pdf for convenience, without loss of generality to the similar problem posed for other regions of support.) Let an n-level quantizer $Q^{(n)}(\cdot)$ be defined in terms of a set of $n-1$ positive step sizes $\{\alpha_i\}_{i=1}^{n-1}$ (defining $\alpha_0 = \infty$) and a set of n nonnegative reconstruction offsets $\{\delta_i\}_{i=0}^{n-1}$, as shown in Fig. 1. The $n+1$ decision thresholds of the quantizer $\{t_i^{(n)}\}_{i=0}^{n}$ are given by

$$t_i^{(n)} = \sum_{j=i}^{n-1} \alpha_j, \qquad i = 0, \cdots, n \tag{1}$$

which we define as zero for $i = n$, and the n output values of the quantizer $\{y_i^{(n)}\}_{i=0}^{n-1}$ are given by

$$y_i^{(n)} = t_{i+1}^{(n)} + \delta_i. \tag{2}$$

The n-level scalar quantizer $Q^{(n)}(\cdot)$ is defined as a functional mapping of an input value $x > 0$ onto an output representation $Q^{(n)}(x)$, such that

$$Q^{(n)}(x) = y_i^{(n)}, \quad \text{for } t_{i+1}^{(n)} < x \leq t_i^{(n)}. \tag{3}$$

Note that the i index subscript *decreases* to the right of the zero-input location, and that the quantizer is defined in terms of its step sizes $\{\alpha_i\}_{i=1}^{n-1}$ and reconstruction offsets $\{\delta_i\}_{i=0}^{n-1}$ rather than the more common convention using the thresholds $\{t_i^{(n)}\}_{i=0}^{n}$ and output values $\{y_i^{(n)}\}_{i=0}^{n-1}$. If we consider some large positive number N, this modified definition allows a single set of $N-1$ step sizes and N reconstruction offsets to be sufficient to define an important *family* of N different quantizers $\{Q^{(n)}(\cdot)\}_{n=1}^{N}$ with each quantizer in the family defined by (1)–(3).

We assume a difference-based distortion measure $d(\Delta)$, which increases monotonically and smoothly (although not necessarily symmetrically) as its argument deviates from zero, and has a finite slope for a finite argument. The expected quantizer distortion is

$$\begin{aligned} D_f^{(n)} &= E_X\{d(X - Q^{(n)}(X))\} & (4)\\ &= \sum_{i=0}^{n-1} \int_{t_{i+1}^{(n)}}^{t_i^{(n)}} d(x - y_i^{(n)}) f(x)\, dx & (5) \end{aligned}$$

and the probability of each output value $y_i^{(n)}$ is

$$p_i^{(n)} = \int_{t_{i+1}^{(n)}}^{t_i^{(n)}} f(x)\, dx. \tag{6}$$

The output probabilities determine the output entropy of the quantizer, a lower bound on the expected bit rate required to encode the output values with arbitrarily small error

$$H_f^{(n)} = -\sum_{i=0}^{n-1} p_i^{(n)} \log_2 p_i^{(n)} \text{ bits per sample.} \tag{7}$$

Modern coding methods such as arithmetic coding can achieve rates close to this lower bound [9], [13].

Using a Lagrange multiplier $\lambda \geq 0$ to specify the incremental relative importance of distortion and entropy, the optimal n-level ECSQ is defined by the step sizes and reconstruction offsets (or alternately, the thresholds and output values) which minimize the *objective function*

$$J_f^{(n)} = D_f^{(n)} + \lambda H_f^{(n)}. \tag{8}$$

Such a quantizer is optimal in the sense that no other scalar quantizer with n levels can obtain lower distortion with equal or lower output entropy. A solution of this problem can be used to solve a constrained-entropy or constrained-distortion problem unless there is no value of $\lambda \geq 0$ for which minimization of (8) results in equality to the constraint value[14], [10], [11] (and the constraint is not achieved by $\lambda = 0$ for an entropy constraint or $\lambda = \infty$ for a distortion constraint). If the minimization of (8) requires some of the n step sizes to be zero or infinite in value, we say that no optimal ECSQ exists with n levels, since zero or infinite step sizes imply equivalence to a quantizer with fewer levels (at least unless an infinite number of steps have widths approaching zero).

One necessary condition for optimality is that the optimal value of each reconstruction offset δ_i^* must minimize the distortion contribution of its associated step interval [6]

$$\begin{aligned} \delta_i^* &= \min_{\psi}{}^{-1} \left\{ \int_{t_{i+1}^{(n)}}^{t_i^{(n)}} d(x - t_{i+1}^{(n)} - \psi) f(x)\, dx \right\} & (9)\\ &= \min_{\psi}{}^{-1} \left\{ \int_0^{\alpha_i} d(x - \psi) f(x + t_{i+1}^{(n)})\, dx \right\} & (10) \end{aligned}$$

where the "*" superscript indicates optimality. Any value of ψ which minimizes the distortion contribution is equally optimal for use as δ_i^* if the minimizing value is not unique. Note that since reconstruction offsets affect only the quantizer distortion and not the bit rate, they can be optimized for distortion alone. For example, the most commonly used distortion measure is mean-square error (MSE), denoted here by $d_{\text{mse}}(\Delta)$ and given by $d_{\text{mse}}(\Delta) = |\Delta|^2$. In the MSE case, (10) becomes

$$\delta_i^* = \int_0^{\alpha_i} x f(x + t_{i+1}^{(n)})\, dx \Bigg/ \int_0^{\alpha_i} f(x + t_{i+1}^{(n)})\, dx. \quad (11)$$

A second necessary condition for optimality is that the optimal value of each step width α_{i+1}^* must locate the decision threshold $t_i^{(n)}$ where the two nearest reconstructions have equal rate-distortion penalty [11]

$$d(\alpha_{i+1}^* - \delta_{i+1}) - \lambda \log_2 (p_{i+1}^{(n)}) = d(-\delta_i) - \lambda \log_2 (p_i^{(n)}). \quad (12)$$

A more analytical way of justifying this expression is by setting to zero the partial derivative of the objective function $J_f^{(n)}$ with respect to the threshold $t_i^{(n)}$, while holding the other thresholds and the output values constant.

Typically, prior methods of ECSQ design have consisted of multidimensional optimization algorithms which iteratively refine a set of estimates for the quantizer parameters. Usually, this has consisted of repeatedly alternating between values obtained using the above two necessary conditions for optimality, using some equivalent to (9) to determine appropriate output values, and some equivalent to (12) to determine the threshold locations [4]–[7], [9]–[11]. However, such prior methods have operated by optimizing the output values $\{y_i^{(n)}\}_{i=0}^{n-1}$ and thresholds $\{t_i^{(n)}\}_{i=0}^{n}$ directly, rather than indirectly by use of step widths and reconstruction offsets. Note that such a prior method, when changing a particular threshold value, affects the probabilities and distortion contributions of just two adjacent steps of the quantizer. However, if we use the indirect definition and change the width α_i of one step, the probabilities and distortion contributions of all steps to the right (steps $j = 0, 1, \cdots, i$) are affected.

Given the number of levels n, the optimal ECSQ for $\lambda = 0$ has the lowest possible distortion. In this case, if the distortion measure is symmetric, (12) becomes merely

$$\alpha_{i+1}^* = \delta_{i+1} + \delta_i. \quad (13)$$

This condition was assumed by Nitadori [1] for his method of optimal $\lambda = 0$ design (i.e., when the only constraint is n) for a Laplacian-distributed source using the squared-error distortion measure $d_{\text{mse}}(\Delta)$. In contrast, an optimal ECSQ as defined above is optimal for dual constraints: the given number of levels in the quantizer (n) and the output entropy resulting from the specified Lagrange multiplier λ.

III. Exponential Random Variables

The key to analysis of the exponentially distributed random variable lies in its *memoryless* property, which can be expressed as follows: given that an exponentially distributed random variable X has a value exceeding some fixed nonnegative threshold t, the conditional pdf of $X - t$ is the same as the pdf of the original random variable X

$$f_e(x) = \mu^{-1} e^{-x/\mu}, \quad x > 0, \mu > 0. \quad (14)$$

Without loss of generality, assume $\mu = 1$.

As a result of the memoryless property, we can build an optimal quantizer one step at a time. First, define a function $\gamma(\alpha, \delta)$, which measures the distortion contribution over the left-most step of width α and reconstruction offset δ

$$\gamma(\alpha, \delta) = \int_0^{\alpha} d(x - \delta) e^{-x}\, dx. \quad (15)$$

The optimal reconstruction offset δ_i^* in this case depends only on the step width α_i. Using (14), the memoryless property of the exponential pdf allows the substitution

$$f_e(x + t_{i+1}^{(n)}) = e^{-t_{i+1}^{(n)}} f_e(x)$$

in (10), which allows us to remove the term $e^{-t_{i+1}^{(n)}}$ from the minimization. Thus we denote the optimal value δ_i^* as $\delta(\alpha_i)$, by defining the function

$$\delta(\alpha) = \min_{\psi}{}^{-1} \{\gamma(\alpha, \psi)\}. \quad (16)$$

Now, only the step widths are needed to specify the optimal quantizer. Define the minimum value of $\gamma(\alpha, \delta)$ as

$$\gamma(\alpha) = \gamma(\alpha, \delta(\alpha)). \quad (17)$$

Denote the distortion, output entropy, and objective function of the quantizer as $D_e^{(n)}$, $H_e^{(n)}$, and $J_e^{(n)}$, where the subscript indicates the exponential pdf of the source. Define also the Bernoulli entropy $B(p)$ as the per-letter entropy of a Bernoulli process (a binary discrete memoryless source) with success probability p, given by

$$B(p) = -p \log_2 (p) - (1 - p) \log_2 (1 - p), \quad \text{for } 0 < p < 1$$

(otherwise 0). We can then use (14) to show the key recursion relation

$$J_e^{(n+1)} = \gamma(\alpha_n, \delta_n) + \lambda B(e^{-\alpha_n}) + e^{-\alpha_n} J_e^{(n)}. \quad (18)$$

Note that only the final term of (18) depends on $\{\alpha_i\}_{i=1}^{n-1}$, so the optimal $(n+1)$-level ECSQ can be found by first determining the optimal n-level ECSQ, and then solving for α_n^*. The design process can be initialized with $n = 1, \alpha_0 = \infty, \delta_0 = \delta(\infty)$, and $J_e^{(1)} = \gamma(\infty)$, and can then be recursed to create any number of levels. The recursive relationship is a direct consequence of the memoryless property: given that the input random variable has a value greater than the left-most nonzero threshold $t_n^{(n+1)}$ which is equal to α_n (an event which occurs with Bernoulli success probability $e^{-\alpha_n}$), the remainder of the quantizer defines an n-level quantizer for an exponential random variable having *the same* pdf as the original input. In order for the entire quantizer to be optimal, the i-level quantizer defined by the size of the steps to the right of each of its decision thresholds $t_i^{(n+1)}$ must also be optimal for *the same* Lagrange multiplier λ.

We now show an equivalent alternative to the method discussed above, which does not necessarily require computing

$J_e^{(n)}$ in order to determine α_n^*. Instead, we can start with $\alpha_0 = \infty$ and proceed sequentially by obtaining α_{i+1}^* using only the value of α_i^*, for $i = 0, \cdots, n-1$, to completely determine the optimal quantizer. Using (12) and noting that

$$p_{i+1}^{(n)} = e^{-t_{i+2}^{(n)}}(1 - e^{-\alpha_{i+1}}) \tag{19}$$

$$p_i^{(n)} = e^{-(t_{i+2}^{(n)}+\alpha_{i+1})}(1 - e^{-\alpha_i}) \tag{20}$$

we obtain

$$d(\alpha_{i+1}^* - \delta(\alpha_{i+1}^*)) - \lambda\left[\frac{\alpha_{i+1}^*}{\ln(2)} + \log_2(1 - e^{-\alpha_{i+1}^*})\right] = d(-\delta(\alpha_i^*)) - \lambda \log_2(1 - e^{-\alpha_i^*}). \tag{21}$$

The right-hand side depends only on α_i^*, and the left-hand side only on α_{i+1}^*. Thus knowing α_i^* may be sufficient to determine α_{i+1}^* (and we can start the process off with $\alpha_0 = \infty$). Equation (21) may, however, be fulfilled for multiple values of α_{i+1}, in which case the best among them must be chosen as α_{i+1}^* according to the resulting $J_e^{(i+2)}$.

For example, (21) has two solutions for α_{i+1} when distortion is measured by mean-square error and $\lambda > 0$. The smaller solution corresponds to finding a $t_{i+1}^{(i+2)}$ which maximizes $J_e^{(i+2)}$ and the larger solution corresponds to minimizing it (for $\lambda = 0$, there is only one solution). The smaller solution can be avoided by first obtaining a low rough estimate of the solution for $\lambda = 0$, and then considering only solutions larger than that. Since the left-hand side of (21) is easily shown to be convex with respect to α_{i+1} when using MSE distortion, no more than two solutions exist and there can be only one with a positive first derivative (which is the correct solution α_{i+1}^*).

One quantizer of particular interest is the n-level uniform threshold quantizer (UTQ), which we define as a quantizer having all its finite step widths equal in value, i.e., $\alpha_i = \tilde{\alpha}$ for $i = 1, \cdots, n-1$ (note that $\alpha_0 = \infty$). Since the best performance for such a quantizer is obtained by having $\delta_i = \delta(\alpha_i)$ for each i, we assume that all but one of the reconstruction offsets are equal to some constant value as well, i.e., $\delta_i = \tilde{\delta}$ for $i = 1, \cdots, n-1$. The performance of a UTQ with this type of offsets can thus be computed in terms of three parameters, a step size $\tilde{\alpha}$, a reconstruction offset $\tilde{\delta}$, and the rightmost reconstruction offset δ_0, yielding

$$J_e^{(n)}(\tilde{\alpha}, \tilde{\delta}, \delta_0) = \left(\frac{1 - e^{-(n-1)\tilde{\alpha}}}{1 - e^{-\tilde{\alpha}}}\right)[\gamma(\tilde{\alpha}, \tilde{\delta}) + \lambda B(e^{-\tilde{\alpha}})] + e^{-(n-1)\tilde{\alpha}}\gamma(\infty, \delta_0). \tag{22}$$

A. Optimal Finite-Alphabet ECSQ Design Algorithm for an Exponential Source

If an n-level optimal ECSQ exists for an exponentially distributed source using Lagrange multiplier λ, then it can be found as follows:

1) Set $i = 0$ and $\delta_i = \delta(\infty)$. Note $\alpha_i = \infty$ and $J_e^{(i+1)} = \gamma(\infty)$.
2) If $i = n-1$ then **Stop**—the quantizer design is complete. Otherwise, set α_{i+1} to α_{i+1}^*, the solution of (21) which minimizes $J_e^{(i+2)}$ in (18) with $\delta_{i+1} = \delta(\alpha_{i+1})$.
3) Increment i and return to Step 2.

B. An "All or Nothing" Theorem of Optimal ECSQ for an Exponential Source

We next present a theorem which determines whether an optimal n-level ECSQ exists and also defines the optimal scalar quantizer with no restriction on the number of levels. First, we define the function

$$\lambda_e(\alpha) = [(1 - e^{-\alpha})\gamma(\infty) - \gamma(\alpha)]/B(e^{-\alpha}) \tag{23}$$

and the parameter

$$\lambda_e^{(\max)} = \max_{\alpha}\{\lambda_e(\alpha)\}. \tag{24}$$

For $\lambda > \lambda_e^{(\max)}$, no optimal entropy-constrained quantizer exists for any number of levels $n > 1$. For $0 \leq \lambda < \lambda_e^{(\max)}$, an optimal quantizer exists for all possible numbers of levels $n \geq 1$ and has an objective function $J_e^{(n)}$ that is strictly decreasing with n. If the restriction on the number of levels in the quantizer is removed, the optimal quantizer becomes an infinite-level UTQ with step width $\tilde{\alpha}$, where $\tilde{\alpha}$ satisfies

$$[d(\tilde{\alpha} - \delta(\tilde{\alpha})) - d(-\delta(\tilde{\alpha}))]/\tilde{\alpha} = \lambda/\ln(2). \tag{25}$$

This UTQ has entropy and distortion given by

$$H_e^{(\infty)}(\tilde{\alpha}) = B(e^{-\tilde{\alpha}})/(1 - e^{-\tilde{\alpha}}) \tag{26}$$

$$D_e^{(\infty)}(\tilde{\alpha}) = \gamma(\tilde{\alpha})/(1 - e^{-\tilde{\alpha}}). \tag{27}$$

For $\lambda = \lambda_e^{(\max)}$, an optimal quantizer exists for any number of levels n if and only if $\lambda_e(\alpha) = \lambda_e^{(\max)}$ for some finite α, in which case the objective function of the optimal quantizer is $J_e^{(n)} = \gamma(\infty)$, regardless of n. This optimal n-level ECSQ is a UTQ with step width $\tilde{\alpha}$, where $\tilde{\alpha}$ satisfies $\lambda_e(\tilde{\alpha}) = \lambda_e^{(\max)}$ and also satisfies (25) and

$$\Upsilon(\tilde{\alpha}, \delta(\infty)) = 1 \tag{28}$$

where the function $\Upsilon(\alpha, \beta)$ is defined by

$$\Upsilon(\alpha, \beta) = (e^{\alpha} - 1) \cdot \exp\left\{-\alpha\left[\frac{d(\alpha - \delta(\alpha)) - d(-\beta)}{d(\alpha - \delta(\alpha)) - d(-\delta(\alpha))}\right]\right\}. \tag{29}$$

As proof, we provide the following: When considering adding a level to an n-level quantizer, an objective function improvement can always be obtained for $\lambda < \lambda_e^{(\max)}$ by splitting the infinite-length rightmost step into two steps. Although this may not give the best performance (as will the algorithm described above), it establishes the existence of an optimal $(n+1)$-level quantizer, since it shows that *any* n-level quantizer can be improved by adding another level. As more and more steps are added to the quantizer, the step sizes decrease monotonically toward a steady-state step size $\tilde{\alpha}$ which satisfies the optimality requirements when adjacent steps have this same width, and the quantizer performance approaches that of an infinite-level UTQ. In fact, the performance must approach that of the infinite-level UTQ at least exponentially fast with n, as a consequence of its objective function being bounded from above by (22), a bound which approaches infinite-level UTQ performance exponentially fast with n. For $\lambda = \lambda_e^{(\max)}$, the *ad hoc* rightmost step split and the optimal procedure both result in quantizers with the same performance, as the optimal finite step size will satisfy optimality requirements whether the step size to its right is either equally wide or is infinitely wide. For $\lambda > \lambda_{\rm e}^{(\max)}$, even the first split to make a two-level quantizer is not worth the entropy penalty, and an improvement of any given quantizer can always be obtained by replacing its rightmost two steps with a single infinitely wide step, thus no optimal quantizer exists for $n > 1$. (Note that if $\lambda_{\rm e}^{(\max)}$ can be obtained by a finite step size, the operational rate-distortion curve of optimal ECSQ must be nonconvex.)

The Appendix shows the following: that $\lambda_{\rm e}^{(\max)}$ is infinite if the distortion measure $d(\Delta)$ increases more than linearly for large Δ (e.g., if $d(\Delta)$ is strictly convex); that $\lambda_{\rm e}^{(\max)}$ is finite but is reached only by infinite α for a linearly increasing distortion measure; and that $\lambda_{\rm e}^{(\max)}$ is finite and can be reached with a finite α only if the distortion measure increases less than linearly as its argument deviates from zero [15], [16] (the only case for which a constrained-rate or constrained-distortion ECSQ optimization problem may not be solvable by the objective function minimization method, since this case implies a nonconvex optimal performance curve).

C. Exponential-pdf Quantizers for MSE

For example, when using the squared-error distortion measure $d_{\rm mse}(\Delta) = |\Delta|^2$, we obtain $\lambda_{\rm e}^{(\max)} = \infty$, and

$$\gamma_{\rm mse}(\alpha, \delta) = (\delta^2 - 2\delta + 2)(1 - e^{-\alpha}) - \alpha e^{-\alpha}(\alpha - 2\delta + 2) \tag{30}$$

$$\delta_{\rm mse}(\alpha) = 1 - \alpha e^{-\alpha}/(1 - e^{-\alpha}) \tag{31}$$

$$D_{\rm e}^{(\infty)}(\alpha) = 1 - \alpha^2 e^{-\alpha}/(1 - e^{-\alpha})^2. \tag{32}$$

For $\lambda = 0$ (the "Nitadori," or "Lloyd–Max" quantizer design case), each optimal step size α_{i+1}^* can be obtained from (21) as a function of the prior step size α_i^* by solving

$$(\alpha_{i+1}^* - \nu_i)e^{(\alpha_{i+1}^* - \nu_i)} = -\nu_i e^{-\nu_i} \tag{33}$$

where $\nu_i = 1 + \delta_{\rm mse}(\alpha_i^*)$. The solution can be expressed using the Lambert W function, yielding

$$\alpha_{i+1}^* = \nu_i + W(-\nu_i e^{-\nu_i}) \tag{34}$$

where $W(\cdot)$ denotes the real-valued solution to the transcendental equation $we^w = z$ for w as a function of z such that $W(z) \geq -1$ in the region of interest, $-1/e \leq z \leq -2/e^2$. The Lambert W function can be evaluated efficiently by methods described in [17], and can be roughly approximated in the region of interest by the first terms of the series

$$\begin{aligned} W(z) = {} & -1 + q - \tfrac{1}{3} q^2 + \frac{11}{72} q^3 - \frac{43}{540} q^4 + \frac{769}{17280} q^5 \\ & - \frac{221}{8505} q^6 + \frac{680863}{43545600} q^7 - \frac{1963}{204120} q^8 \\ & + \frac{226287557}{37623398400} q^9 - \frac{5776369}{1515591000} q^{10} + \cdots \end{aligned} \tag{35}$$

where $q = \sqrt{2(ez+1)}$ [17]. (This series can be obtained by reversion of the power series formed by defining $r = W(z) + 1$ and using the expression $(r-1)e^r = ez$ with $e^r = \Sigma_{j=0}^{\infty} r^j/j!$.) The number of terms given above is sufficient to approximate the W function value to within an error of less than 10^{-4} in the region of interest. Other interesting distortion measures include absolute error $d(\Delta) = |\Delta|$, for which $\lambda_{\rm e}^{(\max)} = \ln(2)$ is finite but is reached only by infinite α, and $d(\Delta) = 1 - e^{-|\Delta|}$, which has a finite $\lambda_{\rm e}^{(\max)}$ at a finite α.

IV. Laplacian Random Variables

We can build upon the results of Section III to derive optimal quantizers for random variables having a Laplacian pdf

$$f_L(x) = e^{-|x|\sqrt{2/\sigma^2}}/\sqrt{2\sigma^2}. \tag{36}$$

Without loss of generality, assume $\sigma^2 = 2$.

Consider first the step of the quantizer which has an output value ε associated with the $x = 0$ input value. The boundaries of the step are defined by two thresholds $t_l \geq 0$ and $t_r \geq 0$ such that $t_l + t_r > 0$, where any input x such that $-t_l \leq x \leq t_r$ is associated with output ε. For the trivial case of $n = 1$, both t_l and t_r are infinite. For $n > 1$, at least one must be finite. Denote the number of levels to the left and right of zero as n_l and n_r, such that $n = n_l + n_r + 1$.

Again, we can examine the behavior of the quantizer in a stepwise fashion. Define the distortion contribution

$$\eta(t_l, t_r, \varepsilon) = \int_{-t_l}^{t_r} d(x - \varepsilon)e^{-|x|}/2\, dx \tag{37}$$

and the ternary entropy $T(p, q) = B(p) + (1-p)B(q/(1-p))$. The optimal value ε^* for ε is a function only of t_l and t_r

$$\varepsilon(t_l, t_r) = \min_{\psi}{}^{-1}\{\eta(t_l, t_r, \psi)\}. \tag{38}$$

We decompose the quantizer for this source into three smaller "subquantizers." Given that a Laplacian-distributed input random variable X is greater than t_r, the conditional pdf of $X - t_r$ is exponential. Similarly, given that X is less than $-t_l$, the conditional pdf of $-X - t_l$ is exponential.

Thus denoting the objective function of the overall quantizer as $J_L^{(n)}$, where the subscript indicates the Laplacian pdf, in a manner similar to (18) we can obtain

$$\begin{aligned} J_L^{(n)} = {} & \eta(t_l, t_r, \varepsilon) + \lambda T(\tfrac{1}{2}\, e^{-t_l}, \tfrac{1}{2}\, e^{-t_r}) \\ & + \tfrac{1}{2}\, (e^{-t_l} \hat{J}_{\rm e}^{(n_l)} + e^{-t_r} J_{\rm e}^{(n_r)}) \end{aligned} \tag{39}$$

where $\hat{J}_{\rm e}^{(n_l)}$ denotes the objective function of an n_l-level quantizer for an exponential pdf with distortion measure $\hat{d}(\Delta) = d(-\Delta)$, and $J_{\rm e}^{(n_r)}$ is defined as before. If t_r is infinite, then n_r is zero and we arbitrarily define $J_{\rm e}^{(n_r)}$ as $\gamma(\infty)$. Similarly, if t_l is infinite, then n_l is zero and $\hat{J}_{\rm e}^{(n_l)}$ is $\hat{\gamma}(\infty)$. Again we have reduced the optimization problem to a stepwise procedure—first we design the left and right subquantizers (per Section III), and then simply determine the zero-input step thresholds t_l and t_r which minimize $J_L^{(n)}$.

Again we may consider an alternative to direct minimization of $J_L^{(n)}$ which may allow us to determine the best values for t_l and t_r using only the values of the innermost left and right step sizes $\hat{\alpha}_{n_l-1}$ and α_{n_r-1}. For a Laplacian source, (12) becomes two conditions

$$d(t_r^* - \varepsilon(t_l^*, t_r^*)) - \lambda\left[1 + \frac{t_r^*}{\ln(2)} + \log_2\left(1 - \frac{e^{-t_l^*} + e^{-t_r^*}}{2}\right)\right] = d(-\delta(\alpha_{n_r-1})) - \lambda \log_2(1 - e^{-\alpha_{n_r-1}}) \qquad (40)$$

and an analogous mirror-image relation for the left threshold t_l.

If the distortion measure is symmetric, the optimal quantizer should also be symmetric (or nearly so). Thus if n is odd, $n_l = n_r, t_l^* = t_r^*$, and $\varepsilon^* = 0$, and the two relations (40) are the same (if $\lambda = 0, t_l^* = t_r^* = \delta(\alpha_{n_l-1}^*)$). If n is even, two equivalent mirror-image solutions exist, one having $n_l = n_r + 1, 0 \leq t_l^* < t_r^*$, and $\varepsilon^* > 0$ (if $\lambda = 0, t_l^* = 0, t_r^* = \alpha_{n_l-1}^*$, and $\varepsilon^* = \delta(a_{n_l-1}^*)$).

As the Lagrange multiplier is increased slowly from zero, there may be up to four critical values of λ (two for symmetric distortion measures) at which sudden changes occur in the optimal rate and distortion, as the left and right subquantizers each collapse to a single level when their $\lambda_e^{(\max)}$ is reached, and then vanish altogether as t_l and t_r become infinite.

A. Bounding the Center Step Size

We now define an important test ratio θ which can provide a bound on the size of the center step for the Laplacian pdf. Note that if we were to construct a $(i+2)$-level quantizer for the exponential pdf, then the distortion at the edge of the optimized step with output value y_{i+1} would be $d(\alpha_{i+1}^* - \delta(\alpha_{i+1}^*))$, the probability of the output y_{i+1} would be $p_{i+1}^{(i+2)} = (1 - e^{-\alpha_{i+1}^*})$, the probability of the output y_i would be $p_i^{(i+2)} = e^{-\alpha_{i+1}^*}(1 - e^{-\alpha_i^*})$, and (12) would be fulfilled. Consider now a situation in which we somehow maintain the optimality of the reconstruction values and keep the distortion values in (12) the same but change the input pdf over the two steps, changing the step probabilities to p'_{i+1} and p'_i rather than $p_{i+1}^{(i+2)}$ and $p_i^{(i+2)}$, and define the probability ratio

$$\theta = \frac{p'_{i+1}/p'_i}{p_{i+1}^{(i+2)}/p_i^{(i+2)}}. \qquad (41)$$

Then, after examining (12) for $\lambda > 0$, we see that the first step of the quantizer α_{i+1} should be widened if $\theta > 1$, kept the same size if $\theta = 1$, and narrowed if $\theta < 1$.

Now consider a quantizer for the Laplacian pdf with a symmetric distortion measure, wherein $n_r = n_l$ (when n is odd). Let us examine the rate-distortion balance for a threshold value $t_r = \alpha_{n_r}^* - \delta(\alpha_{n_r}^*)$. This yields

$$p'_{n_r} = (1 - e^{-(\alpha_{n_r}^* - \delta(\alpha_{n_r}^*))}) \qquad (42)$$

$$p'_{n_r-1} = e^{-(\alpha_{n_r}^* - \delta(\alpha_{n_r}^*))}(1 - e^{-\alpha_{n_r-1}^*})/2 \qquad (43)$$

so that

$$\theta = 2\left(\frac{e^{(\alpha_{n_r}^* - \delta(\alpha_{n_r}^*))} - 1}{e^{\alpha_{n_r-1}^*} - 1}\right). \qquad (44)$$

B. Optimal Infinite-Level Laplacian-pdf Quantizers

As the restriction on the number of levels is relaxed, the optimal left and right subquantizers may become UTQ's as dictated by Section III (although their step sizes may be different if the distortion measure is not symmetric), but the zero-input step width $t_l^* + t_r^*$ generally will not approach the same value as the width of the other steps. Thus the optimal quantizer *is not* a UTQ. It differs by having a different-sized "dead-zone" around zero.

Consider such a quantizer for a symmetric distortion measure. For an infinite-level quantizer, we can then define $t^* = t_l^* = t_r^*$ and observe that $\varepsilon^* = 0$. Then we can dispense with the use of λ altogether, by substituting (25) into (40). This creates a single equation that relates the step size α of the right and left subquantizers to t^* (half the dead-zone width)

$$\Upsilon(\alpha, t^*) = 2(e^{t^*} - 1) \qquad (45)$$

where the function $\Upsilon(\alpha, \beta)$ is defined by (29).

C. Laplacian-pdf Quantizers for MSE

If, for example, we use the squared-error distortion measure $d_{\text{mse}}(\Delta) = |\Delta|^2$, we can insert (31) into (44) to show that for $\lambda > 0$, the ratio $\theta < 1$ always. This proves that the optimal dead-zone (center step) width is *always less than* $2(\alpha_{n_r}^* - \delta(\alpha_{n_r}^*))$ under these conditions. (The opposite conclusion was mistakenly reported in [12].) Thus *the reconstruction values of the optimal quantizer are actually closer together near the middle of the quantizer* than they are farther away from the center.

We now turn our attention to quantizers with a large (effectively infinite) number of levels. We find, using (45) that *the dead zone is always larger than the size of the other steps*, but we have also shown that the dead-zone size must be less than $2(\alpha_{n_r}^* - \delta(\alpha_{n_r}^*))$. Define the dead-zone ratio z such that

$$z = t/(\alpha - \delta(\alpha)) \qquad (46)$$

where α is the step size for the two infinite-level UTQ subquantizers to the right and left of the center dead-zone step. We have found that the optimal dead-zone ratio z^* always lies in the range of $0.95 < z^* < 1$. Its limiting value is unity at very high and very low bit rates, and its behavior is shown as a function of step size in Fig. 2, and as a function of bit rate in Fig. 3.

Berger noted that an infinite-level UTQ with a decision threshold exactly at zero and optimal offsets (a "mid-rise" UTQ with step width α and offset δ, specified by $t_l = 0, t_r = \alpha, \varepsilon = \delta = \delta(\alpha)$) fulfilled the necessary conditions for optimality given above [8]. However, he noted that such quantizers could not achieve an entropy below one bit per sample, and also that they did not perform as well as did a "mid-tread" UTQ with optimal offsets (specified by $t_l = t_r = \alpha/2, \varepsilon = 0, \delta = \delta(\alpha)$). He thus proposed the mid-tread UTQ with optimal offsets as a nearly optimal scalar quantizer. This mid-tread UTQ, which we shall call the *uniform threshold with optimal reconstruction quantizer* (UTORQ), has a dead-zone ratio which is a function of the step size α, yielding $z = \alpha/[2(\alpha - \delta(\alpha))]$. A mid-tread UTQ also has a very simple

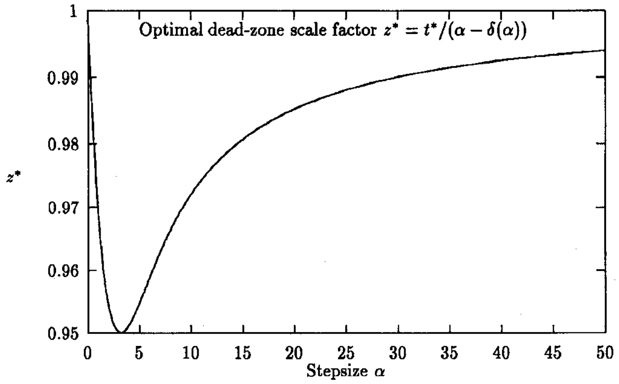

Fig. 2. Optimal dead-zone scale factor z^* as a function of step size α.

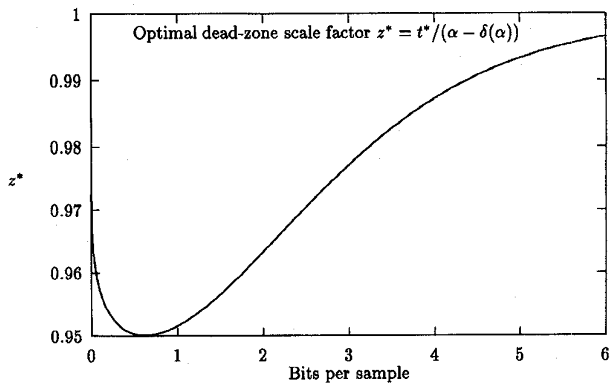

Fig. 3. Optimal dead-zone scale factor z^* as a function of bit rate.

encoding rule, consisting of just scaling and rounding to the nearest integer.

Actually, the infinite-level optimal ECSQ also has a fairly simple encoding and decoding rule, since all steps of the quantizer except the center step have the same step size and the same reconstruction offset. However, some may consider the design and operation of the truly optimal ECSQ to be more complex than practical considerations may allow, and thus would desire a quantizer with a simpler design method and a simpler encoding or decoding rule. For this reason, we now investigate what we call a *uniform reconstruction quantizer* (URQ). The URQ has a very simple decoding rule consisting of just multiplying the quantization index by the quantizer step size. Thus all reconstruction values of the URQ are equally spaced.

Although we have specified the output values of the URQ, we have not yet described an encoding rule for it. The optimal encoding rule would be difficult to design and operate, as the constraints we have introduced by specifying the decoding rule would mean that the best thresholds near the center of the quantizer should not be evenly spaced. Instead, we propose a fairly simple suboptimal thresholding rule which preserves the optimality of the reconstruction values, and describe the type of quantizers we investigate here as the *uniform reconstruction with unity ratio quantizer* (URQ). The encoding rule is simply that the right and left center thresholds are $t = t_l = t_r = \alpha - \delta(\alpha)$, and the rest of the thresholds construct the right and left subquantizers as UTQ's with step size α. Therefore, the URURQ has a constant *unity dead-zone ratio* $z_c = 1$. We know that these thresholds are not optimal, as we have shown that the optimal value of $t = t_l = t_r$ should be strictly less than $\alpha - \delta(\alpha)$. However, the URURQ is simple to design and operate, provides a lower bound on the performance of the best URQ, and has a decoding rule that is compatible with the optimal URQ encoding rule. We will show that the URURQ has performance close to that of the optimal ECSQ, and is significantly superior to that of the UTORQ. In light of the excellent performance of the URURQ, not much can be gained by attempting to find a better URQ encoding rule.

Another important quantizer is the *uniform quantizer* (UQ), which can be viewed either as a UTQ with suboptimal offsets or as a URQ with suboptimal thresholds. The UQ is a UTQ which has its reconstruction values located at the middle of each step, i.e., $\delta = \alpha/2$. A UQ with step size α has the same entropy as a corresponding UTORQ, but has an additional MSE penalty of $e^{-\alpha/2}(\delta(\alpha) - \alpha/2)^2$ due to its suboptimal reconstruction offsets. However, the UQ is commonly used, and its encoding and decoding rules are very simple. It combines the simplicity of the UTQ encoding rule with the simplicity of the URQ decoding rule..The UQ has also been shown by Gish and Pierce to be asymptotically optimal for a variety of source pdf models (including the exponential and Laplacian pdf models) and a variety of distortion measures at high bit rates (i.e., as the bit rate approaches infinity, the optimal ECSQ approaches a UQ) [18].

It is also possible for us to try to approximate the optimal dead-zone ratio with some approximating function. The optimal dead-zone ratio is, of course, specified by finding the solution of (45). However, solving or approximating the solution to (45) closely may be viewed as overkill for designing an ECSQ which is nearly optimal. For this reason, we also consider one very simple approximation—quantizers having a fixed constant dead-zone ratio z_c, which we shall call a *constant dead-zone ratio* quantizer (CDZRQ). Since the URURQ has $z_c = 1$, it is a special case of a CDZRQ. For $z_c \neq 1$, we can no longer use the simple URQ decoding rule allowable for the URURQ, but we may obtain better performance.

Fig. 4 shows the performance of Optimal ECSQ and UQ with respect to the distortion-rate bound $D_L(R)$ for the Laplacian source. The x-axis of the plot shows the entropy rate $R = H_L$ in bits per sample, and the y-axis shows the quantization signal-to-noise ratio

$$\text{QSNR} = 10\log_{10}(\sigma^2/D_L^{(\infty)})\ \text{dB} \tag{47}$$

where $D_L^{(\infty)}$ is the quantizer mean-square error distortion. The points plotted for the distortion-rate bound $D_L(R)$ were computed using the Blahut–Arimoto algorithm [19]–[21] using a discretization of the source in a manner similar to that described in [9]. We only show optimal ECSQ and UQ in Fig. 4, as the performance of UTORQ and URURQ would be visually indistinguishable from that of optimal ECSQ on such a plot.

We see from Fig. 4 that ECSQ has performance very close to the $D_L(R)$ bound at low bit rates, and then gradually lags in performance at higher bit rates, as noted in [10]. In fact, just as we were able to show that the maximum Lagrange multiplier for the exponential pdf was infinite for convex

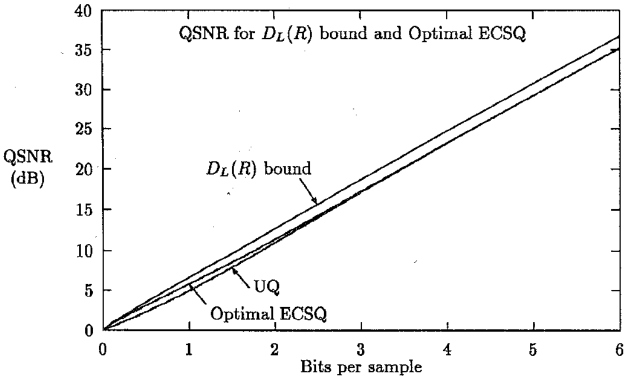

Fig. 4. $D_L(R)$ bound and ECSQ performance.

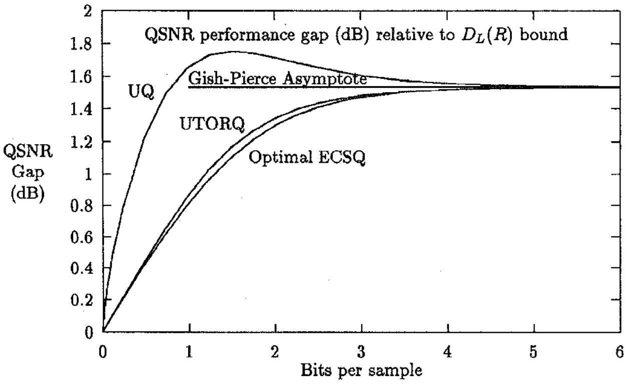

Fig. 5. ECSQ QSNR performance gap relative to $D_L(R)$ bound.

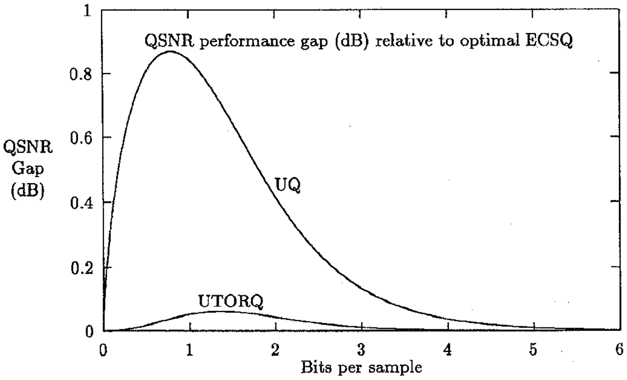

Fig. 6. ECSQ QSNR performance gap relative to optimal ECSQ.

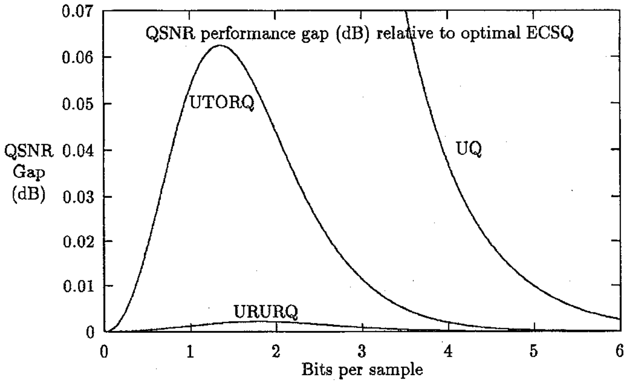

Fig. 7. ECSQ QSNR performance gap relative to optimal ECSQ.

distortion measures, we can do the same for the Laplacian pdf. This implies two important conclusions:

1) That the slope of the $D_L(R)$ bound for convex distortion measures (such as MSE) is infinite at $R = 0$ (since the slope must be at least as large in magnitude for $D_L(R)$ as it is for ECSQ).
2) That the ECSQ performance curve fits the $D_L(R)$ bound closely near zero, since both its value and its slope are the same as those of $D_L(R)$ at the limit.

It is also worth noting that we can show that for the absolute error distortion measure $d(\Delta) = |\Delta|$, the maximal ECSQ Lagrange multiplier for the Laplacian pdf is $\ln(2)$. This shows that the ECSQ performance curve for absolute error also fits its $D_L(R)$ bound closely at low bit rates, since the bound $D_L(R) = \sigma 2^{-R}/\sqrt{2}$ has that same slope [19].

As observed by Goblick and Holsinger for Gaussian sources, and later derived precisely for more general pdf models by Gish and Pierce, the performance gap approaches $\log_2 \sqrt{\pi e/6} = 0.2546$ bits per sample, or $20 \log_{10} \sqrt{\pi e/6} = 1.5329$ dB at high bit rates [22], [18]. In order to examine the performance gap more closely, we plot just the gap itself in Fig. 5, as given by $10 \log_{10} (D_L^{(\infty)}/D_L(R))$. Fig. 5 shows the performance gaps for both the optimal ECSQ and the UTORQ. It does not show the performance gap for URURQ, because its performance would be visually indistinguishable from that of optimal ECSQ on such a plot.

In order to better show the comparative performance of the scalar quantizers, we plot in Figs. 6–8 just the QSNR gap in the performance of various quantizers, measured with respect to the QSNR of the optimal ECSQ. Fig. 6 compares the performance of UQ to that of UTORQ. While the UQ is up to 0.87 dB worse than optimal ECSQ, the UTORQ stays within 0.0625 dB of optimality. However, the URURQ is even better. It significantly outperforms the UTORQ, so much so that its performance would be essentially indistinguishable from optimality in Fig. 6. The URURQ performance is shown alongside that of UTORQ in Fig. 7. Although the QSNR of the UTORQ becomes as much as 0.0625 dB lower than that of the optimal ECSQ near 1.36 bits per sample, the URURQ stays within about 0.0022 dB of optimality at all bit rates (as shown in Fig. 7).

Gish and Pierce showed that the optimal ECSQ should approach a UQ at high bit rates for a wide range of sources and distortion measures (a UQ being a UTQ with reconstruction values located midway between each pair of thresholds). This would seem to call into question our assertion that the optimal quantizer is *not* a UTQ in general, and that the URURQ is an improvement upon the UTQ. However, we note that using (31) it is easily shown that

$$\delta_{\text{mse}}(\alpha) \rightarrow \alpha/2 \quad \text{as } \alpha \rightarrow 0. \tag{48}$$

Thus the reconstruction values of the UTORQ and URURQ will approach the midpoints of each step at high bit rates, and the URURQ dead-zone size $2(\alpha - \delta(\alpha))$ will approach the same width α as the other steps. Therefore, the URURQ and UTORQ approach the UQ at high bit rates, as does the

TABLE I
QSNR FOR $D_L(R)$ MSE BOUND AND VARIOUS SCALAR QUANTIZERS

Rate (bps)	$D_L(R)$	Opt-ECSQ	UQ	UTORQ	URURQ	$z_c = 0.973$
0.015625	0.183	0.168115	0.075633	0.168105	0.168115	0.168115
0.031250	0.333	0.303178	0.148466	0.303131	0.303177	0.303178
0.062500	0.601	0.543407	0.291451	0.543193	0.543402	0.543407
0.125000	1.082	0.970287	0.574113	0.969308	0.970267	0.970284
0.250000	1.953	1.734472	1.143077	1.730069	1.734387	1.734457
0.500000	3.562	3.133696	2.332980	3.115493	3.133349	3.133622
1.000000	6.631	5.820695	4.977373	5.767671	5.819458	5.820445
2.000000	12.668	11.371078	10.952085	11.326802	11.368971	11.370924
4.000000	24.711	23.193305	23.155672	23.191333	23.193113	23.193067
8.000000	48.793	47.260480	47.260317	47.260479	47.260479	47.260418

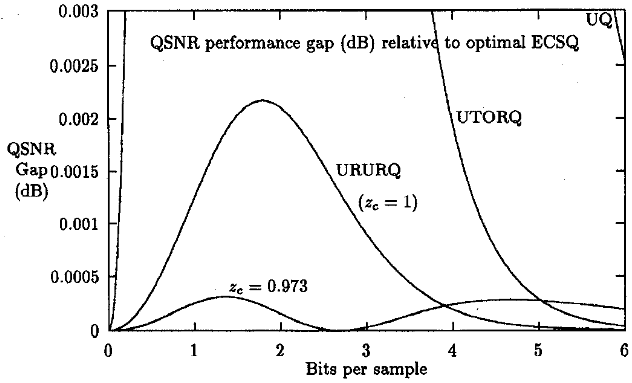

Fig. 8. ECSQ QSNR performance gap relative to optimal ECSQ.

optimal ECSQ. This limiting tendency of the optimal ECSQ to approach a UQ also holds for many other distortion measures, as can be shown using the reasoning found in [18].

We can also design a CDZRQ by searching for the best constant dead-zone ratio z_c in the minimax sense—minimizing the maximum QSNR penalty. (However, a CDZRQ with $z_c \neq 1$ does not approach a UQ at high bit rates.) The performance of a minimax CDZRQ, specified by $z_c = 0.973$, is shown in Fig. 8. Its QSNR performance is always within 0.0003 dB of optimality, and its performance would be indistinguishable from optimality if plotted in any of the other figures.

In order to facilitate comparative research efforts, we provide some numerical results for these various quantizers in Table I.

V. Conclusions

We have described a noniterative approach to the design of ECSQ's for two common sources by exploiting the memoryless property of the exponential pdf. The new method is extremely fast and is optimal for a general difference-based distortion measure and for a restricted or unrestricted number of quantization levels. In addition to designing and measuring the performance of optimal ECSQ, we have evaluated the performance of the popular UTQ method and have described improved simplified quantizers, the URURQ and CDZRQ, which have better performance than the UTQ.

Appendix
$\lambda_e^{(\max)}$

This Appendix proves the following assertions regarding $\lambda_e^{(\max)}$ of (24):

1) that $\lambda_e^{(\max)}$ is infinite if the distortion measure $d(\Delta)$ increases more than linearly as its argument deviates from zero;
2) that $\lambda_e^{(\max)}$ is finite and can be reached with a finite α only if the distortion measure increases less than linearly as its argument deviates from zero; and
3) that $\lambda_e^{(\max)}$ is finite but is reached only by infinite α in (23) for a linearly increasing distortion measure.

We attribute the key steps of the proof of the first two of these assertions to aid provided by Ramchandran and Orchard [15], and the key steps of the proof of the third assertion to aid provided by Chen and Chou [16]. (The assertions were originally formed by examining $\lambda_e(\alpha)$ for various examples of distortion measures.)

We begin by noting from (24) that

$$\lambda_e^{(\max)} \geq \lim_{\alpha \to \infty} \lambda_e(\alpha)$$

and that the denominator $B(e^{-\alpha})$ of (23) obeys

$$B(e^{-\alpha}) \to \alpha e^{-\alpha}/\ln(2) \quad \text{as } \alpha \to \infty. \tag{49}$$

We then note that the numerator of (23) can be separated into three terms as $\gamma(\infty) - \gamma(\alpha) - e^{-\alpha}\gamma(\infty)$, where the final term $e^{-\alpha}\gamma(\infty)$ approaches zero much more rapidly than the denominator for large α, provided $\gamma(\infty)$ is finite. We thus focus on the remaining expression $\gamma(\infty) - \gamma(\alpha)$. We define the property "increasing more than linearly" by $d(\Delta)/\Delta \to \infty$ and "increasing less than linearly" by $d(\Delta)/\Delta \to 0$ as $\Delta \to \infty$.

We now show that $\gamma(\infty) - \gamma(\alpha)$ becomes much greater than $\alpha e^{-\alpha}$ for large α when the distortion measure increases more than linearly, causing $\lambda_e^{(\max)}$ to be infinite. Since $\gamma(\alpha) \leq \gamma(\alpha, \delta)$ for any δ, we have

$$\gamma(\infty) - \gamma(\alpha) \geq \gamma(\infty) - \gamma(\alpha, \delta(\infty)) \tag{50}$$

$$= \int_{\alpha}^{\infty} d(x - \delta(\infty)) e^{-x}\, dx \tag{51}$$

$$= d(\alpha - \delta(\infty)) e^{-\alpha} + \int_{\alpha}^{\infty} d'(x - \delta(\infty)) e^{-x}\, dx \tag{52}$$

$$> d(\alpha - \delta(\infty)) e^{-\alpha} \tag{53}$$

using integration by parts, where $d'(\Delta)$ is the first derivative of $d(\Delta)$, which we have assumed exists and is positive and finite for finite $\Delta > 0$. Thus if $d(\Delta)$ increases more rapidly than linearly for large Δ, then the numerator of (23) dominates (provided $\delta(\infty)$ is finite), yielding $\lambda_e^{(\max)} = \infty$.

We take a similar approach to the converse assertion. Consider a distortion measure $d(\Delta)$ which increases less than linearly for large Δ. Since we have assumed that the slope is finite for a finite argument, and since the distortion measure increases less than linearly for large Δ (making the slope finite for an infinite argument as well), we can find a finite maximum slope constant C_m and a distortion offset constant C_o, such that $d(\Delta) \leq C_m(C_o + C_m\Delta)$ and $d'(\Delta) \leq C_m$ for $\Delta \geq 0$. We also note that $\lambda_e(\alpha)$ is finite for any finite α, and obtain

$$\begin{aligned}\gamma(\infty) - \gamma(\alpha) &\leq \gamma(\infty, \delta(\alpha)) - \gamma(\alpha) &(54)\\ &= \int_{\alpha}^{\infty} d(x - \delta(\alpha))e^{-x}\,dx &(55)\\ &= d(\alpha - \delta(\alpha))e^{-\alpha} + \int_{\alpha}^{\infty} d'(x - \delta(\alpha))e^{-x}\,dx &(56)\\ &\leq C_m[(C_o + \alpha - \delta(\alpha))e^{-\alpha} + e^{-(\alpha - \delta(\alpha))}] &(57)\\ &\leq C_m[\alpha + C_o + e^{\delta(\infty)}]e^{-\alpha} &(58)\end{aligned}$$

and thus

$$\lim_{\alpha \to \infty} \lambda_e(\alpha) \leq C_m \ln(2) \qquad (59)$$

and therefore, $\lambda_e^{(\max)}$ is finite.

For the borderline case in which distortion increases precisely linearly (although not necessarily symmetrically), we can describe the distortion measure in terms of a slope scaling $C_m > 0$, an offset constant C_o, and a skew $b > 0$ as

$$d_{\text{lin}}(\Delta) = \begin{cases} C_o - C_m\Delta, & \text{for } \Delta < 0\\ C_o + bC_m\Delta, & \text{for } \Delta \geq 0. \end{cases} \qquad (60)$$

From this we can obtain

$$\begin{aligned}\gamma_{\text{lin}}(\alpha, \delta) &= C_o(1 - e^{-\alpha}) + C_m[(1 + b)e^{-\delta} - (1 + be^{-\alpha})\\ &\quad \cdot (1 - \delta) - b\alpha e^{-\alpha}], \qquad \text{for } 0 \leq \delta \leq \alpha &(61)\\ \delta_{\text{lin}}(\alpha) &= \ln\left(\frac{1 + b}{1 + be^{-\alpha}}\right) &(62)\\ D_e^{(\infty)}(\alpha) &= C_o + C_m\left[(1 + be^{-\alpha})\ln\left(\frac{1 + b}{1 + be^{-\alpha}}\right)\right.\\ &\quad \left. - \alpha e^{-\alpha}\right] \Big/ (1 - e^{-\alpha}) &(63)\\ \lambda_e(\alpha) &= C_m\{[b\alpha - (1 + b)\ln(1 + b)]e^{-\alpha}\\ &\quad + (1 + be^{-\alpha})\ln(1 + be^{-\alpha})\}/B(e^{-\alpha}). &(64)\end{aligned}$$

Beginning with

$$\lim_{\alpha \to \infty} \lambda_e(\alpha) = bC_m \ln(2) \qquad (65)$$

$$\lim_{\alpha \to 0} \lambda_e(\alpha) = 0 \qquad (66)$$

we need to prove that $\lambda_e(\alpha) < bC_m \ln(2)$ for $\alpha \in (0, \infty)$. Defining $p = e^{-\alpha}$ and multiplying both sides of the inequality by $B(p)$, we find that this is equivalent to proving for $p \in (0, 1)$ that

$$(1 + bp)\ln(1 + bp) - p(1 + b)\ln(1 + b) + b(1 - p)\ln(1 - p) < 0. \qquad (67)$$

The function composed of the first and second terms of this expression is concave and has zero-valued limits at zero and one, and thus is negative for $p \in (0, 1)$, and third term of the expression is also always negative. Thus the entire expression is negative and the assertion is proved.

Acknowledgment

The author wishes to thank K. Ramchandran, M. T. Orchard, F. Chen, H. Chou, K. Popat, and G. Bjøntegaard for help with questions relating to this work, and the PictureTel research staff for comments and support. The helpful comments of the anonymous reviewers are also greatly appreciated.

References

[1] K. Nitadori, "Statistical analysis of ΔPCM," *Electron. Commun. in Japan*, vol. 48, pp. 17–26, Feb. 1965.

[2] H. Lanfer, "Maximum signal-to-noise-ratio quantization for Laplacian-distributed signals," *Informations-Und Systemtheorie in der Digitalen Nachrichtentechnik (Inform. Syst. Theory in Digital Commun.)*, vol. 65, p. 52, 1978. NTG-Report, VDE-Verlag GmbH, Berlin, Germany.

[3] P. Noll and R. Zelinski, "Comments on 'Quantizing characteristics for signals having Laplacian amplitude probability density function'," *IEEE Trans. Commun.*, vol. COM-27, pp. 1259–1260, Aug. 1979.

[4] W. C. Adams, Jr., and C. E. Giesler, "Quantizing characteristics for signals having Laplacian amplitude probability density function," *IEEE Trans. Commun.*, vol. COM-26, no. 8, pp. 1295–1297, 1978.

[5] M. D. Paez and T. H. Glisson, "Minimum mean-squared-error quantization in speech PCM and DPCM systems," *IEEE Trans. Commun.*, vol. COM-20, pp. 225–230, Apr. 1972.

[6] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 7–12, Mar. 1960.

[7] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 129–137, Mar. 1982.

[8] T. Berger, "Minimum entropy quantizers and permutation codes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 149–157, Mar. 1982.

[9] K. Popat, "Scalar quantization with arithmetic coding," M.S. thesis, MIT, Cambridge, MA, June 1990.

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

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

[12] G. J. Sullivan, "Optimal entropy constrained scalar quantization for exponential and Laplacian random variables," in *IEEE Int. Conf. Acoust., Speech, Signal Proc. (ICASSP)*, Apr. 1994, pp. V-265–268.

[13] J. C. Darragh, "Subband and transform coding of images," Ph.D. dissertation, Univ. of California, Los Angeles, May 1989. Tech. Rep. UCLA-ENG-89-53.

[14] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 36, pp. 1445–1453, Sept. 1988.

[15] K. Ramchandran and M. T. Orchard, personal communication, Nov. 1993.

[16] F. Chen and H. Chou, personal communication, Apr. 1995.

[17] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert *W* function." to appear, *Advances in Computational Mathematics*, http:// pineapple.apmaths.uwo.ca/~rmc/ papers/ LambertW.ps.Z

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

[19] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[20] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 460–473, Apr. 1972.

[21] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 14–20, Jan. 1972.

[22] T. J. Goblick and J. L. Holsinger, "Analog source digitization: A comparison of theory and practice," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 323–326, Apr. 1967.