THE OTHER ASYMPTOTIC THEORY OF LOSSY SOURCE CODING

D. L. NEUHOFF

ABSTRACT. Rate-distortion theory, as initiated by Shannon in his celebrated 1948 paper, is a well known theory of lossy source coding. It is an asymptotic theory in the sense that the performance it prescribes is approachable only in the limit as code dimension increases. Less well known is the "other" asymptotic theory of lossy source coding, which goes by the names of high-rate, high-resolution and asymptotic quantization theory. This theory prescribes the performance of codes with a given dimension and asymptotically large rate. The purpose of the present paper is to compare and contrast the two theories, and to highlight some recent results in high-rate quantization theory. It is the thesis of this paper that high-rate quantization theory has surpassed rate-distortion theory in its relevance to practical code design, because of its ability to identify key characteristics of good codes and to analyze the performance of codes with complexity-reducing structure.

1. INTRODUCTION

Source coding is the process of encoding to produce a binary representation of the data from some source and decoding to create a reproduction of the data from the binary representation. In this paper, we are concerned with the data from discrete-time, real-valued, stationary sources – the most illuminating case. As these cannot be encoded and decoded without introducing errors, the codes that are used are often called *lossy*.

The efficiency or compression of a lossy source code is measured by its *rate*, which is the average number of bits produced by the encoder per source symbol. The accuracy or fidelity of such a code is characterized by its *distortion*, which is a measure of the average loss associated with substituting the reproductions for the original data. Another important aspect of a source code is the complexity of implementing its encoder and decoder, as for example measured by the amount of storage required and the number of arithmetic operations required per sample.

The principal goals of a source coding theory include finding the following (for a given source):

(1) the least distortion of codes with rate R;
(2) the least distortion of codes with dimension k and rate R;

Key words and phrases. Lossy Source Coding, Vector Quantization, Rate-Distortion Theory, Rate-Distortion Function, High-Rate Quantization Theory, Asymptotic Quantization Theory, High-Resolution Quantization Theory, Bennett's Integral, Zador's Formula.

This work was supported by NSF grant NCR-9105647.

(3) the least distortion of codes with complexity c (with respect to some appropriate measure of complexity) and rate R;
(4) the least distortion of codes in a given class C with rate R;
(5) the key characteristics of codes that are optimal (or nearly so) in the above senses; and
(6) methods to design optimal (or nearly optimal) codes.

In this paper, we compare and contrast how two competing theories of lossy source coding attempt to serve these goals. The most well known is Shannon's rate-distortion theory. The "other", lesser-known theory has been called high-rate, high-resolution and asymptotic quantization theory.

2. Background

We focus on *vector quantizers* (VQ's). These are characterized by a *dimension* k, a *size* N, a *quantizing partition* $S = \{S_1, \ldots, S_N\}$ of $\Re^k$ into N *cells*, a *codebook* of *quantization points* $C = \{v_1, \ldots, v_N\} \subset \Re^k$ and a set of *binary codewords* $B = \{b_1, \ldots, b_N\}$, each having $\lceil \log_2 N \rceil$ bits. (The restriction to fixed length binary codewords is for simplicity of discussion and implementation.) When a source sequence $x = (x_1 \ldots x_k)$ lies in cell S_i of the quantizing partition, the encoder produces b_i as the binary representation of x, and the decoder produces $y = v_i$ as the reproduction of x. Overall, the VQ is characterized by the *quantization* rule $Q(x) = v_i$ when $x \in S_i$.

The rate of a vector quantizer is $R = k^{-1} \log_2 N$. (We choose this instead of $\lceil \log_2 N \rceil / N$ to simplify discussion.) As a measure of distortion, we choose mean squared error $D = k^{-1} E \|X - Q(X)\|^2$, where $\|x - y\|$ denotes k-dimensional Euclidean distance and where the expectation is with respect to the k-dimensional source distribution, which we presume to be characterized by a density $p_k(x)$. (Later, we mention generalizations to other distortion measures.)

Given a codebook, the conceptually simplest method of encoding with the best quantizing partition is to compute the squared distance between the given source vector x and each quantization point. The encoder outputs the binary codeword corresponding to the closest quantization point. The decoder uses the binary codeword to address a table containing the quantization points. This *full search* method of quantization requires storage and arithmetic complexity proportional to $N = 2^{kR}$, which grows exponentially with kR, the dimension-rate product. Thus complexity is a serious issue in vector quantization. Indeed, finding low complexity codes is a principal goal of modern source coding.

As indicated by the first of the aforementioned goals, given a source, we would like to find the least distortion of any VQ with rate R. Viewed as a function of R, this least distortion is called the VQ Operational Distortion-Rate Function (ODRF)[1] and denoted $\delta(R)$. Therefore, the first goal is equivalent to finding $\delta(R)$. Similarly, each of the second through fourth goals is equivalent to finding a corresponding ODRF. For example, $\delta_k(R)$ denotes the least distortion of k-dimensional VQ's with rate R. In summary, a principal goal of source coding theory is to find the ODRF's of various classes of codes.

[1]An ODRF, sometimes called an OPTA function, need only be "attainable" in the sense that there are codes whose performance comes arbitrarily close to it.

Of special interest, is finding how the ODRF's of various classes of codes relate to their complexity. For instance, one would like to know how close $\delta_k(R)$ is to $\delta(R)$ for small to moderate values of k, because the complexity of coding increases rapidly with k.[2] As another example, the ODRF for a class of structured vector quantizers will indicate how its complexity-reducing structure affects its performance.

3. Rate-Distortion Theory

Shannon's rate-distortion theorem, the third theorem of his celebrated 1948 paper [1] (and expanded upon in his 1959 paper [2] and generalized by others [3, 4]) achieves the first goal by showing that for stationary, ergodic sources (satisfying a moment finiteness condition)

$$\delta(R) = \mathcal{D}(R)$$

where $\mathcal{D}(R)$ is the information theoretic *distortion-rate function* (DRF)[3], defined as follows (c.f. [3, 5, 6]):

$$\mathcal{D}(R) \triangleq \lim_{k\to\infty} \mathcal{D}_k(R) \ .$$

where

$$\mathcal{D}_k(R) \triangleq \inf_{q\in Q_k(R)} \frac{1}{k} E\|X-Y\|^2 \ ,$$

where $Y = (Y_1 \ldots Y_k)$ is the output of the *test channel* q with X as the input, where $Q_k(R) \triangleq \left\{q(y|x) : k^{-1} I_q(X;Y) \le R\right\}$, is a collection of k-dimensional test channels, and where $I_q(X;Y) \triangleq \int p_k(x) q(y|x) \log_2 \left(p_k(x)q(y|x)/p(y)\right) dx$ is the *information* between input and output of the test channel q. In other words, rate-distortion theory relates the operational quantity $\delta(R)$ that we wish to know to the information theoretic quantity $\mathcal{D}(R)$. The latter can be evaluated analytically (c.f. [5, 6]) or numerically [7].

It is well known that, except for degenerate cases,

$$\delta_k(R) > \mathcal{D}_k(R) \ .$$

For small k, this bound is usually rather weak, and the kth-order distortion-rate function cannot be viewed as having much operational significance.

There are several well-known bounds to $\mathcal{D}(R)$ and $\mathcal{D}_k(R)$ (c.f. [5, 6]). An upper bound derives from known parametric formulas for $\mathcal{D}_k(R)$ and $\mathcal{D}(R)$ for Gaussian sources and from the fact that among sources with a given covariance function, the Gaussian distortion-rate function is largest. The simplest and most widely used lower bounds are the Shannon lower bounds:

$$\begin{aligned} \mathcal{D}(R) &\ge \frac{1}{2\pi e} 2^{2(h-R)} \triangleq \mathcal{D}_{slb}(R) \\ \mathcal{D}_k(R) &\ge \frac{1}{2\pi e} 2^{2(h_k-R)} \triangleq \mathcal{D}_{k,slb}(R) \end{aligned}$$

where h is the source *differential entropy-rate* and h_k is the *kth-order differential entropy*, as defined, respectively, by $h \triangleq \lim_{k\to\infty} h_k$ and $h_k \triangleq k^{-1} \int p_k(x) \log_2 p_k(x) dx$.

[2] It is straightforward to show that $\delta_k(R)$ approaches $\delta(R)$ from above as k tends to infinity (by showing subadditivity).

[3] Actually, Shannon showed the equivalent result that the least rate achievable by codes with distortion D is given by the rate-distortion function $\mathcal{R}(D)$, which is the inverse of $\mathcal{D}(R)$.

Both of these bounds are asymptotically tight in the sense that the ratio of the DRF to the lower bound goes to one as R tends to infinity.

4. High-Rate Quantization Theory

The central result of high-rate quantization theory is Zador's theorem [8, 9], which shows that

$$\lim_{R\to\infty} \frac{\delta_k(R)}{Z_k(R)} = 1 \tag{4.1}$$

where $Z_k(R)$ is the Zador-Gersho function, defined as

$$Z_k(R) \triangleq c_k \alpha_k 2^{-2R} \tag{4.2}$$

where $\alpha_k \triangleq \left(\int p_k(x)^{k/(k+2)} dx\right)^{(k+2)/k}$ depends on the source distribution and c_k, called the *tiling moment* of k-dimensional space, is the least *normalized moment of inertia* (n.m.i.) among k-dimensional polyhedra that tesselate. The n.m.i. of a cell P about a point y is $k^{-1} \int_P \|x-y\|^2 dx \ \mathrm{vol}(P)^{-(k+2)/k}$. It follows from (4.1) that

$$\lim_{R\to\infty} \frac{\delta(R)}{Z(R)} = 1$$

where

$$Z(R) \triangleq \lim_{k\to\infty} Z_k(R) \ .$$

Another important result of high-rate quantization theory is Bennett's integral for scalar quantizers [10] and its recent extension to VQ's [11]. The latter shows that for a VQ with many, mostly small cells, and neighboring cells having similar sizes and shapes

$$D \cong \frac{1}{N^{2/k}} \int \frac{m(x)}{\lambda(x)^{2/k}}\, p_k(x) dx \ , \tag{4.3}$$

where $\lambda(x)$ is the point density and $m(x)$ is the *inertial profile* of the VQ. As suggested by its name, $\lambda(x)$ is the density of quantization points in the vicinity of x, normalized to integrate to one. ($\lambda(x) \cong (N\mathrm{vol}(S_i))^{-1}$ when $x \in S_i$.) The inertial profile $m(x)$ is a function that equals, approximately, the normalized moment of inertia of the cells in the vicinity of x.[4]

Bennett's integral shows that the inertial profile (cell shapes) and the point density (cell sizes) are key characteristics of a VQ with many points. It is natural, then, to seek the best inertial profile and point density. Although never formally proved, there is a widely believed conjecture due to Gersho [9] that, for large N, most cells of the best k-dimensional quantizers are congruent, approximately, to the tesselating polytope with minimum normalized moment of inertia. Assuming this to be the case, the best inertial profile is the constant function $m(x) = c_k$. Assuming a constant inertial profile such as this, either calculus of variations or

[4] Eq. (4.3) is made precise [11] by showing that $N^{2/k} D_N$ converges to the integral in (4.3), where $\{D_N\}$ are the distortions of a sequence of quantizers whose point densities and inertial profiles converge to λ and m, respectively, as the size N increases to infinity.

Hölder's inequality can be used to show that Bennett's integral is minimized by the point density

$$\lambda_k^*(x) = \frac{p_k(x)^{k/k+2}}{\int p_k(x)^{k/k+2}dx} \ . \tag{4.4}$$

Substituting $\lambda_k^*(x)$ and c_k into Bennett's integral yields the Zador-Gersho formula[5].

5. Comparisons

We now compare and contrast rate-distortion theory and high-rate quantization theory.

(1) The two theories are complementary: Rate-distortion theory prescribes the best possible performance of codes with a given rate and asymptotically large dimension. High-rate quantization theory prescribes the best possible performance of codes with a given dimension and asymptotically large rate. That is,

$$\delta_k(R) \cong \begin{cases} \mathcal{D}(R) \text{ for large } k \text{ and any } R \\ Z_k(R) \text{ for large } R \text{ and any } k. \end{cases}$$

(2) History: Both theories originated in 1948. Rate-distortion theory had a spectacular beginning and rapidly became well known. It was mostly developed by Shannon [1, 2] and is now fairly mature. See [5] (pp.291,293) for a survey of other early contributors, and [12] for a survey of more recent work. High-rate quantization theory had humble beginnings in Bennett's paper for scalar quantizers [10], developed slowly through the years through the efforts of many [13-21,9,22-29], and is not so well known. Recent results have significantly expanded its usefulness [11,30-32], and it is still developing.

(3) Underlying Principles: Rate-distortion theory is a deep and elegant theory based on the law of large numbers. High-rate quantization theory is a simpler, less elegant theory based on geometric characterizations and integral approximations over small cells.

(4) Siblings: Rate-distortion theory has siblings, also based on the law of large numbers, in almost lossless block source coding, channel coding and multiuser information theory. Siblings of high-rate quantization theory are the error probability analyses in channel coding based on minimum distances and average power analyses for the additive Gaussian channel based on the continuous approximation.

(5) Distortion measures: Rate-distortion theory requires a per-letter fidelity criteria, whereas high-rate quantization theory requires a difference distortion measure. Both, however, provide the most results for the squared-error distortion measure. Each requires the source to satisfy a finite moment condition, related to the specific distortion measure.

(6) Sources: Rate-distortion theory is intended primarily for stationary, ergodic sources, although there are generalizations to asymptotically stationary and nonergodic sources (c.f. [12]). High-rate quantization theory is, fundamentally, a theory devoted to finite-dimensional random vectors, as opposed to random processes. However, for stationary sources, one may let dimension increase and obtain nice limiting formulas. Both theories give nice analytical formulas for Gaussian sources.

[5] Zador [8] found the form of (4.2); Gersho [9] recognized that the multiplicative constant is c_k.

(7) Code structures: Rate-distortion theory applies primarily to vector quantizers, or *block codes* as they were originally called. However, there are converse theorems that show that no codes of any kind with rate R or less achieve distortion less than $\mathcal{D}(R)$. And other code structures, such as variable-length, sliding-block, tree and trellis have been shown to contain codes achieving $\mathcal{D}(R)$.

High-rate quantization theory was also developed primarily for vector quantizers. However, its ability to focus on vector quantizers with a specific dimension is a major strength. Another, is that it can be tailored to specific classes of VQ's. For example, it has long been used to find the ODRF's of various scalar quantizer based codes such as PCM, transform coding and even DPCM, which is not a VQ. It has also been used to analyze lattice codes. And the recent extension of Bennett's integral to vector quantizers has lead to finding the ODRF's for tree-structured VQ [31] and two-stage VQ [32].

(8) Performance vs. complexity: Rate-distortion theory specifies the fundamental limits to the performance of lossy source coding without regard to cost or complexity. On the other hand, high-rate quantization theory, which analyzes the performance of codes with a given dimension and, sometimes, with a given structure, may be viewed as taking complexity into account, or at least, quantitatively showing the effects of complexity reducing structure. For example, in lattice quantization, the point density is a constant (far from optimal), in TSVQ the point density is good but the cell shapes are no better than cubes [31], and in two-stage VQ, the point density is affected but the cell shapes are optimal [32].

(9) Relationships: The following relations hold between distortion-rate functions and Zador-Gersho functions (c.f.[6]). The inequalities marked with "•" become tight as dimension k increases. Those marked with "+" become tight as R increases.

$$\begin{array}{ccccc} \mathcal{D}(R) & \overset{+}{\geq} & \mathcal{D}_{slb}(R) & = & Z(R) \\ \wedge\!\mid \bullet & & \wedge\!\mid \bullet & & \wedge\!\mid \bullet \\ \mathcal{D}_k(R) & \overset{+}{\geq} & \mathcal{D}_{k,slb}(R) & \overset{\bullet}{<} & Z_k(R) \end{array}$$

Thus, the Shannon lower bound and Zador's function are closely related.

(10) How large is "sufficiently large"? Both theories predict the performance of optimal codes as a certain parameter (k or R) becomes large. We now examine how large these parameters must be for the predictions to be reasonably accurate. Consider the case of an IID Gaussian source. By designing VQ's with the LBG algorithm [33] at low rates and by using high-rate theory at high rates, we find that a dimension of 8 to 14 is needed to give distortion within 1 dB of $\mathcal{D}(R)$. Thus, one may say that rate-distortion theory is applicable for VQ's with dimensions of the order of 10 or greater. For sources with memory, the dimension needs to be larger, roughly by the "length" of the memory.

On the other hand, designing VQ's with the LBG algorithm shows that, as a general rule, the Zador-Gersho formula is very accurate (to within tenths of a dB) for rates 3 or greater. It has also been found that Bennett's integral and other asymptotic analyses are accurate for roughly the same range of rates (c.f. [31]). Thus, one may say that high-rate quantization theory is accurate for rates 3 and up. Moreover, it is generally accurate enough at rate 2 to give indicative results.

(11) Code design: Each theory leads to its own design philosophy. On the one hand, rate-distortion theory shows that good high-dimensional codes have quanti-

zation points that are typical of the output distribution of the test channel that achieves the distortion-rate function. Indeed, one can obtain a good code by randomly generating such quantization points. In the special case of IID Gaussian sources, the quantization points can be uniformly distributed either in a sphere or on the surface of a sphere. More generally, the points can be uniformly distributed in the region of *typical sequences*, where the source density is approximately uniform (see the *asymptotic equipartition property* (AEP) [5]).

On the other hand, high-rate quantization theory indicates that one obtains a good code with dimension k by making the quantization cells as spherical as possible, i.e. they should be shaped like the polyhedron with n.m.i. equal to c_k, and by causing the quantization points to be distributed according to the optimal point density, λ_k^*. For example, one can design essentially optimal scalar quantizers by choosing $y_i = F^{-1}((2i-1)/2N)$ [10, 13, 27], where $F(x) = \int_{-\infty}^{x} \lambda_k^*(u)du$ is the distribution function corresponding to the optimal point density.

These two design philosophies are not inconsistent. When both k and R are large,high-rate theory shows that the points of an optimal k-dimensional VQ are distributed as $\lambda_k^*(x)$, which in this case is approximately equal to $p_k(x)$. It follows, then, from the AEP, that the points of an optimal k-dimensional VQ are spread uniformly throughout the typical sequences, just as rate-distortion theory indicates.

Which design philosophy is more useful? At low rates (say 1 bit per sample or less), one has no choice but to look to rate-distortion theory. But at moderate to high rates, we argue that the high-rate philosophy is the better choice. The reason is that codes designed by the rate-distortion theory approach require very large dimensions. To see this consider an IID Gaussian source, a desired rate R and a VQ of dimension k with 2^{kR} points distributed uniformly throughout a spherical support region. This is the ideal code suggested by rate-distortion theory. One may lower bound its distortion by assuming that source vectors outside the spherical support region are quantized to the closest point on the surface of the sphere, and by assuming that the cells within the support region are perfect k-dimensional spheres. In this case, at moderate to large rates (say rate 10), after choosing the diameter of the support region to minimize this lower bound to distortion, it has been found [34] that the dimension k must be larger than 250 in order that the mean squared error is within 1dB of the distortion-rate function.

We conclude that the fact that VQ's with dimension about 10 can have distortion within 1 dB of the distortion-rate function is not due to the rate-distortion theory design philosophy, the AEP, nor the use of ideal spherical codes. Rather it is due to the fact that good codes with small to moderate dimensions have appropriately tapered point densities, as suggested by high-rate theory.[6]

12. The nature of the error process: From rate-distortion theory, we know that good VQ's with large dimension have codewords typical of the output of the test channel that achieves the distortion-rate function. One can use this fact to learn something about the error process resulting from an optimal high-dimensional VQ. For example, for a Gaussian source with memory, one discovers how the spectrum of the noise relates to that of the source (c.f. [5]).

[6] It is likely that the originally perceived need for high dimensions (based on the AEP) delayed the serious consideration of vector quantizers. The breakthrough was the paper by Linde, Buzo and Gray [33], which showed that VQ's with moderate dimension were both good and feasible.

Recently, in high-rate theory there has been developed an asymptotic formula for the density of the error that results from a VQ with many mostly small cells, with neighboring cells having similar sizes and shapes [30]. Taking the second moment of this density yields, of course, Bennett's integral. The form of the density depends intimately on the point density and the cell shapes. The original motivation for finding the asymptotic error density was a high-rate analysis of two-stage VQ [32], where a second stage VQ is fed the error resulting from a first-stage VQ. With knowledge of the first-stage error density, Bennett's integral may be used to determine the overall distortion. It turned out that the formula is very sensitive to the shapes of the quantization cells. Thus, it has also proved useful in understanding the cell shapes of other VQ's, for example TSVQ [31].

13. Succesive approximation: Many vector quantizers operate in a *successive approximation* or *progressive* fashion, whereby a low-rate coarse quantization is followed by a sequence of finer and finer quantizations, which add to the rate. Tree-structured, multi-stage and hierarchical VQ's are examples of such. Successive approximation is useful in situations where the decoder needs to produce rough approximations of the data from the first bits it receives and, subsequently, to refine the approximation as more bits are received. Moreover, successive approximation VQ's are often structured in a way that makes them simpler than unstructured VQ's. Indeed, the three examples just cited are known more for their good performance with low complexity than for their progressive nature. An important question is whether the distortion of a successive refinement code will be larger than a one shot code. On the one hand, a recent rate-distortion theory analysis [35] has shown there are situations where successive approximation can be done without loss of optimality. On the other hand, recent high-rate analyses of TSVQ [31] and two-stage VQ [32] have quantified the loss of these particular codes. Thus, both theories have something to say about successive refinement.

6. Conclusions

As indicated in the abstract, the thesis of this paper is that high-rate quantization theory has surpassed rate-distortion theory in its relevance to practical code design and analysis, because of its ability to identify key characteristics of good codes and to be specialized to particular codes with complexity-reducing structure.

High-rate quantization theory also provides results concerning the performance of scalar and vector quantizers with variable-length binary codes [8,18,5(p.173),20,9,22], which space has precluded discussing.

References

1. C.E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.,* vol. 27, pp. 397-423, 623-656, 1948.
2. C.E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat'l. Conv. Rec. part 4,* pp. 142-163, 1959.
3. R.G. Gallager, *Information Theory and Reliable Communication,* New York: Wiley, 1968.
4. T. Berger, "Rate distortion theory for sources with abstract alphabets and memory," *Information and Control,* vol. 13, pp. 254-273, 1968.
5. T. Berger, *Rate Distortion Theory,* Englewood Cliffs, NJ: Prentice-Hall, 1971.
6. R.M.Gray, *Source Coding Theory,* Boston: Kluwer, 1990.
7. R.E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inform. Theory,* vol. IT-18, pp. 460-473, July 1972.

8. P. Zador, "Development and Evaluation of Procedures for Quantizing Multivariate Distributions," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1964.
9. A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory,* vol. IT-23, pp. 373-380, July 1979.
10. W.R. Bennett, "Spectra of quantized signals," *B.S.T.J.,* vol. 27, pp. 446-472, July 1948.
11. S. Na and D.L. Neuhoff, "Bennett's integral for vector quantizers, and applications," *1990 IEEE Int'l Symp. on Inform. Theory*, Jan. 1990.
12. J.C. Kieffer, "A survey of the theory of source coding," to appear in *IEEE Trans. Inform. Theory.*
13. P.F. Panter and W. Dite, "Quantization distortion in pulse count modulation with nonuniform spacing of levels," *Proc. IRE,* vol. 39, pp. 44-48, Jan. 1951.
14. S.P. Lloyd, "Least squares quantization in PCM," Bell Laboratories Memorandum 1957, also published in *IEEE Trans. Inform. Theory,* vol. IT-28, pp. 129-137, March 1982.
15. M.P. Schutzenberger, "On the quantization of finite-dimensional messges," *Inform. and Control,* vol. 1, pp. 153-158, 1958.
16. J.J.Y. Huang and P.M. Schultheiss, "Block quantization of correlated Gaussian variables," *IEEE Trans. Commun.,* vol. COM-11, pp. 289-296, Sept. 1963.
17. V.R. Algazi, "Useful approximations to optimum quantization, "*IEEE Trans. Commun.,* vol. COM-14, pp. 297-301, June 1966.
18. H. Gish and J.R. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory,* vol. IT-14, pp. 676-683, Sept. 1968.
19. P. Elias, "Bounds on performance of optimum quantizers," *IEEE Trans. Inform. Theory,* vol. IT-16, pp. 172-184, March 1970.
20. I. Csiszar, "Generalized entropy and quantization problems," 6th Prague Conference, 1973.
21. R.M. Gray and A.H. Gray, "Asymptotically optimal quantizers," *IEEE Trans. Inform. Theory,* pp. 143-144, Jan. 1977.
22. J.A. Bucklew and N.C. Gallagher, "Some properties of uniform step size quantizers," *IEEE Trans. Inform. Theory,* vol. IT-26, pp. 610-613, Sept. 1980.
23. Y. Yamada, S. Tazaki and R.M. Gray, "Asymptotic performance of block quantizers with difference distortion measures," *IEEE Trans. Inform. Theory,* vol. IT-26, pp. 6-14, Jan. 1980.
24. J. A. Bucklew, "Companding and random quantization in several dimensions," *IEEE Trans. Inform. Theory,* vol. IT-27, pp. 207-211, March 1981.
25. J.A. Bucklew and G.W. Wise, "Multidimensional asymptotic quantization theory with r th power distortion measures," *IEEE Trans. Inform. Theory,* Vol. IT-28, pp. 239-247, Mar. 1982.
26. N.C. Gallagher, J.A. Bucklew, "Properties of minimum mean squared error block quantizers," *IEEE Trans. Inform. Thy.,* vol. IT-28, pp. 105-107, Jan. 1983.
27. S. Cambanis and N. Gerr, "A simple class of asymptotically optimal quantizers," *IEEE Trans. Inform. Theory,* vol. IT-29, pp. 664-676, Sept. 1983.
28. J.A. Bucklew, "A note on optimal multidimensional companders," *IEEE Trans. Inform. Theory,* vol. IT-29, pp. 279, March 1983.
29. J.A. Bucklew, "Two results on the asymptotic performance of quantizers," *IEEE Trans. Inform. Theory,* vol. IT-30, pp. 341-348, Mar. 1984.
30. D.H. Lee and D.L. Neuhoff, "The asymptotic distribution of the error in scalar and vector quantizers," *1990 IEEE Int'l Symp. Inform. Theory*, San Diego, Jan. 1990.
31. D.L. Neuhoff and D.H. Lee, "On the Performance of Tree-Structured Vector Quantization," *ICAASP91*, vol. 4, pp. 2277-2280, Toronto, May 1991.
32. D.H. Lee and D.L. Neuhoff, "An Asymptotic Analysis of Two-Stage Vector Quantization," *1991 IEEE Int'l Symp. on Inform. Theory,* p. 316, Budapest, June 1991.
33. Y. Linde, A. Buzo and R.M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.,* vol. COM-28, pp. 84-95, Jan. 1980.
34. A. Balamesh and D. L. Neuhoff, unpublished notes, 1992.
35. W.H.R. Equitz and T.M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory,* vol. IT-37, pp. 268-275, Mar. 1991.

Department of Electrical Engineering and Computer Science, Ann Arbor, Michigan 48109
E-mail address: neuhoff@eecs.umich.edu