Correspondence

Piecewise Uniform Vector Quantizers

FEDERICO KUHLMANN MEMBER, IEEE, AND
JAMES A. BUCKLEW, MEMBER, IEEE

Abstract — **The companding model for quantizer design and analysis has been widely applied in the scalar quantization case. However, if the signal to be quantized is a vector, then the optimum companding system can be designed for only a limited number of distributions. On the other hand, multidimensional piecewise linear companders can be designed for any signal density, generating quantizers that are uniform on each region of the compander. These systems, while not optimal, can have asymptotic performance arbitrarily close to the optimum. Furthermore, their analysis and implementation can be simpler than those of optimal systems. Piecewise linear companders for asymptotic multidimensional quantization are analyzed, and a method for their design is suggested.**

I. Introduction

Digital processing of discrete-time signals has become a major research area, in spite of the fact that many information-bearing signals appear originally as waveforms which are continuous in time as well as in amplitude. Hence quantizers play an important role in the theory and practice of modern-day signal processing. Extensive results have been developed on scalar quantization and, more recently, on vector quantization. The advantage of vector (or multidimensional) quantization over scalar quantization is implicitly shown by the proofs of the rate-distortion theorems of information theory. Unfortunately, rate-distortion theory does not provide "constructive proofs" but results entirely of the "existence" nature. The benefits of multidimensional schemes (for a fixed finite dimension) were explicitly studied (for the asymptotic case of fine quantization) first by Zador [1], and subsequently by Gersho [2].

Although there is great promise inherent in the theory, the problem of how to design large dimensional quantizers for a variety of source distributions has proven to be the major stumbling block in implementation. Designing an optimal vector quantizer is basically equivalent to finding a partition of the vector space and assigning a representative point to each partition such that a predefined distortion measure between input and output is minimized. Unfortunately, the optimal partitions in higher dimensional spaces are unknown for even the simplest source distributions and the most common distortion measures. (The exception is that in two-dimensional space for fine quantization of a uniform distribution and a mean square error distortion measure, hexagons are the optimal partitions.) An important design algorithm is reported by Linde *et al.* [5], in which the resulting quantizer yields a local minimum for the distortion. This algorithm appears to be the method of choice among researchers for the design of low rate (small number of bits per sample) multidimensional quantizers. Other significant work has been done on quantizers for some specialized sources such as multidimensional Gaussian [3] or Laplacian [4]. Our somewhat more general approach is presented as a design technique for high-rate multidimensional quantizers.

We approach multidimensional quantizer design by formulating the problem as a distortion minimization over an output point density function. This method is a generalization of Bennett's results [6] in which he models a zero memory nonuniform quantizer by the cascade connection of a zero memory nonlinearity, a uniform scalar quantizer, and the inverse of the first nonlinearity. Such a scheme is termed "companding," because the first nonlinearity is a compressor while the second performs an expander operation. In the scalar case any nonuniform quantizer can be modeled in this fashion. Furthermore, a fixed compressor/expander characteristic can yield the asymptotically optimal rate of decay to zero of the error as the number of output levels goes to infinity. An expression for the minimum asymptotic quantization noise attainable by any vector quantizer of a given dimension was derived by Gersho [2]. However, it is not always possible to represent an asymptotically optimal multidimensional quantizer with this block companding scheme. It is shown in [7] that the optimal compressor characteristic must be conformal almost everywhere, a condition which cannot always be met. The problem we address in this correspondence is the design of nearly optimal vector quantizers, using the block companding model but restricting the class of compressors to be piecewise linear. This class of compressor characteristics is also interesting in its own right, as it has some robustness properties (described in [8]).

In Section II we define the problem and develop the main ideas to be used for its solution. The results concerning quantizer design with compressors for piecewise constant densities are developed in Section III. Examples of optimum vector compander designs for two-dimensional quantizers are presented in Section IV, and Section V is devoted to a discussion of the robustness properties of these quantizers.

II. Statement of the Problem

Let the k-dimensional random vector x with probability density function $p(x)$ be the input to the k-dimensional quantizer. The quantizer partitions R^k into N disjoint and exhaustive regions B_i, $i=1,2,\cdots,N$, and quantizes each input vector x by means of the following mapping:

$$Q(x) = y_i, \qquad \text{if } x \in B_i$$

where usually $y_i \in B_i$. The number N represents the number of quantizer output levels. The performance of the quantizer is measured by the per-dimension (or per-sample) distortion

$$D = \frac{1}{k} E\{ \|x - Q(x)\|_2^r \}$$

where $\|\cdot\|_2$ denotes the Euclidean distance norm, $x \in \boldsymbol{R}^k$ is the quantizer input, and $Q(x)$ is the quantizer output. Let $Q_0(x)$ denote the optimum quantizer. Its asymptotic performance, as derived by Zador [1] and Gersho [2], is given by

$$D_0 = C(k,r) N^{-r/k} \| p(x) \|_{k/k+r} \tag{1}$$

where $\| p(x) \|_\alpha = [\int p(x)^\alpha \, dx]^{1/\alpha}$, and $C(k,r)$ is a constant which only depends on k and r (for example, $C(1,2)$ is $1/12$).

Now let $g(x)$ be a piecewise constant approximation to $p(x)$. Assume that $p(x)$ has compact support S. If the support of $p(x)$

Manuscript received August 19, 1986; revised July 1, 1987.

F. Kuhlmann is with the Electrical and Computer Engineering Department, University of Wisconsin, Madison, WI 53706, on leave from the University of Mexico.

J. A. Bucklew is with the Electrical and Computer Engineering Department, University of Wisconsin, 1415 Johnson Drive, Madison, WI 53706.

IEEE Log number 8824498.

is unbounded, one can choose a sufficiently large compact subset of the support and restrict $p(x)$ to that region, thus having, in addition to the quantizer noise, a truncation error. We will assume, however, that the truncation noise is smaller than a prescribed value and will thus be concerned in the following only with the so-called quantizer "granular" noise. For example, if a scalar zero-mean unit variance Gaussian random variable is truncated to ± 4 or ± 6, the differences between the optimal quantizer performance for the untruncated and the truncated random variables is less than 0.2 and 0.006 dB, respectively (for values of N between 8 and 128) [9]. Let $g(x)$ have the same compact support S as $p(x)$. The piecewise constant density $g(x)$ is related to $p(x)$ by the following. Let $g(x)$ have M segments and denote by C_i the compact region over which $g(x)$ has its ith value. Let m_i be the measure of C_i. The density $g(x)$ is then given by

$$g(x) = \frac{p_i}{m_i}, \qquad x \in C_i,\ i = 1,2,\cdots,M \tag{2}$$

where $p_i = \int_{C_i} p(x)\,dx$. It is easy to see that $g(x)$ is a valid density function.

Since $g(x)$ has a constant value over each of the regions $C_i, i = 1,2,\cdots,M$, the optimal (asymptotic) quantizer for $g(x)$ will be piecewise uniform, i.e., for each region C_i the quantizer is uniform and in each region the number of quantizer output levels depends upon the total number of levels N and the particular shape of $g(x)$ (i.e., the values of p_i and m_i). Note that, by uniform, we mean here that the point density function of the quantizer is uniform and not that it is necessarily a rectangular lattice. The minimum asymptotic distortion resulting from quantizing the random variable y corresponding to $g(x)$ with its optimal quantizer (call it Q_g) is then given by

$$D_g = \frac{1}{k} E\left\{ \|y - Q_g(y)\|_2^r \right\} = C(k,r) N^{-r/k} \|g(x)\|_{k/k+r}. \tag{3}$$

By definition, if we now try to quantize x with the quantizer Q_g, we find that

$$\frac{1}{k} E\left\{ \|x - Q_0(x)\|_2^r \right\} \le \frac{1}{k} E\left\{ \|x - Q_g(x)\|_2^r \right\}$$

where Q_0 is the optimal N-level quantizer for the data. However, as $M \to \infty$ (under some mild technical assumptions such as the maximum probability of any cell approaches zero and the cell boundaries remain approximately proportional), it can be shown that

$$\lim_{M \to \infty} g(x) = p(x) \text{ a.e.} - x$$

and

$$\lim_{M \to \infty} \|g(x)\|_{k/k+r} = \|p(x)\|_{k/k+r}.$$

Here and throughout this correspondence we require that $M \to \infty$ in a way such that the ratio of the sides of the regions remains constant. Two particular cases for the above ideas are the following. If $p(x) = 1/m, x \in S$, with $m = \int_S dx$, then obviously $M = 1$ and $g(x) = p(x), x \in S$. This corresponds to an originally uniformly distributed input. On the other hand, if $p(x) = 1/m_i$ for $x \in C_i$, then $g(x)$ can also be identical to $p(x)$ for M finite (corresponding to an input random variable which is piecewise uniform). These ideas imply that a near optimum quantizer for $p(x)$ can be designed by finding an optimum quantizer for an input vector with probability density $g(x)$. As the approximation to $p(x)$ by $g(x)$ becomes more accurate, the asymptotic distortion approaches its minimum value. In the next section we examine the design of optimum quantizers for these piecewise constant densities.

III. Compressors for Piecewise Constant Densities

Assuming that $g(x)$ is known, the design of the optimum quantizer consists in finding the number of quantization levels that correspond to each of the M partitions. This is accomplished by rewriting $\|g(x)\|_{k/k+r}$ as follows:

$$\|g(x)\|_{k/k+r} = \left[\sum_{i=1}^{M} \left(\frac{p_i}{m_i} \right)^{k/k+r} m_i \right]^{k+r/k} = \left[\sum_{i=1}^{M} p_i^{k/k+r} m_i^{r/k+r} \right]^{k+r/k}, \tag{4}$$

substituting (4) into Zador's expression (3),

$$D = C(k,r) N^{-r/k} \left[\sum_{i=1}^{M} p_i^{k/k+r} m_i^{r/k+r} \right]^{k+r/k}, \tag{5}$$

and rewriting (5) as

$$D = C(k,r) N^{-r/k} \left[\sum_{i=1}^{M} p_i^{k/k+r} m_i^{r/k+r} \right] \left[\sum_{j=1}^{M} p_j^{k/k+r} m_j^{r/k+r} \right]^{r/k}$$

$$= C(k,r) N^{-r/k} \left[\sum_{i=1}^{M} p_i p_i^{-r/k+r} m_i^{r/k+r} \right] \cdot \left[\sum_{j=1}^{M} p_j^{k/k+r} m_j^{r/k+r} \right]^{r/k},$$

which can be rearranged to form

$$D = \sum_{i=1}^{M} p_i C(k,r) \left[\frac{m_i^{k/k+r}}{N p_i^{k/k+r}} \sum_{j=1}^{M} p_j^{k/k+r} m_j^{r/k+r} \right]^{r/k}. \tag{6}$$

On the other hand, we can obtain the corresponding distortion by noting that for $x \in C_i$, $g_i(x) = g(x | x \in C_i) = 1/m_i$, and $g_i(x) = 0$ if $x \notin C_i$. Using (3) again yields

$$D_i = C(k,r) N_i^{-r/k} \|g_i(x)\|_{k/k+r} = C(k,r) N_i^{-r/k} \left[\frac{1}{m_i^{k/k+r}} m_i \right]^{k+r/k} = C(k,r) \left(\frac{m_i}{N_i} \right)^{r/k}. \tag{7}$$

In the above, N_i is the number of levels in C_i, and we note that

$$\sum_{i=1}^{M} N_i \le N. \tag{8}$$

The total distortion D_T can be written as the expected value of the D_i in (7) and thus

$$D_T = \sum_{i=1}^{M} p_i D_i = \sum_{i=1}^{M} p_i C(k,r) \left(\frac{m_i}{N_i} \right)^{r/k}. \tag{9}$$

Recall that D in (6) is the minimum quantization distortion. Thus, by setting $D_T = D$, we can directly solve for the optimum assignment of the numbers of quantization levels. Since all the quantities involved are non-negative, it follows that (by a Lagrange multiplier type of minimization)

$$\frac{m_i}{N_i} = \frac{m_i^{k/k+r}}{N p_i^{k/k+r}} \sum_{j=1}^{M} p_j^{k/k+r} m_j^{r/k+r},$$

and therefore

$$N_i = \frac{N p_i^{k/k+r} m_i^{r/k+r}}{\sum_{j=1}^{M} p_j^{k/k+r} m_j^{r/k+r}}. \tag{10}$$

We note that the N_i must be integers, so we must round (10) up or down (subject to (8) being satisfied). For the special case where $m_i = m$, we have

$$N_i = \frac{N p_i^{k/k+r}}{\sum_{j=1}^{M} p_j^{k/k+r}}. \tag{11}$$

If we do a uniform partition, i.e., $p_i = 1/M$, then the optimum assignment is

$$N_i = \frac{N m_i^{r/k+r}}{\sum_{j=1}^{M} m_j^{r/k+r}}. \tag{12}$$

Equations (11) and (12) can be used to get an estimate of the required number of regions M for the performance of Q_g to be similar to D_0. In the general case, substituting (11) in (9) we find that

$$D_0 \le D_T \le C(k,r)\left(\frac{V}{N}\right)^{r/k}\left(\sum_{i=1}^{M} p_i^{k/k+r}\right)^{k+r/k}$$

where $V = mM = \int_S dx$. Substituting (12) in (9) we obtain

$$D_0 \le D_T \le C(k,r)\frac{1}{M}N^{-r/k}\left(\sum_{i=1}^{M} m_i^{r/k+r}\right)^{k+r/k}$$

Equation (10) then suggests a method for designing near-optimum quantizers given $p(x)$: first, partition the support S into M regions; then quantize each region using a multidimensional uniform quantizer with the number of quantization levels specified by (10). Each region C_i is mapped by a simple translation into the input space of the N-level optimum quantizer and the measure of each region is scaled appropriately so that the region is uniformly quantized with N_i levels. After quantization, the inverse mapping is used to obtain the output for the near optimum quantizer.

To this point we have not discussed a method of partitioning the input space. Given that we want to partition the support S into M regions, there are two important points to be considered. The first point involves the optimum quantization of each region. As stated before, each region is quantized with a uniform k-dimensional quantizer. However, because of distortion near the edges of regions we can calculate the performance only for a large number of quantization levels. For example, a region with only two levels may not be very well quantized. Therefore, the number of quantization levels N must be large relative to M so that all of the regions are finely quantized.

Second, we observe that S can be partioned into M regions in an infinite number of ways. Thus the performance of the piecewise compander is a function of the number of regions and their shapes. Ideally, we want to choose the partitioning method that results in the minimum distortion for the near-optimum quantizer. However, the shapes of the regions also determine the difficulty in implementing the piecewise compander. Therefore, a trade-off exists between the ease of implementation and the performance of the near-optimum quantizer.

When N is large we can partition the input space into a large number of regions (subject to N/M remaining large). In this situation the shapes of the individual regions have only a small effect on the quantizer performance. Thus we can partition the input space into hypercubes to simplify the implementation. By quantizing each component of the input vector with a one-dimensional quantizer, the hypercube that contains the input vector is easily determined. When N is small the number of regions must also be small. For this case an efficient partitioning of the input space may be necessary to obtain near-optimal performance. The piecewise compander is then implemented using a search to locate the regions corresponding to each input vector.

At this point we want to recall that the piecewise linear compander will be used with the true input density $p(x)$. One way to determine how well the optimum quantizer and the piecewise linear compressor used with the true density match is to compare, for each region C_i, the distortion resulting from the optimum quantizer and from the piecewise linear one, when the input is the random vector with density $p(x)$.

Following Gersho's development [2], it can be shown that the optimum quantizer allocates N_i^* output levels to the ith region, where

$$N_i^* = N\|p(x)\|_{k/k+r}^{-k/k+r}\int_{C_i} p(x)^{k/k+r}\,dx.$$

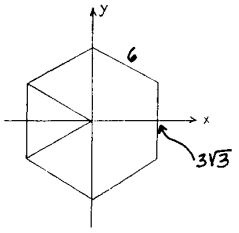
On the other hand, if $p_i(x)$ is the density of x, conditioned on $x \in C_i$, then on C_i the average asymptotic minimum distortion is

$$D_i^* = C(k,r)(N_i^*)^{-r/k}\|p_i(x)\|_{k/k+r}.$$

Upon observing the difference $D_i - D_i^*$, where D_i is given by (7), it can be determined which regions (if any) should be modified or further subdivided.

IV. Example

We now present an example of the design and the performance analysis of a two-dimensional piecewise linear compressor. We consider x to be bivariate Gaussian, with covariance equal to the identity matrix and mean zero. Since the input support must be contained in the support of the pieceiwse linear compressor, and for a finite number of regions the latter is finite, we truncate the input to the circle with radius $3\sqrt{3}$. The support of the piecewise linear compressor was chosen to be a regular hexagon with $a = 6$, as shown in Fig. 1(a). The regions m_i were selected to be trapezoids with base parallel to the base of each component equivalent triangle as shown in Fig. 1(b). The compressor was designed under the equal m_i assumption, i.e., $m = 6a^2\sqrt{3}/4M$, where M is the number of regions. The performance is measured in terms of mean-squared error.



(a)

(b)

Fig. 1. (a) Support of compressor. (b) Regions of compressor.

The number N_i of output levels per region was determined by means of (11), and these numbers were then used to determine the distortion D_N (normalized by N). The resulting distortions ("granular" quantization noise only) are presented in Table I (for

TABLE I
PIECEWISE LINEAR COMPRESSOR PERFORMANCE

M	D_N
300	2.043
180	2.050
120	2.063
90	2.081
60	2.131
30	2.397
12	3.946

reference, the minimum normalized distortion, calculated using (1) for an untruncated, bivariate Gaussian density with zero mean and the above covariance, is 2.015).

It is noteworthy that even with a small number of regions there is a reasonable performance degradation. For example, with $M = 30$ (i.e., only five regions per triangle) there is a loss of only 0.75 dB, while two regions per triangle ($M = 12$) yields less than a 3-dB loss. All dB comparisons are with reference to the asymptotically optimal error constant 2.015.

The normalized distributions of levels (N_i/N) for different values of M, and for a plane cutting the $x-y$ plane at $y = 0$, are shown in Fig. 2. We observe in this figure that larger numbers of regions tend to modify N_i/N more drastically for those regions located close to the origin, i.e., they try to improve the quantization procedure over the regions with a larger probability mass. To illustrate this point, we present in the numerical values of N_i/N for the $M = 12$, 30, and 60 cases (Table II).

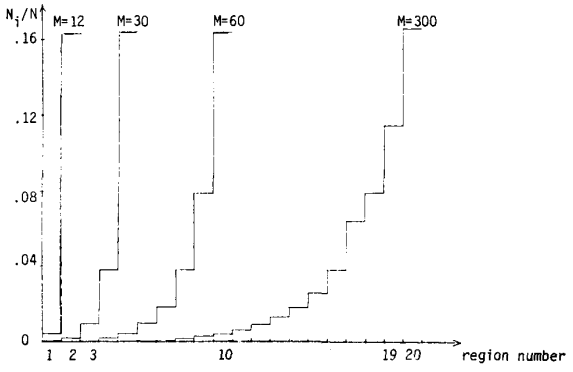


Fig. 2. Level distributions for $y = 0, x \le 0$.

TABLE II
N_i/N FOR DIFFERENT REGIONS

M	1	2	3	4	5	6	7	8	9	10
12	0.004	0.162	—	—	—	—	—	—	—	—
30	0.0004	0.0017	0.0069	0.0294	0.1283	—	—	—	—	—
60	0.0001	0.0003	0.0005	0.0011	0.0028	0.0047	0.0096	0.0198	0.0414	0.0869

It is convenient to point out that the larger the support of the piecewise linear compressor, the larger the distortion for a given number of regions. If the support of the true density is unbounded, a good strategy is to choose the support of the piecewise linear compressor as small as possible without introducing intolerable truncation errors. If, in our example, the support of the compressor is reduced so that $a = 5$, then the distortion (normalized) for $M = 12$ is reduced to 3.011, and an improvement of 1.17 dB with respect to the value given in Table I is achieved; this value is only 1.74 dB worse than the optimum value. For $M = 60$, the distortion is 2.059, or about 0.15 dB better than the value in Table I, and only 0.09 dB worse than the optimum. To quantize an input signal, one must first determine which region the particular realization falls into (say k), and then quantize it uniformly with the corresponding N_k levels.

V. COMMENTS ON THE ROBUSTNESS OF PIECEWISE CONSTANT COMPRESSORS

Optimum quantizer design is primarily based on exact knowledge of the statistical model for the data to be quantized. However, if the statistical model is not completely known, it is of practical interest to study the performance degradation of the quantizer designed for certain input statistics but used for a different set of statistics. The first important results in this direction were published by Morris and Vandelinde [10] and Bath and Vandelinde [11], [12] for zero memory quantization. Swaszek and Thomas [13] analyzed the design problem when the input statistics are only known through a histogram, and Kazakos [8] analyzed the robustness problem for vector quantizers. To place our piecewise constant compressor in this context, recall from (3) that $D_g = C(k,r)N^{-r/k}\|g(x)\|_{k/k+r}$ and that, according to (4),

$$\|g(x)\|_{k/k+r} = \left| \sum_{i=1}^{M} p_i^{k/k+r} m_i^{r/k+r} \right|^{k+r/k}.$$

Thus it can be seen that the piecewise uniform quantizer distortion is a function of the true input density $p(x)$ only through the quantities p_i and $m_i, i = 1,2,\cdots, M$. In other words, if Q_g is used to quantize any random vector with probability density $f(x)$ such that $\int_{C_i} f(x)\,dx = p_i$, $i = 1,2,\cdots, M$, then the distortion is given by (3). In [8], it is shown that a piecewise linear compressor does yield a solution to a minimax type of formulation when it is known that the inpout density $f(x)$ is such that $\int_{C_i} f(x)\,dx = p_i$, i.e., when $f(x)$ is a member of the so-called p-point uncertainty class.

These ideas lead to the following conclusion of practical interest. If a nonuniform quantizer is to be designed for a nominal density which is known only through its p_i and m_i, the piecewise linear compander design will yield good performance (if not optimal) and will allow the input density to change somewhat without modifying the distortion (unless the p_i or the m_i change).

VI. DISCUSSION AND CONCLUSION

We point out what has been accomplished and what remains to be done. We have presented a technique inspired by one-dimensional piecewise linear compander theory and tried to generalize it to several dimensions. One divides the support of the data probability distribution into regions (usually hypercubes) and then implements a multidimensional uniform quantizer in each region.

The form of the optimal multidimensional uniform quantizer is not known (even for asymptotically large numbers of levels) for dimensions greater than two. Good multidimensional uniform quantizers are known, however, for certain dimensions (e.g., Sloane [14] has given a fast computational method for eight-dimensional quantizers). For correlated or dependent data, a general rule of thumb is that most of what is gained by going to multidimensional quantizers is attributable to the form of the compressors. Whether one uses a one-dimensional product lattice as the k-dimensional uniform or the "optimal k-dimensional uniform" (if it were known) only gives an asymptotic factor of improvement (as $k \to \infty$) of $2\pi e/12 \approx 1.42$.

The main contribution of this work is to point out a method of implementing vector quantizers on nonuniform data distributions in the fine quantization or high rate limit. The methodology is general in that we do not have to assume a particular kind of nonuniform distribution, e.g., Gaussian or Laplacian. The main drawback to our method is that we do not have a method for deciding how to change the regions in which to divide the support of an arbitrary data distribution. In our example of a two-dimensional Gaussian distribution, the method was *ad hoc*

and based on symmetry considerations. More should be done to correct this drawback in low-dimensional cases. However, for fine quantization, we feel that this method is the only viable alternative for arbitrary data distributions.

REFERENCES

[1] P. Zador, "Development and evaluation of procedures for quantizing multivariate distributions," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1964 (University Microfilm 64-9855).

[2] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 373–380, July 1979.

[3] J. A. Bucklew and N. C. Gallagher, Jr., "Quantization schemes for bivariate Gaussian random variables," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 667–671, Nov. 1979.

[4] T. R. Fischer, "A pyramid vector quantizer," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 568–583, July 1986.

[5] 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.

[6] W. R. Bennett, "Spectra of quantized signals," *Bell Syst. Tech. J.*, vol. 27, pp. 446–472, July 1948.

[7] J. A. Bucklew, "Companding and random quantization in several dimensions," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 207–211, Mar. 1981.

[8] D. Kazakos, "On the design of robust quantizers," in *1981 IEEE Telecommunications Conf. Rec.*, New Orleans, LA, Dec. 1981, pp. F4.5.1–F4.5.4.

[9] F. Kuhlmann and G. L. Wise, "Design considerations for asymptotically optimal piecewise linear companders," in *1981 IEEE Telecommunications Conf. Rec.*, New Orleans, LA, Dec. 1981, pp. F4.4.1–F4.4.4.

[10] J. M. Morris and V. D. Vandelinde, "Robust quantization of discrete-time signals with independent samples," *IEEE Trans. Commun. Technol.*, vol. COM-22, pp. 1897–1901, Dec. 1974.

[11] W. G. Bath and V. D. Vandelinde, "Robust quantizers designed using the companding approximation," in *Proc. 18th IEEE Conf. Decision and Control*, Ft. Lauderdale, FL, Dec. 1979, pp. 484–487.

[12] ____, "Robust memoryless quantization for minimum signal distortion," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 296–306, Mar. 1982.

[13] P. Swaszek and J. B. Thomas, "Quantization in unsure statistical environments," in *Proc. 15th Annu. Johns Hopkins Conf. Information Sciences and Systems*, Baltimore, MD, Mar. 1981.

[14] J. H. Conway and N. J. A. Sloane, "A fast encoding method for lattice codes and quantizers," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 820–824, Nov. 1983.

An Upper Bound on the Bit Error Probability of Combined Convolutional Coding and Continuous Phase Modulation

GÖRAN LINDELL
AND CARL-ERIK W. SUNDBERG, SENIOR MEMBER, IEEE

***Abstract*—Continuous phase modulation (CPM) is a class of digital constant-amplitude modulations with good combined power and bandwidth efficiency. We study the bit error probability properties of signals consisting of convolutional coding combined with partial response multilevel CPM. It is assumed that the channel is an additive white Gaussian noise channel and that the receiver performs coherent maximum likelihood sequence detection by means of the Viterbi algorithm. An upper bound on the bit error probability is derived, using the average generating function technique. The upper bound is evaluated numerically for a number of coded multilevel full-response CPM schemes. Simulation results are also presented. It is concluded that the free Euclidean distance is the best one-parameter description of the error probability for the considered class of signals for high signal-to-noise ratios. However, it is interesting to observe that the upper bound results show that the free distance alone yields pessimistic bit error probability behavior for some interesting cases.**

Manuscript received July 19, 1985; revised May 26, 1987. This work was supported by the National Swedish Board of Technical Development under Grant 83-3334. This work was partially presented at the International Symposium on Information Theory, Brighton, England, June 1985, and at the Global Telecommunications Conference (GLOBECOM '85), New Orleans, LA, December 1985.

G. Lindell is with Telecommunication Theory, University of Lund, Box 118, S-22100, Lund, Sweden.

C-E. W. Sundberg was with Telecommunication Theory, University of Lund, Lund, Sweden. He is now with AT&T Bell Laboratories MH-2C-483, 600 Mountain Avenue, Murray Hill, NJ 07974.

IEEE Log Number 8824504.

I. INTRODUCTION

An abundance of constant-amplitude digital modulation schemes has been proposed and analyzed in the recent literature. Many of these schemes fit in as special cases of the continuous phase modulation (CPM) class [1]–[5]. Memory is introduced into the transmitted signal by means of continuous phase and also by means of partial-response smoothing (correlative encoding). Uncoded partial-response CPM yields good combined power and bandwidth efficiency [1]. Continuous phase frequency shift keying (CPFSK) is a special case of the CPM class [1].

Here we shall consider combined convolutional coding and multilevel CPM, which introduces more memory in the transmitted signal. See [6]–[14] for background information on combined coding and modulation. An exact constant amplitude is maintained at all time instants in the coded CPM schemes, in contrast with the amplitude modulation and phase shift keying modulation considered in [6]. We will consider transmission over an additive white Gaussian noise channel only. Furthermore, it is assumed that the receiver is an ideal coherent maximum likelihood sequence detector (MLSD) using the Viterbi algorithm.

The transfer function technique for evaluating upper bounds on the error probability of convolutional codes was introduced in [15] and is described in detail in [16]–[18]. It was first applied to conventional linear convolutional codes. This technique was later extended to the more general cases of, e.g., nonlinear trellis codes, in [19], [20]. Also, symbol error probability for uncoded continuous phase modulation is evaluated in [21]. In [22]–[24], a detailed performance evaluation is presented for some trellis coded AM and PSK systems. The work in [16]–[24] is all based on [15] and an average transfer function technique is used.

In this correspondence we have applied the average transfer function technique to the case of coded continuous phase modulation [25], [26]. We have derived a general expression for an upper bound on the bit error probability for partial-response CPM with finite memory, a rational modulation index, an M-ary natural binary mapper, and a general convolutional code. An ideal Viterbi detector with infinite path memory is assumed. The upper bound is evaluated numerically for a number of interesting coded multilevel full-response CPFSK schemes. Simulation results for a Viterbi detector with finite path memory are also presented.

It is shown in [1] and [12] that CPM systems combined with convolutional codes are both more power and bandwidth efficient than uncoded CPM. The previous analysis is based on asymptotic error probability performance for high signal-to-noise ratios given by the minimum Euclidean distance. We shall show that these conclusions also hold for a range of signal-to-noise ratios when the upper bound on the bit error probability is used for performance evaluation. The gains given by calculations based on the free distance are actually sometimes pessimistic. An interesting result from the detailed numerical evaluation is that the minimum Euclidean distance error event for some coded CPM schemes only occurs for rather few transmitted signals. This follows from the fact that combined convolutional coding and CPM forms a nonlinear trellis code.

In Section II we give a brief system description of coded CPM and in Section III we briefly give the upper bound. In Section IV the numerical evaluations of the upper bound on the bit error