ADVANCES IN ELECTRONICS AND ELECTRON PHYSICS, VOL 72

Lattice Quantization

JERRY D. GIBSON

Department of Electrical Engineering
Texas A&M University
College Station, TX

and

KHALID SAYOOD

Department of Electrical Engineering
University of Nebraska–Lincoln
Lincoln, NE

I. Introduction

We are all familiar with the process of analog–to–digital (A–to–D) conversion, whereby continuous-time, continuous-amplitude signals are converted into a sequence of binary words suitable for storage or for manipulation in digital form. The process of A–to–D conversion consists of three distinct operations: sampling, quantization, and coding. When we sample a signal, we represent the continuous-time signal by a set of sample values taken at discrete time instants, and, as long as these samples are taken at a uniform rate greater than the Nyquist rate, there is no appreciable loss in fidelity as a result of this sampling operation. Quantization generates a

discrete-amplitude representation of the continuous-amplitude sample values, but unlike sampling, quantization produces a non-recoverable loss in fidelity. The combined operations of sampling and quantization generate a sequence of discrete-time, discrete-amplitude values, and this sequence is changed into digital form by coding, which assigns a distinct binary word to each allowable discrete-amplitude level at the quantizer output.

A standard A-to-D converter is an example of a uniform, scalar quantizer, and if the fidelity of the digitized output is not adequate, we simply select another A-to-D converter with more bits (quantization levels). When we do this, we are increasing the number of bits/sample required to store the digitized sequence or to transmit the sequence over a communications link. For two common and important sources, speech and images, the number of bits/sample (called the rate and denoted by R) required by straightforward A-to-D conversion to achieve acceptable fidelity can be excessive for many applications. As a result, a research area called *data compression* has emerged, which has as its goal the representation of a source sample with as few bits as possible while still maintaining adequate fidelity for the particular application at hand. Familiar examples of data compression systems are delta modulation (DM), logarithmic-pulse code modulation (log–PCM), and differential PCM (DPCM). See Jayant and Noll (1984) for details concerning these systems.

Vector quantization is a relatively new data compression technique which has been and continues to be the subject of intense research interest and which also is beginning to find applications in practical systems. The fundamentally different characteristic of vector quantization, as opposed to scalar quantization where each scalar sample is quantized individually, is that a block or vector of scalar quantities is formed and this vector is quantized as a single entity. The reader can perhaps imagine that this approach could prove useful if the scalar components of the vector are dependent or correlated, but it may be surprising to note that quantizing and coding of blocks (or vectors) always yields better theoretical performance than scalar quantization, even if the vector components (the scalars) are uncorrelated or independent. This last fact is a result from a branch of information theory called rate distortion theory, originally delineated by Claude Shannon (1948; 1959). Research in vector quantization has been pursued vigorously only within the last ten years primarily because of the following three reasons: (1) Although Shannon's results provide bounds on the performance of optimal data compression systems, they do not provide guidance as to how vector quantizers might be designed; (2) early rate distortion theory results were primarily concerned with Gaussian sources, and optimal block coding of these sources only offers an asymptotic gain of about 0.25 bit/sample over scalar quantization followed by optimal noiseless coding, which did not provide sufficient motivation for further investigations by researchers; and (3) block coding or vector quanti-

zation/coding requires operations in multidimensional space, which is not only mathematically more difficult than scalar quantization, but it also implies substantially increased complexity over a scalar approach.

Recent research has uncovered various vector quantizer design techniques, all of which are based upon one of two approaches, the iterative design procedure often called the Linde, Buzo Gray (LBG) algorithm (Linde, Buzo and Gray, 1980) or the specification of uniform quantizers by using lattices. The former approach generates a locally optimal vector quantizer design, but the quantization/encoding problem may be formidable. The lattice-based approach can greatly simplify the quantization operation, but the resulting quantizers are only optimal for uniformly distributed sources or asymptotically optimal as the number of output points becomes large. Nevertheless, experimental results using the various vector quantizer designs have indicated that substantial performance improvements are available with vector quantizers, and these results, coupled with further rate distortion theoretic results and studies of the asymptotic performance of vector quantizers have combined to intensify research and development efforts concerning vector quantization.

The present chapter attempts to introduce the concept of vector quantization and to describe how lattices can be used to advantage in the vector quantization process. Several excellent survey/tutorial articles have previously appeared in the literature (Gersho and Cuperman, 1983; Gray, 1984; Makhoul, Roucos, and Gish, 1985), and these papers are highly recommended. The development in this chapter differs from these papers in that their emphasis is on the LBG algorithm (see Gersho and Cuperman, 1983; Gray, 1984) while we are concerned almost wholly with the use of lattices in vector quantization. Additionally, these articles were written with the goal of minimizing the number of equations in their presentation in order to reach a wider audience. There is some overlap with Makhoul, Roucos and Gish, 1985, particularly on the topics of rate distortion theoretic and asymptotic performance results. Other perspectives on vector quantization are available in the tutorial/survey chapters written by Gersho (1986), Swaszek (1986), and Adoul (1987).

We begin our presentation with discussions of scalar quantization and vector quantization in Sections II and III, followed by reasons for the consideration of lattice-based vector quantizers in Section IV. Section V defines the various lattices of interest and develops their important properties, while Section VI illustrates the application of these lattices to designing vector quantizers. The utility of lattices for devising fast quantization algorithms is demonstrated in Section VII. Performance comparisons among scalar quantizers and the best known vector quantizers are present in Section VIII, including theoretical results, experimental results on synthetic sources, and

experimental results for speech and images. Current research areas and variations of lattice quantizers are developed in Section IX, followed by a few summary thoughts and conclusions in Section X.

II. Scalar Quantization

A scalar quantizer is a quantizer that discretizes only a single input sample at a time. An L-level scalar quantizer $Q(x)$ is determined by specifying $L+1$ values $x_0 < x_1 < \cdots < x_L$, called *step points* or *decision levels*, that partition the real line $\mathcal{R}$, and a set of L output points $y_1, y_2, \ldots, y_L$, such that if the input sample x satisfies $x_{i-1} \le x < x_i$, then $Q(x) = y_i$. A typical quantizer input-output characteristic for L even is shown in Fig. 1(a), which can be equivalently represented by the one-dimensional diagram in Fig. 1(b), where the hash marks are step points and the dots are output "levels" or points. Although the quantizer representation in Fig. 1(b) is not as familiar as the one Fig. 1(a), the Fig. 1(b) diagram generalizes easily to two dimensions. For L even, a symmetric, uniform L-level quantizer has step points $0, \pm\Delta, \pm 2\Delta, \ldots, \pm(L/2-1)\Delta$ with $x_0 = -\infty$ and $x_L = +\infty$, and output points $\pm\Delta/2, \pm 3\Delta/2, \ldots, \pm(L-1)\Delta/2$, where Δ is called the step size. An L-level, symmetric, nonuniform scalar quantizer has step points $0, \pm\eta_i\Delta$, $i = 1, 2, \ldots, (L/2)-1$, with $x_0 = -\infty$ and $x_L = +\infty$, and output points $\pm\xi_j\Delta$, $j = 1, 2, \ldots, L/2$, where the constants $\{\eta_i\}$ and $\{\xi_j\}$ can be selected to yield the desired quantizer characteristic (Jayant and Noll, 1984).

For the design or performance analysis of a quantizer, the input samples are regarded as sequences of random variables, since the quantizer has no way of knowing exactly what the next sample will be. If the input signal or source samples are independent and identically distributed, each with the probability density function (pdf) $f_x(x)$, and if the chosen error measure is $g[\hat{x} - x]$, then the average distortion can be expressed as

$$D = \int_{-\infty}^{\infty} g[\hat{x} - x] f_x(x)\, dx, \tag{1}$$

where $\hat{x}$ denotes the quantizer output for input value x. Limiting consideration to the squared error distortion measure, $g[\hat{x} - x] = (\hat{x} - x)^2$, we can write the distortion in terms of the step points and output levels as

$$D = \sum_{i=1}^{L} \int_{x_{i-1}}^{x_i} [y_i - x]^2 f_x(x)\, dx. \tag{2}$$

The entropy of the quantizer output is given by

$$H(Y) = -\sum_{i=1}^{L} p_i \log_2 p_i \tag{3}$$

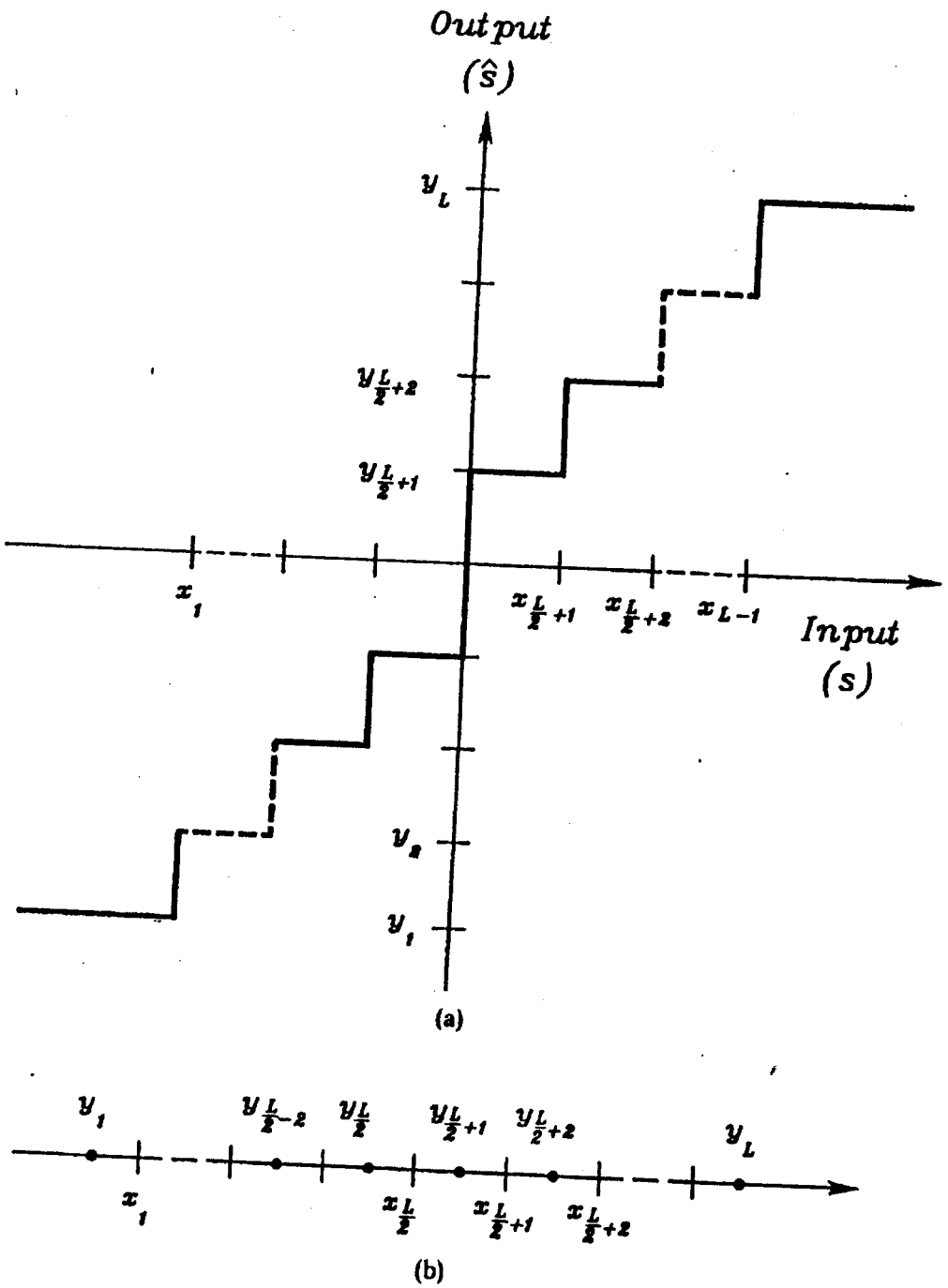

FIG. 1(a). Symmetric midriser quantizer (L Even). (b). One-dimensional quantizer characteristic.

where $p_i = P[x_{i-1} \leq x < x_i]$. Since the entropy of a discrete random variable is the minimum rate in bits/sample required to represent that variable for storage or transmission applications, D in Eq. (2) and $H(Y)$ in Eq. (3) specify the (minimum) rate necessary to achieve a distortion D with the chosen quantizer.

Assigning codewords to the quantizer output levels is an example of what is called noiseless source coding, which is so-named because the coding process is invertible. For an L-level quantizer, with L an integer power of 2, the simplest approach to choosing codewords is to assign a binary word of length $b = \log_2 L$ to each quantization level. If L is odd or $L \neq 2^b$, $\log_2 L$ is no longer an integer and the procedure must be modified. For example, if we have a three-level quantizer, we could assign a two digit binary word to each of the three quantization levels, thus yielding a rate of 2 bits/sample, but this would leave one binary word, or 25% of the codewords, unused. Another possibility would be to group the quantizer outputs into groups of three ternary symbols which are then mapped into five digit binary words, which is a rate 5/3 bit/sample source code. This latter code has less than 16% of the codewords unused, and is clearly more efficient than the former 2 bits/sample code. Other such codes could be pursued, but since the minimum rate required by a fixed-length code to represent three levels is $\log_2 3 = 1.585$ bits/sample, there is less than 0.1 bit/sample yet to be gained with this approach for the three-level quantizer.

If the probabilities of the quantizer output levels are known or can be estimated, an entropy coding technique such as Huffman coding can be used. This method assigns short codewords to highly probable levels and longer codewords to less probable levels, thus yielding a short average codeword length, $\bar{n}$. Since this procedure maps fixed-length blocks into variable-length binary words, buffering is necessary at both the transmitter and receiver. It is important to note that grouping quantizer output levels into blocks before using the Huffman procedure can produce a smaller $\bar{n}$, since the average codeword length $\bar{n}$ of the Huffman code satisfies

$$H(Y) \leq \bar{n} < H(Y) + \frac{1}{M}, \tag{4}$$

for a block of M quantizer output levels. Thus, coding pairs of output levels ($M = 2$) is at least as good as coding one level or sample ($M = 1$) at a time, and for large M, $\bar{n}$ can be made arbitrarily close to $H(Y)$. How large M has to be for efficient encoding depends upon the probabilities of the particular quantizer output levels of interest.

Other details concerning methods for scalar quantizer design and the various approaches to coding, that is, assigning binary words to the quantizer output points, can be found in Jayant and Noll (1984) and Gallager (1968). However, a few observations are in order. First, designing a scalar quantizer consists of partitioning the real line into a finite set of disjoint and exhaustive intervals and assigning a single output value to each interval. Second, efficient coding methods require that the output points be collected in groups or blocks, and in order to approach the minimum possible rate, delay and/or complexity may be significant.

III. Definitions and Motivation for Optimal Vector Quantization

In this section, we present some technical details concerning the performance improvement possible using optimal vector quantization, but we begin by carefully defining what is meant by vector quantization. For simplicity in the sequel, we abbreviate both vector quantization and vector quantizer by VQ. Whether VQ stands for vector quantization or vector quantizer should be evident from the context.

Let $\mathbf{X}$ be an N-component source vector with joint pdf $f(\mathbf{X}) = f(x_1, x_2, \ldots, x_N)$. An N-dimensional VQ is a function $\mathbf{Q}(\mathbf{X})$ that maps $\mathbf{X} \in \mathcal{R}^N$ into one of L output points with each output point corresponding to an output vector $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_L$, belonging to $\mathcal{R}^N$. The quantizer is completely specified by listing the L output vectors and their corresponding partitions of $\mathcal{R}^N$ into L disjoint and exhaustive regions denoted by $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_L$, so that $\mathbf{Q}(\mathbf{X}) = \mathbf{Y}_i$ if $\mathbf{X} \in \mathcal{P}_i$ for $i = 1, 2, \ldots, L$. An N-dimensional VQ is sometimes called a block quantizer with block length N. Throughout this chapter we confine the discussion to the mean squared error (MSE) per sample distortion measure given by

$$D = \frac{1}{N} E\{d(\mathbf{X} - \mathbf{Y})\} = \frac{1}{N} E\|\mathbf{X} - \mathbf{Q}(\mathbf{X})\|^2$$
$$= \frac{1}{N} \sum_{i=1}^{L} \int_{\mathbf{X} \in \mathcal{P}_i} \|\mathbf{X} - \mathbf{Y}_i\|^2 f(\mathbf{X})\, d\mathbf{X} \tag{5}$$

where $\|\cdot\|$ denotes the usual ℓ_2 norm.

For purposes of transmission or storage the output vectors $\mathbf{Y}_i$ are assigned a binary codeword $\mathbf{c}_i$ of length b_i bits. The average codeword length is thus

$$\bar{b} = \sum_{i=1}^{L} b_i P(\mathbf{X} \in \mathcal{P}_i) \text{ bits/vector} \tag{6}$$

so the average rate in bits/sample is

$$R = \frac{\bar{b}}{N}, \tag{7}$$

and hence

$$\frac{1}{N} H(\mathbf{Y}) \le R \le \frac{1}{N} \log_2 L. \tag{8}$$

The design of a VQ for a given distortion measure and input vector pdf requires the selection of the partitions $\mathcal{P}_i$ and the output vectors $\mathbf{Y}_i$, $i = 1, 2, \ldots, L$, often called the VQ codebook, such that the partitions are nonoverlapping and cover $\mathcal{R}^N$. In the scalar case ($N = 1$), the problem is

relatively simple since partitioning of $\mathcal{R}^1$ consists of choosing nonoverlapping intervals along the real line. For $N > 1$, the partitions $\mathcal{P}_i$ can take on any shape, and hence, there are infinitely many candidates for the set of optimum partitions. Even when the N-dimensional VQ is uniform, which implies that the $\mathcal{P}_i$ are just translates of the same shape, there are many different kinds of partitions which cover $\mathcal{R}^N$. For example, all triangles, quadrilaterals, and hexagons can be used to partition $\mathcal{R}^2$ (Gersho, 1979). This availability of many possible shapes for the partition makes the design procedure more complicated for $N > 1$, but it also provides a possible performance advantage over scalar quantization.

The performance of a VQ is completely determined by two quantities, the average distortion D in Eq. (5) and the required rate R in Eq. (7). If we wished to find the optimum performance possible using an N-dimensional VQ, we could take either one of two possible approaches (Gray and Davisson, 1974). We could fix the acceptable distortion D, and find the VQ that requires the minimum rate R which has distortion less than or equal to D, or we could fix the maximum rate R and find the VQ that achieves the smallest distortion D with rate less than or equal to R. However, both of these approaches, as described, imply that we must design an optimum N-dimensional VQ, and we do not know how to do this as yet. The ultimate bound on the performance of vector quantizers, indeed, on any data compression system, is provided by rate distortion theory as originally developed by Shannon (1948; 1959). The utility of rate distortion theory stems from the fact that the optimum performance theoretically attainable for any data compression system can be computed without actually designing such a system; in fact, all that is needed is a characterization of the source and a specification of the distortion measure.

To define the rate distortion function of a source, which specifies the minimum possible rate for a given distortion D, consider a discrete-time, continuous amplitude, stationary source that produces a sequence of scalar random variables, x_i, $i = 1, 2, \ldots$. A block of n source samples $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is represented as $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ by the source coder with probability density function $f_{Y|X}(\mathbf{y}|\mathbf{x})$, and so for the single-letter (sample) fidelity criterion

$$d(\mathbf{x} - \mathbf{y}) = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2, \tag{9}$$

the average distortion is given by

$$E\{d(\mathbf{x} - \mathbf{y})\} = \int\int f_X(\mathbf{x}) f_{Y|X}(\mathbf{y}|\mathbf{x}) d(\mathbf{x} - \mathbf{y})\, d\mathbf{x}\, d\mathbf{y} \tag{10}$$

where $f_X(\mathbf{x})$ is the joint probability density for the components of $\mathbf{x}$. Now, we

are only interested in values of the average distortion less than or equal to D, and thus we define

$$F_D = \{f_{Y|X}(\mathbf{y}\,|\,\mathbf{x}): E[d(\mathbf{x}-\mathbf{y})] \le D\} \tag{11}$$

as those transition probability densities between the source coder input and output which produce an average distortion less than or equal to D. The rate required to transmit these n-sample input blocks $\mathbf{x}$ with an average distortion of D or less is therefore

$$R_n(D) = \frac{1}{n} \inf_{f_{Y|X} \in F_D} \int\int f_X(\mathbf{x}) f_{Y|X}(\mathbf{y}\,|\,\mathbf{x}) \log \frac{f_{Y|X}(\mathbf{y}\,|\,\mathbf{x})}{f_Y(\mathbf{y})}\, d\mathbf{x}\, d\mathbf{y} \tag{12}$$

where

$$f_Y(\mathbf{y}) = \int f_X(\mathbf{x}) f_{Y|X}(\mathbf{y}\,|\,\mathbf{x})\, d\mathbf{x}. \tag{13}$$

Finally, we can define the rate distortion function of the given source as

$$R(D) = \lim_{n \to \infty} R_n(D) \tag{14}$$

which is the effective rate that the source produces information for reproduction with fidelity D. Therefore, $R(D)$ constitutes a lower bound on the rate required by any data compression scheme to achieve an average distortion of D or less (Jayant and Noll, 1984; Shannon, 1959; Gallager, 1968; Gray and Davisson, 1974; Berger, 1971).

While the rate distortion function $R(D)$ has received most of the attention in the information theory literature, it is seldom that a pre-specified average acceptable distortion value is available, and it is much more common in communication systems to know the maximum allowable bit rate. Because of these facts, it is more direct to define a distortion rate function $D(R)$ that is obtained by minimizing the average distortion subject to a constraint on the transmission rate R. The distortion rate function is the inverse of $R(D)$, and it can be defined precisely as follows. The rate required to represent the source coder output is given by the average mutual information between $\mathbf{x}$ and $\mathbf{y}$,

$$I(\mathbf{x},\mathbf{y}) = \int\int f_X(\mathbf{x}) f_{Y|X}(\mathbf{y}\,|\,\mathbf{x}) \log \frac{f_{Y|X}(\mathbf{y}\,|\,\mathbf{x})}{f_Y(\mathbf{y})}\, d\mathbf{x}\, d\mathbf{y}, \tag{15}$$

where $f_Y(\mathbf{y})$ is as shown in Eq. (13). Each choice of $f_{Y|X}(\mathbf{y}\,|\,\mathbf{x})$ gives rise to an average mutual information, and in minimizing the average distortion, we wish to consider only those conditional densities which correspond to a transmission rate less than or equal to some specified rate R. Thus, the admissible set of conditional probability densities is defined by

$$F_R = \{f_{Y|X}(\mathbf{y}\,|\,\mathbf{x}): I(\mathbf{x},\mathbf{y}) \le R\}. \tag{16}$$

Therefore, the minimum average distortion possible when transmitting the n-vector x at a rate of R bits/sample or less is

$$D_n(R) = \frac{1}{n} \inf_{f_{Y|X} \in F_R} E\{d(\mathbf{x} - \mathbf{y})\} \tag{17}$$

with the average distortion expressed by Eq. (10). It follows then that the minimum average distortion when transmitting at a rate of R bits/sample or less is

$$D(R) = \lim_{n \to \infty} D_n(R). \tag{18}$$

Equations (14) and (18) actually constitute the original motivation for considering vector quantizers since they imply that as the block length n of the vector x increases, the performance of the data compression system or source coder approaches the best performance possible, namely $R(D)$ and $D(R)$, respectively. A slightly closer connection with VQ can be made if we avoid the expressions involving average mutual information and consider the average distortion of an N-dimensional VQ as specified in Eq. (5). The distortion-rate approach to designing a VQ is to choose Q(X) to minimize the average distortion in Eq. (5) subject to the rate constraint in Eq. (8). Therefore, we have

$$\begin{aligned} D_N(R) &= \min_{Q(X)} \frac{1}{N} E\{d(\mathbf{X} - \mathbf{Y})\} \\ &= \min_{Q(X)} \frac{1}{N} E\|\mathbf{X} - \mathbf{Q}(\mathbf{X})\|^2 \end{aligned} \tag{19}$$

over all Q(X) which satisfy

$$\frac{1}{N} H(\mathbf{Q}(\mathbf{X})) = \frac{1}{N} H(\mathbf{Y}) \leq R. \tag{20}$$

The distortion rate function can be obtained from Eq. (19) as

$$D(R) = \lim_{N \to \infty} D_N(R). \tag{21}$$

Equation (21), like Eqs. (14) and (18), implies that as the vector length becomes large, the performance of a VQ can be made to approach the best performance possible by any data compression system. Thus, in theory at least, we have a source coder structure, namely vector quantization, which can achieve the performance promised by the rate distortion and distortion rate bounds (Makhoul, Roucos and Gish; 1985).

Of course, we would also like to have some quantitative indication of the performance gain available with VQ in comparison to scalar quantization. Some early and very important results were obtained by Gish and Pierce (1968) who showed that at high rates (large R), the uniform scalar quantizer is

the optimum entropy constrained quantizer, and that the performance of the optimum entropy constrained quantizer is within 0.255 bits/sample of $R(D)$ for the mean squared error distortion measure independent of the source probability density function. Farvardin and Modestino (1984) have demonstrated that this excellent performance by the optimum entropy constrained quantizer is also maintained at low rates for the uniform, Gaussian, Laplacian, and gamma probability densities. In particular, they have found that the optimum entropy constrained quantizer performs within 0.3 bits/sample of $R(D)$ in all cases (that they considered) and for all distortions.

Possible performance gains of 0.3 bits/sample or less seem relatively small, and one may feel that such modest increases in performance are not worth the additional effort and complexity required by VQ. However, VQ may offer subjective performance improvements in many applications not evident in these objective performance measures, and furthermore, entropy-coded scalar quantization has some disadvantages of its own. Entropy coding of the scalar quantizer output always involves delay and usually consists of fixed-length to variable-length coding. Variable-length codes require buffering at both the transmitter and receiver, and they can be extremely susceptible to loss of synchronization in the presence of channel errors. Therefore, it is often desirable to avoid entropy coding, and, as a result, it may be more meaningful to compare the performance of scalar quantizers whose 2^R output levels, R an integer, are encoded by direct assignment of R bit binary words. In such cases, the performance of scalar quantizers at rates of 1 to 3 bits/sample are at least 0.6 bits/sample greater than $R(D)$ for a Gaussian source and 1 bit/sample greater than $R(D)$ for a Laplacian source (Farvardin and Modestino, 1984; Max, 1960; Adams and Giesler, 1978). Certainly, these possible performance gains are significant. It should also be noted that VQ with large dimension or blocklength offers the possibility of avoiding entropy coding altogether. This result follows from Shannon's work (1959) which showed that for a fixed number of output levels, L, the output entropy of the VQ approaches $\log_2 L$ when N is large. Sakrison (1968) gave a geometrical demonstration of this fact for Gaussian sources by proving that for large dimension (N), the optimum quantization vectors fall with high probability on the surface of an N-dimensional sphere and that the output points are uniformly distributed on the sphere's surface. Because of these results, the minimum output rate can be achieved by a straightforward assignment of appropriate-length binary words.

We thus see that vector quantization can theoretically achieve the optimum performance promised by the rate distortion bound and that this performance is possible without entropy coding. Furthermore, the performance increment available with VQ in comparison to scalar quantization is at least 0.3 bits/sample, and possibly even greater if entropy coding is not used on the scalar quantizer output.

IV. Motivation for Lattice Quantization

Now that we have defined vector quantization and have established that optimal vector quantizers can provide significant performance advantages, we are ready to see how optimal vector quantizers can be found. As noted previously, an N-dimensional VQ is completely determined by specifying the partitions $\mathcal{P}_i$ and the output vectors $\mathbf{Y}_i$, $i = 1, 2, \ldots, L$, such that the $\mathcal{P}_i$ are nonoverlapping and completely cover $\mathcal{R}^N$. Two necessary conditions that must be satisfied for an optimal N-dimensional quantizer are that:

(i) the partition of $\mathcal{R}^N$ must be a Dirichlet partition (also called a Voronoi region), that is,

$$\mathcal{P}_i = \{\mathbf{X}: \|\mathbf{X} - \mathbf{Y}_i\| \leq \|\mathbf{X} - \mathbf{Y}_j\| \text{ for each } j \neq i\} \tag{22}$$

and

(ii) the output points must be centroids of their respective regions, so

$$\mathbf{Y}_i = \{\mathbf{Y}: \int_{\mathcal{P}_i} \|\mathbf{X} - \mathbf{Y}\|^2 f(\mathbf{X})\, d\mathbf{X} \text{ is minimum}\}. \tag{23}$$

One important approach for VQ design is based upon using training sequences representative of the vectors to be quantized and the LBG algorithm, which is a version of the K-means algorithm in the pattern recognition literature (Linde, Buzo and Gray, 1980; Makhoul, Roucos and Gish, 1985; MacQueen, 1967). This iterative algorithm can be shown to converge to at least a local optimum, and global optimality can be approximated by repeatedly running the algorithm with different initialization vectors. This algorithm (in general) produces a nonuniform partition of $\mathcal{R}^N$ and a nonuniform distribution of VQ output points. This design procedure is performed totally off-line, but the computational and storage requirements of this process are still not insignificant. In fact, for M training vectors and I iterations of the algorithm, the computational cost is about $NLMI = NMI\, 2^{RN}$ operations and the storage cost is $N(L + M)$. Since M must be at least $10L$, these quantities are large and grow exponentially with an increase in R and N (Makhoul, Roucos and Gish, 1985).

Once we have obtained a VQ codebook, the quantization process consists of calculating the distortion between the current input vector and each output vector and choosing that output vector with minimum distortion. If each distortion calculation requires N operations, then the quantization process requires $NL = N\, 2^{RN}$ operations. The storage cost is also $N\, 2^{RN}$. Since these operations must be accomplished in real time, this computational burden is quite significant. For example, if $R = 2$ bits/sample and $N = 10$, the number of operations is $10 \cdot 2^{20} \cong 10$ million! These computational and storage

requirements are for a full-search VQ, and much research effort is going into reducing these numbers with some loss in performance.

Vector quantizer codebooks designed using the LBG algorithm generally have no discernible structure, and this "random" codebook distribution is what complicates the quantization or encoding process. Lattices in $\mathscr{R}^N$ have considerable structure, and hence, lattice-based quantizers offer the promise of design simplicity and reduced complexity encoding, providing that lattices can be found in high dimensions which yield good quantization performance. A lattice is defined as a set of vectors

$$\Lambda = \{\mathbf{x}:\mathbf{x} = u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + \cdots + u_N\mathbf{a}_N\} \tag{24}$$

where $\mathbf{a}_i$, $i = 1,2,\ldots,N$, are the basis vectors of the lattice and the u_i are integers. We form a VQ from a lattice by selecting L of the lattice points $\mathbf{x}$ to be the output points $\mathbf{Y}_i$ and forming Voronoi regions about these output points so that if a source vector $\mathbf{X} \in \mathscr{V}_i$, then $\mathbf{Y}_i = \mathbf{Q}(\mathbf{X})$.

Now that we have a defined lattice quantizer, how do we find a good lattice quantizer? First, Gersho (1979) has conjectured that for N asymptotically large, the optimal quantizer for a uniformly distributed source will have all of its Voronoi regions (except the boundary regions) congruent to some basic polytope. Second, a quantizer will perform well if its Voronoi region approaches the shape of a sphere. This statement comes from the following argument. It is generally felt that for asymptotic N the best covering of space is a dense packing of nonoverlapping spheres. Since a nonoverlapping covering is not possible for finite N, the best covering will be a covering by spheres with minimal overlap. The nonoverlapping regions are the Voronoi regions. Thus, to find a good lattice quantizer, we are interested in regular lattices and Voronoi regions which best approximate a sphere in $\mathscr{R}^N$. Note that we can also use the second argument to justify approaching the problem as a search for dense sphere packings in $\mathscr{R}^N$ with the lattice points as the sphere centers, called a lattice packing (Sloane, 1984).

Details concerning the design of lattice vector quantizers, or simply lattice quantizers, are pursued in subsequent sections. Here we examine the performance offered by lattice quantizers. That is, since we are restricting the possible output points to lie on a lattice, which is a regular structure, and we know that locally optimal vector quantizers designed with the LBG algorithm can be quite irregular, exactly what is the performance penalty for limiting consideration to lattice quantizers? We begin to answer this question by presenting a version of a derivation due to Sakrison (1979).

We consider an N-dimensional VQ as previously defined in Sec. III with an output rate in bits sample of

$$\frac{1}{N}H(\mathbf{Y}) = -\frac{1}{N}\sum_{i=1}^{L} P[\mathbf{Y} = \mathbf{Y}_i]\log_2 P[\mathbf{Y} = \mathbf{Y}_i]. \tag{25}$$

Since

$$P[\mathbf{Y} = \mathbf{Y}_i] = \int_{\mathcal{A}_i} f(\mathbf{X})\, d\mathbf{X}, \tag{26}$$

which is an N-dimensional integral over the input vector probability density, we can rewrite Eq. (25) as

$$\frac{1}{N} H(\mathbf{Y}) = -\frac{1}{N} \sum_{i=1}^{L} \int_{\mathcal{A}_i} f(\mathbf{X}) \log P[\mathbf{Y} = \mathbf{Y}_i]\, d\mathbf{X}. \tag{27}$$

Under the assumptions of small distortion and a sufficiently smooth input density, the argument of the logarithm in Eq. (27) can be approximated by

$$P[\mathbf{Y} = \mathbf{Y}_i] = \int_{\mathcal{A}_i} f(\mathbf{X})\, d\mathbf{X} \cong V_i f(\mathbf{X}) \tag{28}$$

where V_i is the volume of the *ith* partition, so that Eq. (27) becomes

$$\begin{aligned} \frac{1}{N} H(\mathbf{Y}) &\cong -\frac{1}{N} \sum_{i=1}^{L} \int_{\mathcal{A}_i} f(\mathbf{X}) \log [V_i f(\mathbf{X})]\, d\mathbf{X} \\ &= \frac{1}{N} H(\mathbf{X}) - \frac{1}{N} \sum_{i=1}^{L} \int_{\mathcal{A}_i} f(\mathbf{X}) \log V_i\, d\mathbf{X} \\ &= \frac{1}{N} H(\mathbf{X}) - \frac{1}{N} \sum_{i=1}^{L} P[\mathbf{Y} = \mathbf{Y}_i] \log V_i. \end{aligned} \tag{29}$$

To simplify further we also assume that all of the partitions $\mathcal{A}_i$ are translated versions of the same shape, say $\mathcal{A}$, and that each partition contributes an average distortion equal to D. Then $V_i = V$ for all i, and

$$\frac{1}{N} H(\mathbf{Y}) \cong \frac{1}{N} H(\mathbf{X}) - \frac{1}{N} \log V \tag{30}$$

is the rate in bits/sample required by the vector quantizer to achieve an average distortion of D or less.

What we would like to do now is show that the N-dimensional VQ rate-distortion performance just derived is near the rate distortion bound for the input source sequence x_j, $j = 1, 2, \ldots$. Letting y_i, $i = 1, 2, \ldots$, denote the representation sequence produced by the source coder, then we consider the single letter fidelity criterion

$$d(x - y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2 \tag{31}$$

which allows us to write the average distortion

$$E\{d(x-y)\} = \iint f_x(x) f_{y|x}(y|x) d(x-y)\, dx\, dy. \tag{32}$$

The average mutual information between the source coder input x and the reproduction y is

$$I(x;y) = \iint f_x(x) f_{y|x}(y|x) \log \frac{f_{y|x}(y|x)}{f_y(y)}\, dx\, dy \tag{33}$$

where

$$f_y(y) = \int f_x(x) f_{y|x}(y|x)\, dx, \tag{34}$$

and to find $R(D)$ we wish to minimize $I(x;y)$ over all transition probability densities $f_{y|x}(y|x)$ that yield an average distortion equal to D. Thus, defining the admissible set

$$F_D = \{f_{y|x}(y|x): E[d(x-y)] \le D\}, \tag{35}$$

we can write the rate distortion function for the given source and the chosen fidelity criterion as

$$R(D) = \inf_{f_{y|x} \in F_D} I(x;y). \tag{36}$$

By using the facts that $I(x;y) = H(x) - H(x|y)$, $H(x-y|y) = H(x|y)$, and $H(x-y) \ge H(x-y|y)$, we can manipulate Eq. (36) as follows,

$$\begin{aligned} R(D) &= \inf_{f_{y|x} \in F_D} [H(x) - H(x|y)] \\ &= H(x) - \sup_{f_{y|x} \in F_D} H(x-y|y) \\ &\ge H(x) - \sup_{f_{y|x} \in F_D} H(x-y) \triangleq R_S(D), \end{aligned} \tag{37}$$

where the second equality results since $H(x)$ does not depend on $f_{y|x}(y|x)$. The final result in Eq. (37) is called the Shannon lower bound to the rate distortion function (Shannon, 1959; Berger, 1971), and our particular derivation is due to Sakrison (1979). The Shannon lower bound, denoted here by $R_S(D)$, is particularly useful since Eq. (36) defining $R(D)$ is difficult to evaluate for general sources and distortion measures, but $R_S(D)$ can be calculated in many cases of interest.

To see how close the performance of a uniform VQ comes to $R(D)$, we compare Eq. (30) and $R_S(D)$ in Eq. (37). Note that if we assume that the

sequence of source inputs is independent and identically distributed, then $H(\mathbf{X}) = NH(x)$, so that the first terms in Eq. (30) and $R_S(D)$ are identical. Although it is beyond our development here, it is possible to show that for N large, $H(\mathbf{X} - \mathbf{Y})$ equals $\log V$ with high probability (Shannon, 1948; Sakrison, 1968; Sakrison, 1979), and since $H(x - y) = (1/N)H(\mathbf{X} - \mathbf{Y})$, then the last terms in Eq. (30) and in $R_S(D)$ are equal for large dimensions.

Therefore, we conclude that the performance of a uniform VQ achieves the Shannon lower bound on $R(D)$ as the number of dimensions becomes asymptotically large. This result is encouraging since it says that VQs with identical quantization cells, which are obviously highly structured, are asymptotically optimum. To evaluate the performance for small N, we need only compare $\sup_{f_{y|x} \in F_D} H(x - y)$ and $(1/N)\log V$, where V is the N-dimensional volume of the basic quantizer partition $\mathcal{A}$. Before V can be calculated, however, we must scale $\mathcal{A}$ such that the desired average distortion, say D^*, is achieved. To perform this scaling, we first note from Eq. (28) that the probability of any N-vector $\mathbf{X}$ is uniform throughout the region $\mathcal{A}$, and that if the representation vector is at the center of $\mathcal{A}$, then it follows that the quantization errors are uniformly distributed throughout $\mathcal{A}$. Based upon these observations, we can show that for an average distortion D^*, the uniform quantization region $\mathcal{A}$ in one dimension is the closed interval $[-\sqrt{3D^*}, \sqrt{3D^*}]$, and thus,

$$\frac{1}{N}\log V = \frac{1}{2}\log 12D^* = 1.792 + \frac{1}{2}\log D^* \text{ bits/sample.} \tag{38}$$

Similarly, if in two dimensions we choose $\mathcal{A}$ to be a hexagon, then for an average distortion D^*, the radius of the hexagon is $[24D^*/5]^{1/2}$ and

$$\frac{1}{N}\log V = \frac{1}{2}\log[72\sqrt{3}D^*/10]$$

$$= 1.82 + \frac{1}{2}\log D^* \text{ bits/sample.} \tag{39}$$

(Note that Eq. (30) in Sakrison (1979) appears to be in error). The values in Eqs. (38) and (39) should be compared to $\sup_{f_{y|x} \in F_D} H(x - y)$ for the squared error distortion measure, which is (Berger, 1971; Sakrison, 1979)

$$\sup_{f_{y|x} \in F_D} H(x - y) = \frac{1}{2}\log 2\pi e D^*$$

$$= 2.047 + \frac{1}{2}\log D^* \text{ bits/sample.} \tag{40}$$

Subtracting Eq. (38) from (40), we see that optimal VQ offers a reduction of

0.255 bits/sample over scalar quantization, but comparing Eqs. (39) and (40), the two-dimensional uniform hexagonal quantizer performance is only 0.028 bits/sample better than the scalar quantizer.

These last results have both an encouraging and a discouraging aspect. On one hand, it is encouraging that uniform VQ performance can approach $R(D)$ for large dimensions. On the other hand, however, it is discouraging to find that multidimensional VQ only performs 0.255 bits/sample better than uniform scalar quantization, and that a two-dimensional quantizer provides only a small portion of this available improvement. This last point implies that we will not be able to close the gap between vector quantizer performance and $R(D)$ without going to higher dimensions (N). In fact, Sakrison (1968, 1979) felt that because of this seemingly negligible gain in performance and the complexity involved in the implementation of N-dimensional VQs, that vector quantization "...may never be used in practice". However, as stated at the end of the immediately preceding section, VQ may provide subjective improvements not evident in our mathematical development, and it is possible that we can avoid the use of entropy coding with vector quantization.

In this section, it has been demonstrated that uniform vector quantizers can achieve performance at or near the rate distortion bound, and hence, we conclude that highly structured VQs based upon lattices, which offer the possibility of significant reductions in implementation complexity, are viable alternatives to optimal and locally optimal VQs designed using the LBG algorithm. We are now ready to begin our investigation of lattice quantizers.

V. Lattices

Given a set A of n linearly independent vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$, a lattice in n-dimensional Euclidean space is defined by taking integral linear combinations of the vectors in A as shown in Eq. (24). The set A is called the basis of the lattice Λ, and as is evident from Eq. (24), there is considerable structure associated with a lattice. In fact, most of this section and the two subsequent sections are concerned with presenting and exploiting lattice structural properties. However, it is also clear from Eq. (24) that it is possible to define many different lattices, and we wonder exactly how to proceed to find those lattices which can serve as good quantizers. One approach is to examine lattices which have already been widely studied in other contexts or applications, such as sphere packing, and investigate the performance of quantizers built upon these lattices.

Such has been the approach employed in the lattice quantization literature, and it has led to the study of lattices that are based upon the root systems of Lie algebras and which have the designations $A_n (n \geq 1)$, $B_n (n \geq 1)$,

TABLE V

A GENERATOR MATRIX FOR THE LEECH LATTICE, $\Lambda_{24} = \Lambda_{24}^{*}$ (SLOANE, 1981)[7]

$$\frac{1}{\sqrt{2}}\begin{bmatrix}
4 & & & & & & & & & & & & & & & & & & & & & & & \\
2 & 2 & & & & & & & & & & & & & & & & & & & & & & \\
2 & & 2 & & & & & & & & & & & & & & & & & & & & & \\
2 & & & 2 & & & & & & & & & & & & & & & & & & & & \\
2 & & & & 2 & & & & & & & & & & & & & & & & & & & \\
2 & & & & & 2 & & & & & & & & & & & & & & & & & & \\
2 & & & & & & 2 & & & & & & & & & & & & & & & & & \\
2 & & & & & & & 2 & & & & & & & & & & & & & & & & \\
2 & & & & 0 & & & & 2 & & & & & & & & & & & & & & & \\
2 & & & & & & & & & 2 & & & & & & & & & & & & & & \\
2 & & & & & & & & & & 2 & & & & & & & & & & & & & \\
2 & & & & & & & & & & & 2 & & & & & & & & & & & & \\
1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & & & & & & & & & & & \\
1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & & 1 & & & & & & & & & & \\
1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & & & 1 & & & & & & & & & \\
1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & & & & 1 & & & & & & & & \\
1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & & & & & 1 & & & & & & & \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & & & & 0 & & 1 & & & & & & \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & & & & & & & 1 & & & & & \\
1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & & & & & & & & 1 & & & & \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & & & & & & & & & 1 & & & \\
1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & & & & & & & & & & 1 & & \\
1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & & & & & & & & & & & 1 & \\
-1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}$$

VI. LATTICE QUANTIZER DESIGN

Recall from previous sections that an N-dimensional VQ is completely determined by specifying the partitions $\mathcal{P}_i$ and the output vectors $\mathbf{Y}_i$, $i = 1, 2, \ldots, L$, such that the $\mathcal{P}_i$ are nonoverlapping and completely cover $\mathcal{R}^N$. Two necessary conditions that must be satisfied for an optimal VQ are given in Eqs. (22) and (23) of Section IV. Given one of the lattices in Section V, we design a lattice VQ by letting the lattice points be the output points $\mathbf{Y}_i$, and we construct Voronoi regions or Dirichlet partitions about these lattice points according to Eq. (22). Note that under the assumption that the N-dimensional input pdf $f(\mathbf{X})$ is uniform, then once we construct the Voronoi region around a lattice point, the output points (lattice points) are guaranteed to be the centroids of their respective regions (see Eq. (23)).

Consider the following examples. For the Z^1 lattice the Voronoi regions are simply intervals along the real line of size $\pm\frac{1}{2}$ about each lattice point. Clearly, this partition covers $\mathcal{R}^1$ and the output points are centroids of their

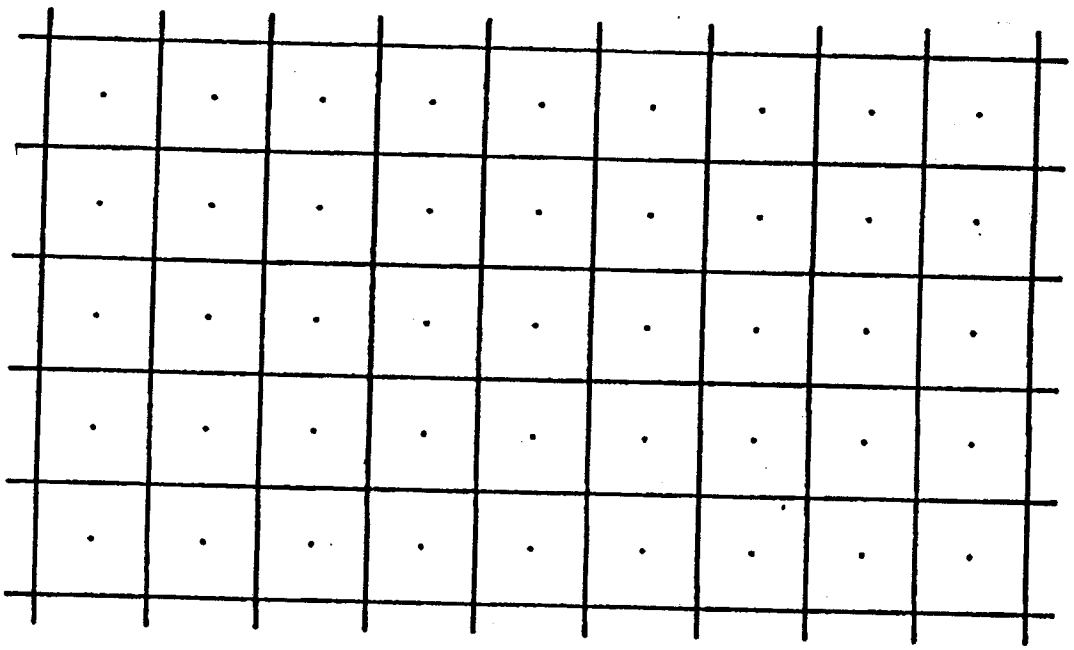

FIG. 12. Voronoi regions and output points for the Z^2 lattice.

respective regions. The Z^2 lattice has Voronoi regions that are squares with the lattice/output points at the center of each square as shown in Fig. 12. Similarly, the output points and Voronoi regions for the A_2 lattice are shown in Fig. 13. The hexagonal partitions in Fig. 13 can also be obtained by considering A_2 as a sphere packing in $\mathscr{R}^2$. In fact, A_2 is the laminated lattice packing in $\mathscr{R}^2$ as mentioned in Section V. The hexagonal partitions can be generated from the laminated lattice by connecting the deep holes. Alternatively, the spheres in the laminated lattice packing can be expanded in

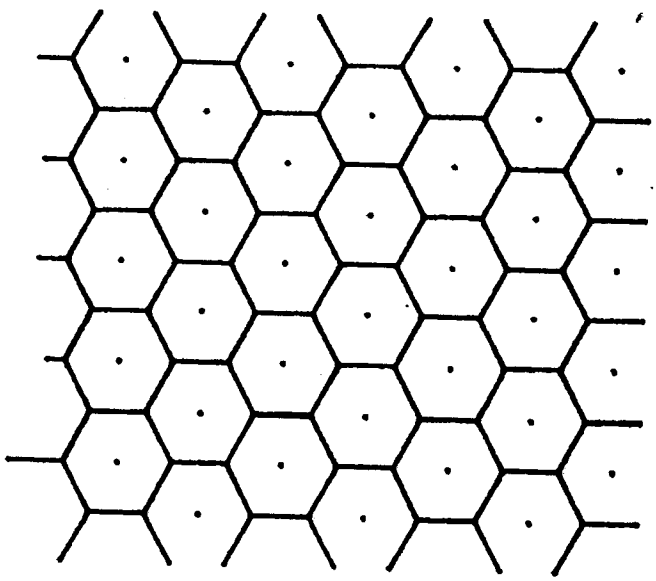

FIG. 13. Voronoi regions and output points for the A_2 lattice.

radius until all of $\mathscr{R}^2$ is just covered, which, incidentally, occurs just when the spheres pass through the deep holes. The expanded spheres overlap, but if we draw chords through each overlapping region, we get the hexagonal covering in Fig. 13. Figure 14 demonstrates the last two interpretations based upon the laminated lattice packing (Sloane, 1984).

In three dimensions, some interesting Voronoi regions are the Voronoi region for the $A_3^* \cong D_3^*$ lattice, called a truncated octahedron, shown in Fig. 15, and the Voronoi region for the $A_3 \cong D_3$ lattice, called a rhombic dodecahedron, shown in Fig. 16. The Voronoi regions are shown inscribed in a cube so that the locations of the surrounding lattice (output) points, shown as dark dots, can be easily discerned. There is an output point at the center of each Voronoi region as indicated, and in each figure there is one lattice (output) point behind the Voronoi region that is suppressed (not seen) to avoid confusion. The lattice points for the $A_3^* \cong D_3^*$ quantizer in Fig. 15 are at the vertices of the cube and at the center of the cube, while in Fig. 16, the output points are at the midpoints of the edges of the cube and at the center of the cube (Gersho, 1979; Conway and Sloane, 1982a and 1984; Barnes and Sloane, 1983; Joshi, 1977). The A_3^* lattice is sometimes called the *body centered cubic lattice*, and the A_3 lattice is sometimes called the *face centered cubic lattice* (Conway and Sloane, 1982a and 1984; Barnes and Sloane, 1983). The origin of these names is not transparent, but the terms come from the following interpretation (Coxeter, 1961). Consider a simple lattice in $\mathscr{R}^3$ whose points have only even coordinates. In fact, it is instructive to consider the cube with vertices at (0, 0, 0), (2, 0, 0), (0, 2, 0), (0, 0, 2), (2, 2, 0), (2, 0, 2), (0, 2, 2), and (2, 2, 2). The center of a face in this cube has two odd coordinates and these points along with the point at the origin are equivalent to the $A_3 \cong D_3$ lattice, thus the name "face centered cubic lattice". Further, the center of the cube or "body"

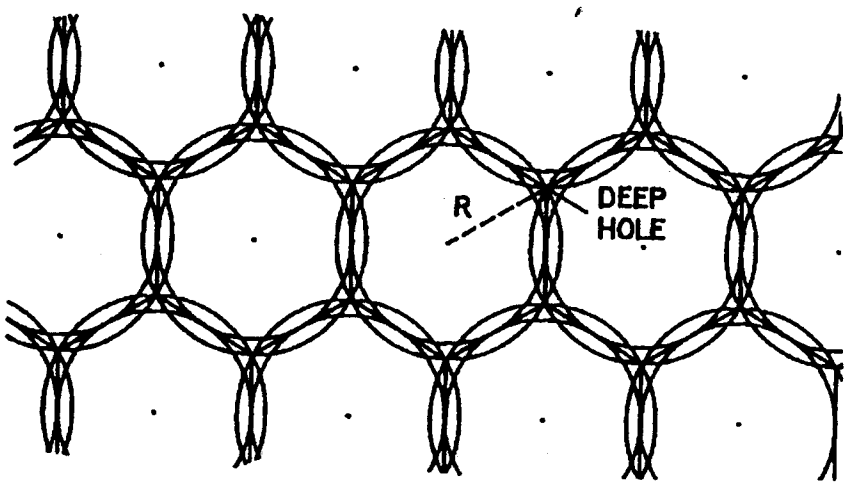

FIG. 14. Generating the hexagonal covering of $\mathscr{R}^2$ from the laminated lattice sphere packing (Sloane, 1984).[8]

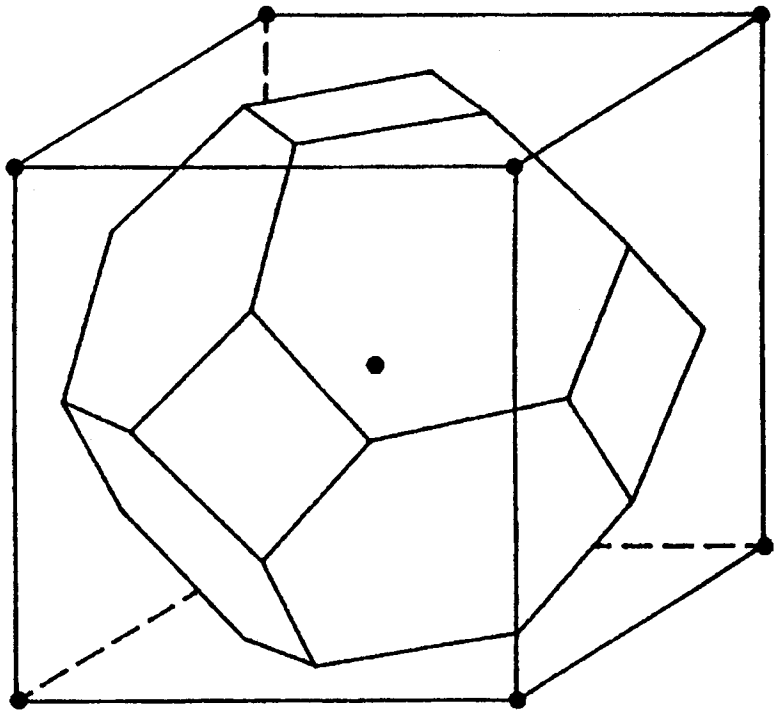

FIG. 15. Voronoi region and surrounding output points for the $A_3^* \cong D_3^*$ lattice quantizer.

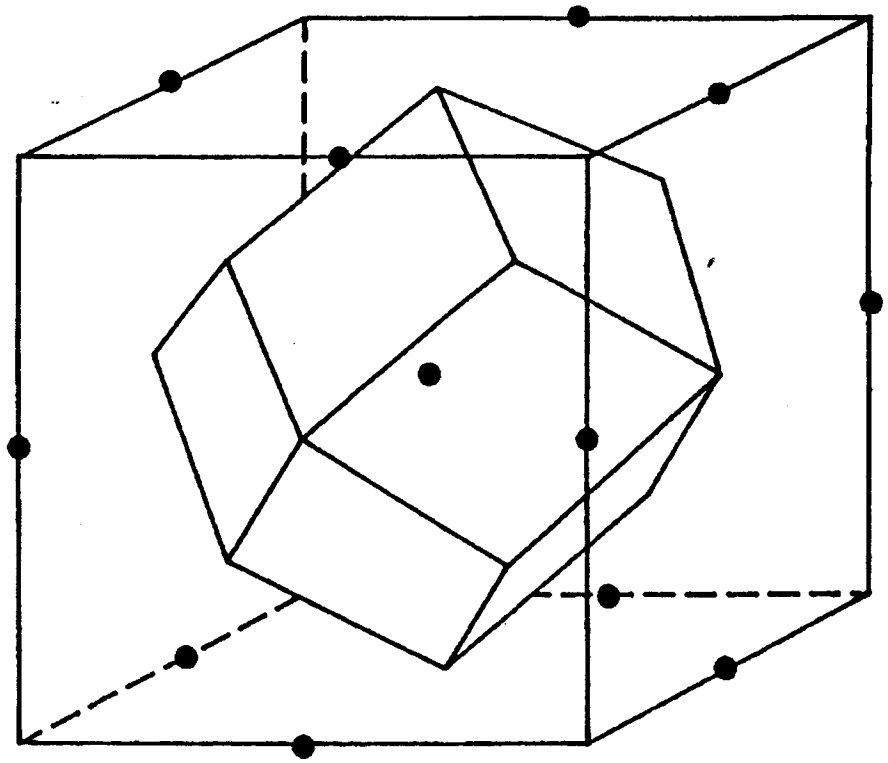

FIG. 16. Voronoi region and surrounding output points for the $A_3 \cong D_3$ lattice quantizer.

has three odd coordinates and these points (using all other cubes nearest the origin) along with the origin are equivalent to the $A_3^* \cong D_3^*$ lattice, motivating the nomenclature "body centered cubic lattice." It is also of interest to note that the term *Brillouin zone* is sometimes used. The Brillouin zone of a lattice Λ is the Voronoi region of the dual or reciprocal lattice Λ^*. Thus, the Brillouin zone of A_3 is a truncated octahedron, which is the Voronoi region of the A_3^* lattice; and similarly, the Brillouin zone of the A_3^* lattice is the rhombic dodecahedron, which is the Voronoi region of the A_3 lattice. We finally point out that the Voronoi region of a lattice, which is a nearest neighbor or Dirichlet partition, may also be called a Wigner-Seitz cell for the same lattice (Coxeter, 1961).

Another way to construct uniform quantizers that is consistent with the finite reflection group technique for generating the root lattices and that has considerable intuitive appeal is to cover space with copies of an admissible polytope. This approach, due to Gersho (1979), is particularly insightful in $\mathscr{R}^2$ and $\mathscr{R}^3$. A convex polytope P is said to tile or cover space or to generate a tessellation of $\mathscr{R}^N$ if a partition of $\mathscr{R}^N$ exists where all regions are congruent to the basic polytope P. Furthermore, Gersho defines the class of admissible polytopes as those polytopes which tessellate (or tile or cover) $\mathscr{R}^N$ and which constitute a Dirichlet partition with respect to the centroids of each region. Thus, admissible polytopes in $\mathscr{R}^2$ are the equilateral triangle, the rectangle, and the regular hexagon. In three dimensions, the space-filling polytopes are the cube (Fig. 17), the hexagonal prism (Fig. 18), the rhombic dodecahedron (Fig. 19), the elongated dodecahedron (Fig. 20), and the truncated octahedron (Fig. 21) (Gersho, 1979; Lyusternik, 1963). Of course, the cube is the Voronoi region for the Z^3 lattice, the rhombic dodecahedron is the Voronoi region for the $A_3 \cong D_3$ lattice, and the truncated octahedron is the Voronoi region for the $A_3^* \cong D_3^*$ lattice.

It is more difficult to conceive admissible polytopes in $\mathscr{R}^N$ for $N > 3$, but Gersho points out that one method is to form cross-products of lower dimensional polytopes. For example, Z^N is the Nth cross product of the

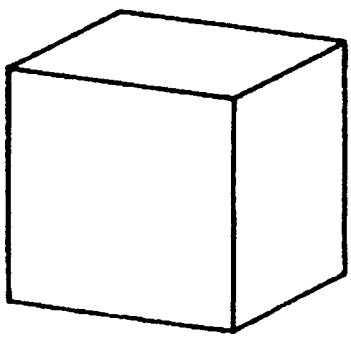

FIG. 17. The cube in $\mathscr{R}^3$.

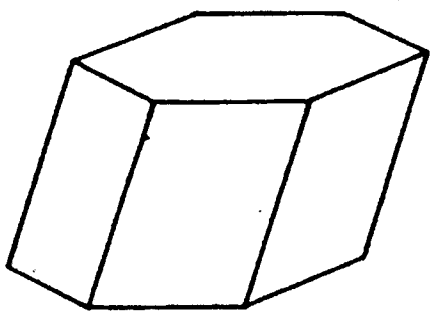

FIG. 18. The hexagonal prism.

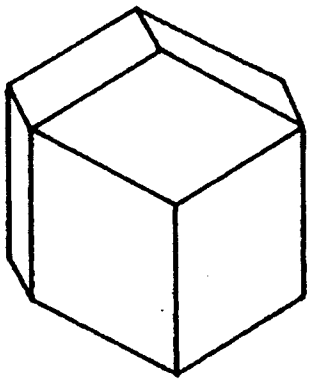

FIG. 19. The rhombic dodecahedron.

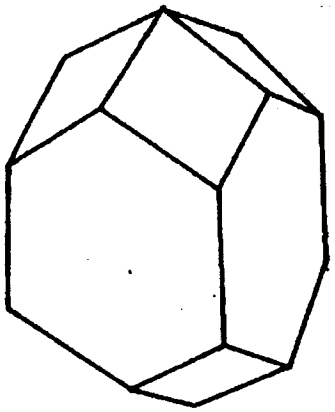

FIG. 20. The elongated dodecahedron.

interval, and the hexagonal prism is the cross product of the regular hexagon in $\mathscr{R}^2$ and the interval. A polytope in $\mathscr{R}^5$ can be generated by the cross product of the regular hexagon in $\mathscr{R}^2$ and the truncated octahedron in $\mathscr{R}^3$. From Section V recall that lattices built from direct products of lower dimensional lattices are called reducible and have a dimension equal to the sum of the dimensions of the lower dimensional systems. A lattice that cannot be written as a direct product of lower dimensional systems is called irreducible. Carrying this nomenclature over to the admissible polytopes, we could say that the cube is reducible but that the truncated octahedron is irreducible.

The descriptions of the Voronoi regions of the root lattices and their duals and the K_{12}, Λ_{16}, and Λ_{24} lattices are far from trivial. In fact, the Voronoi regions for E_6^*, E_7^*, K_{12}, Λ_{16}, and Λ_{24} are not known, but Conway and Sloane (1982a) have determined the Voronoi regions for the lattices $A_n (n \geq 1)$,

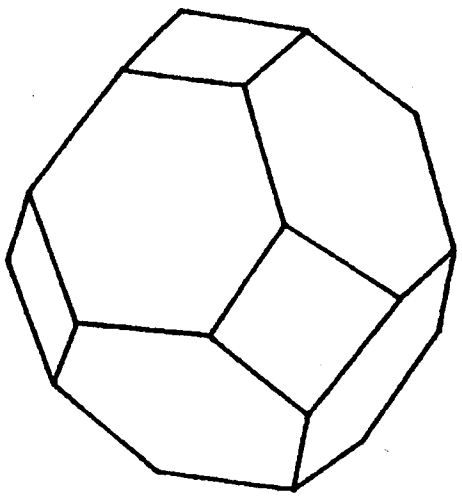

FIG. 21. The truncated octahedron.

$^*_n(n \geq 1)$, $D_n(n \geq 3)$, $D^*_n(n \geq 3)$, E_6, E_7, and $E_8 = E^*_8$. The method they used quires a description of the fundamental simplex for the affine Weyl group of e lattice. This development would take us further into group theory and into ie algebras, and hence, for simplicity, we do not pursue the description of ɔronoi regions for lattices further here. Fortunately, in order to implement a ttice quantizer we do not need to know its Voronoi region; we need only be ɔle to determine which lattice point a given vector is nearest to according to]. (22), which we can do. A knowledge of the Voronoi regions is useful for eoretical performance predictions of lattice quantizers, however, as given by ɔnway and Sloane (1982a).

Therefore, given a particular lattice, we can use it for vector quantization ' simply calculating the nearest lattice point for a particular input vector. owever, we are not quite finished with designing a lattice VQ, since the ttices are generally infinite and we are only interested in L output points, en though L may be large. We must therefore restrict the number of lattice ɔints that are possible output points.

One possible selection rule is to generate a codebook with L vectors that ıs a minimum peak energy, where peak energy is defined as the squared stance of the output point (lattice point or code vector) furthest from the igin. This minimum peak energy rule entails filling the codebook (choosing owable lattice points) with L points from the innermost shells of the lattice, ıere a shell or layer consists of all points that fall a fixed distance from e origin. The number of lattice points in each shell is available from the efficients in the theta function for a given lattice. Sloane (1981) has found the eta functions for the A_n, D_n, D^*_n, E_n, K_{12}, Λ_{16}, and Λ_{24} lattices and has bulated the number of points in the innermost shells for the A_2, D_3, D^*_3, D_4, , E_7, E_8, K_{12}, Λ_{16}, and Λ_{24} lattices. As an example, the A_2 hexagonal lattice th its five innermost shells is shown in Fig. 22.

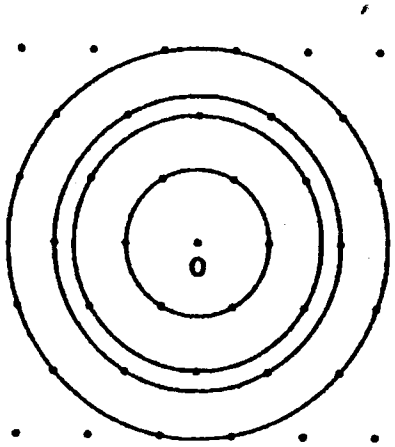

FIG. 22. The five innermost shells for the A_2 lattice containing 1, 6, 6, 6, and 12 points ɔane, 1981).[9]

Another lattice codebook selection rule, due to Conway and Sloane (1983), is based on the Voronoi region of a lattice point. Specifically, for a lattice Λ in $\mathscr{R}^N$, we define the Voronoi region around a lattice point $\mathbf{Y}_i$ as

$$V(\mathbf{Y}_i) = \{\mathbf{x} \in \mathscr{R}^N : \|\mathbf{x} - \mathbf{Y}_i\| \leq \|\mathbf{x} - \mathbf{Y}_j\| \text{ for all } i \neq j\}, \tag{91}$$

and we let $V(0)$ denote the Voronoi region around the origin. We call $V(0)$ the Voronoi region of the lattice. For the positive integers $r = 1, 2, 3, \ldots$, we let V_r denote the Voronoi region for the lattice $r\Lambda$, so that the Voronoi region of Λ is magnified r times. Recall that the volume of the Voronoi region is equal to the determinant of the lattice, so we note that V_r has the volume $\det(r\Lambda) = d(r\Lambda) = r^N d(\Lambda)$. Thus, V_r has r^N times the volume of $V(0)$, and since $V(0)$ contains one lattice point, V_r contains r^N lattice points.

Therefore, we define a *Voronoi code* with $L = r^N$ codewords (or output points) as all vectors $\mathbf{x} - \mathbf{a}$ for $\mathbf{x} \in \Lambda \cap (\mathbf{a} + V_r)$ for some vector $\mathbf{a} \in \mathscr{R}^N$. We denote the Voronoi code by $C_\Lambda(r, \mathbf{a})$ and discard all lattice points not in the code. The vector $\mathbf{a}$ is included to prevent lattice points from falling on the boundary. For some Euclidean code in $\mathscr{R}^N$, $C = \{\mathbf{x}_1, \ldots, \mathbf{x}_L\}$, define the centroid

$$\hat{\mathbf{x}} = \frac{1}{L} \sum_{i=1}^{L} \mathbf{x}_i \tag{92}$$

and average energy

$$\mathscr{E}(C) = \frac{1}{d_m^2 L} \sum_{i=1}^{L} \|\mathbf{x}_i - \hat{\mathbf{x}}\|^2, \tag{93}$$

where d_m is the minimum distance between codewords (lattice points). It is desirable to choose $\mathbf{a}$ so that C has the smallest average energy, which since the Voronoi codes usually have their centroid at $\mathbf{a}$, gives $\hat{\mathbf{x}} = \mathbf{a}$.

Figure 23 illustrates the construction of the Voronoi code $C_{A_2}(4, \mathbf{a})$ with $\mathbf{a} = (-\frac{1}{4}, 0)$. The Voronoi region V_4 for the lattice $4A_2$ is shown by the hexagonal dashed line. The Voronoi code consists of all lattice points falling within the solid hexagonal line, which are the $L = 4^2 = 16$ points with circles around them. Conway and Sloane (1983) present more details on this example and construct Voronoi codes based on the D_4 and E_8 lattices.

These two approaches are rather ad hoc and it seems that better quantizers could be obtained by a more refined selection rule. Unfortunately, our choice of selection rules is limited by the requirement that we maintain the regular structure of the lattice. In the next section we show how this regular structure is used to obtain fast encoding algorithms for lattice quantizers.

vectors of the [24, 12, 8] Golay code (MacWilliams and Sloane, 1977). Therefore, for this lattice, 8192 direct distance calculations must be performed for each input vector which yields a relatively slow quantizing algorithm (Conway and Sloane, 1984).

VIII. Performance Comparisons

Performance evaluations for lattice quantizers are based upon the conjectures mentioned in Section IV which imply that the Voronoi regions of good lattice quantizers are those which best approximate a sphere in $\mathcal{R}^N$. Furthermore, since the best covering of $\mathcal{R}^N$ is a dense packing of nonoverlapping spheres, we may find good lattice quantizers by looking for dense sphere packings in $\mathcal{R}^N$ where the sphere centers are the lattice points. The sphere packing interpretation is very useful for gaining insight into the problem by examining spaces with dimension $N \leq 3$. In one dimension the densest lattice packing is called Z^1 with the lattice points corresponding to the integers. As shown in Fig. 25(a), the "spheres" are line segments of unit length, and the entire space is covered by nonoverlapping spheres. A lattice packing in two dimensions is Z^2, as shown in Fig. 25(b), which has spheres centered at every point in the plane with integer coordinates. The nonoverlapping spheres clearly do not cover $\mathcal{R}^2$. Another two-dimensional lattice packing is the hexagonal or triangular lattice packing, denoted by L_2 and A_2 and illustrated in Fig. 25(c). This packing is constructed by forming one layer of spheres with centers at the integers along the horizontal axis and then adding a layer of spheres that fits in the "slots" of the first layer. The third layer, like the first layer, has sphere centers that are integers in the x-coordinate, and the process is continued. The nonoverlapping spheres in L_2 also do not cover $\mathcal{R}^2$. Which is the denser packing, Z^2 or L_2? The density of a lattice packing is that fraction of the space covered by spheres, and can be calculated by dividing the volume of a sphere by the volume of space nearer to its center than any other center. Thus, for Z^1 the density is 1, for Z^2 the density is $\pi/4 \cong .7954$, and for L_2 the density is $\pi\sqrt{3}/6 \cong .9069$. The denser sphere packing is therefore L_2.

A dense sphere packing is not guaranteed to yield a good quantizer, and hence it is necessary to calculate the distortion associated with each lattice when used as a quantizer. A lattice quantizer can be constructed from the lattice packings Z^2 and L_2 by forming Voronoi, or nearest neighbor, regions about each lattice point (sphere center), which is the output point for the particular region of interest. The Voronoi regions are squares in Fig. 25(b) and hexagons in Fig. 25(c). To find the MSE per dimension, we simply find the average squared error between the output point and all other points in the region. For the Z^1, Z^2, and L_2 quantizers, the MSE per dimension

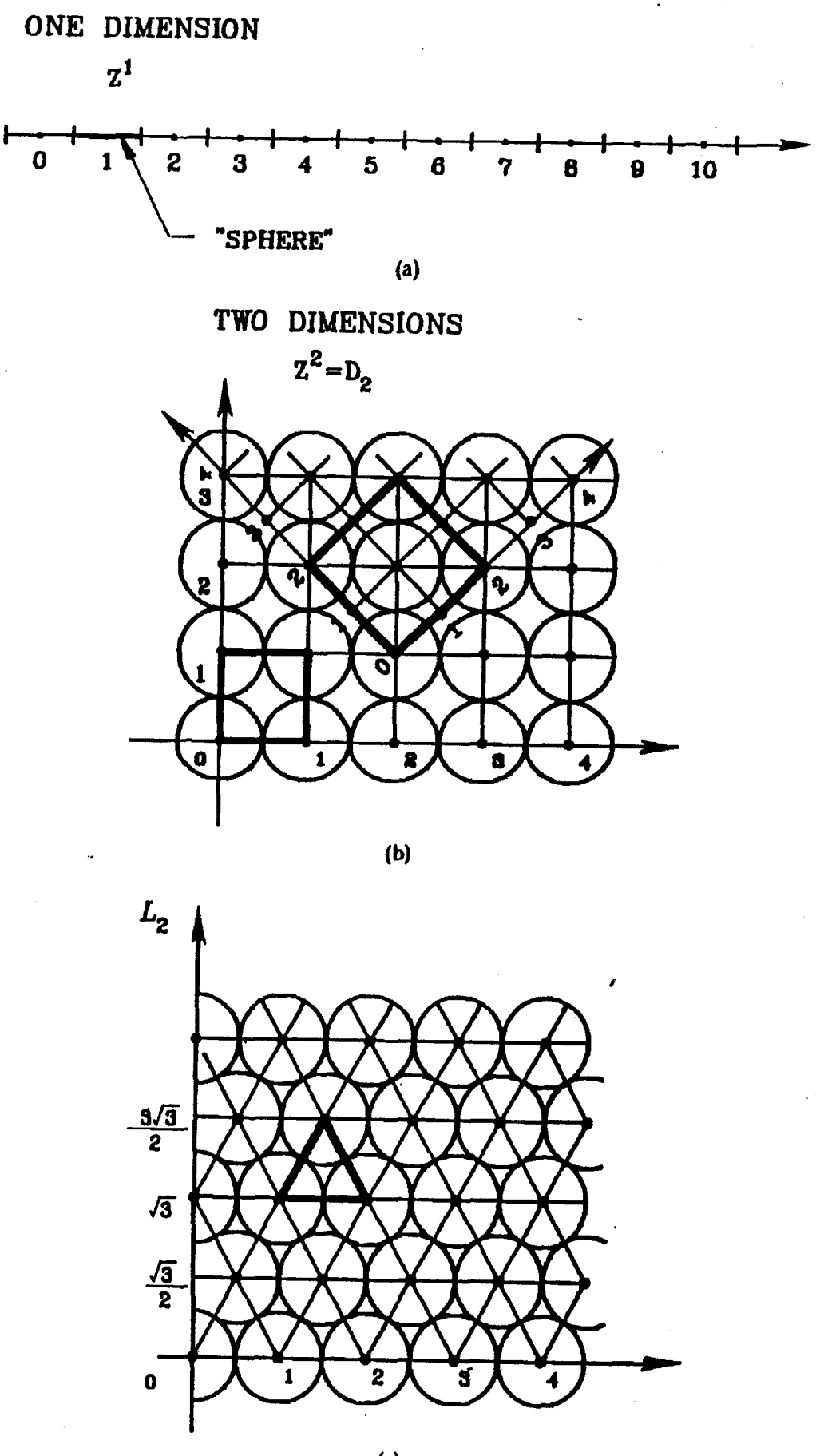

FIG. 25. Sphere packings in one and two dimensions (Sloane, 1984).[11]

can be readily evaluated as $\frac{1}{12} = 0.08333\ldots$, $\frac{1}{12} = 0.08333\ldots$, and $5/36\sqrt{3} = 0.0801875\ldots$, respectively, assuming a uniform input distribution (Makhoul, Roucos and Gish, 1985; Conway and Sloane, 1982a). Although it is simple to calculate the MSE for quantizers in one and two dimensions, the calculation becomes increasingly difficult in higher dimensions. The structure of lattices provides assistance in these cases.

The basis vectors of a lattice Λ can be selected in many different ways, and so there is tremendous flexibility in specifying a lattice quantizer. The generator matrix for Λ is defined as the $N \times N$ matrix

$$M = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_N \end{bmatrix} \tag{135}$$

and the determinant of Λ is

$$\begin{aligned} \det \Lambda &= (\det MM^T)^{1/2} \\ &= |\det M|. \end{aligned} \tag{136}$$

The determinant of a lattice gives an indication of the amount of space represented by a single point of the lattice, so the determinant of a lattice is the volume of that lattice's Voronoi region (Conway and Sloane, 1982a). Furthermore, the density of a lattice (sphere) packing of radius ρ is

$$\Delta = \frac{V_N \rho^N}{\det \Lambda} \tag{137}$$

where

$$V_N = \frac{\pi^{N/2}}{\Gamma((N/2) + 1)} \tag{138}$$

is the volume of the unit sphere in $\mathscr{R}^N$.

Some of the most important lattices for VQ design are the root lattices $A_N (N \geq 1)$, $D_N (N \geq 2)$, and $E_N (N = 6, 7, 8)$ and their duals which yield the densest known sphere packings and coverings for $N \leq 8$ (Conway and Sloane, 1982b). As an example of the calculation of the quantities in Eqs. (136) and (137), consider the two-dimensional lattice $A_2 (L_2)$. The basis vectors for this lattice are $\mathbf{a}_1 = (1, 0)$, $\mathbf{a}_2 = (-\frac{1}{2}, \frac{\sqrt{3}}{2})$, so that

$$M = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$$

and $\det \Lambda = \det M = \sqrt{3}/2$ (Sloane, 1981). With $\rho = 1/2$, we have from Eq. (137) that $\Delta = .9069$, which agrees with the earlier direct calculation. It is also easy to check that with $\rho = 1/2$, the volume of the A_2 lattice's Voronoi region (a hexagon with side $1/\sqrt{3}$) is $\det \Lambda = \sqrt{3}/2$. The dual lattice A_2^* only differs from A_2 by a rotation and scale change, so these lattices are considered equivalent, which is indicated by the notation $A_2^* \cong A_2$.

To compute the MSE per dimension for lattice quantizers in higher dimensions, it is common to rely on Gersho's previously mentioned conjecture that for large L, the Voronoi regions of an optimal quantizer are all congruent to some polytope, say P, and define quantities called the volume, the unnormalized second moment, and the normalized second moment of P, respectively, as

$$\text{vol}(P) = \int_P d\mathbf{x}, \tag{139}$$

$$U(P) = \int_P \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \, d\mathbf{x}, \tag{140}$$

and

$$I(P) = \frac{U(P)}{\text{vol}(P)}, \tag{141}$$

where $\hat{\mathbf{x}}$ is the centroid of P. Using Eqs. (139)–(141), we can then define the dimensionless second moment of P, denoted $G(P)$, to be

$$G(P) = \frac{1}{N} \frac{U(P)}{\text{vol}(P)^{1+2/N}} = \frac{1}{N} \frac{I(P)}{\text{vol}(P)^{2/N}}. \tag{142}$$

Gersho (1979) calls the quantity in Eq. (142) the coefficient of quantization, but it is equivalent to the MSE per dimension for large L as previously calculated for lattices in dimensions 1 and 2 under the assumption of a uniform input distribution.

The connection between $G(P)$ and the MSE per dimension can also be made through a result of Zador's. If the MSE per dimension is

$$D(N) = \frac{1}{N} \int_{\mathcal{R}^N} \|\mathbf{x} - Q(\mathbf{x})\|^2 f(\mathbf{x}) \, d\mathbf{x}, \tag{143}$$

then under rather general assumptions on $f(\mathbf{x})$, Zador (1982) showed that

$$\lim_{L \to \infty} L^{2/N} D(N) = G_N \left(\int_{\mathcal{R}^N} f(\mathbf{x})^{N/(N+2)} \, d\mathbf{x} \right)^{(N+2)/N} \tag{144}$$

where G_N does not depend upon $f(\mathbf{x})$. Therefore, G_N is interpreted to be the minimum MSE per dimension achievable by vector quantization, and

assuming (as Gersho conjectures) that the Voronoi regions are all congruent to some polytope P, then

$$G_N = \min_P G(P) \tag{145}$$

where the minimum is taken over all admissible N-dimensional polytopes (Conway and Sloane, 1982a). Since G_N does not depend upon $f(\mathbf{x})$, any convenient $f(\mathbf{x})$ can be used to find G_N, and hence, $f(\mathbf{x})$ is often chosen to be uniform. If Eq. (145) holds, then we can find G_N by calculating $G(P)$ for all admissible N-dimensional polytopes and selecting the smallest as G_N. If the conjecture does not hold or if we cannot specify all possible admissible polytopes, then we still have an upper bound on G_N by finding $G(P)$ for any admissible P.

For $N = 1$, the optimum uniform quantizer is a uniform partition of the real line and $G_1 = \frac{1}{12} = 0.08333\ldots$. In two dimensions there are many admissible polytopes, including all triangles, quadrilaterals, and hexagons (Gersho, 1979), but the minimum MSE per dimension is achieved by the hexagonal quantizer based upon the A_2 lattice, and $G_2 = 5/36\sqrt{3} = 0.0801875\ldots$ (Conway and Sloane, 1982a; Newman, 1982). Gersho (1979) specified five admissible polytopes in three dimensions, namely, the cube, the hexagonal prism, the rhombic dodecahedron, the elongated dodecahedron, and the truncated octahedron, and found by calculating $G(P)$ for all five polytopes that the truncated octahedron had the smallest $G(P)$ of these five which is 0.0785433.... Table VI lists $G(P)$ for four of the admissible polytopes in $\mathcal{R}^3$. He conjectured that this value was not just an upper bound to G_3, but that the truncated octahedron is the optimal polytope in three dimensions so that $G_3 = 0.0785433\ldots$. This conjecture is proved by Barnes and Sloane (1983) who show that the optimal lattice quantizer in three dimensions is based upon the body centered cubic lattice $D_3^* \cong A_3^*$, which has Voronoi regions that are truncated octahedra.

As the dimension of the VQ increases, the problem centers around finding admissible space-filling polytopes and then evaluating $G(P)$. The principal

TABLE VI

$G(P)$ FOR FOUR POLYTOPES IN $\mathcal{R}^3$ (CONWAY AND SLOANE, 1982)[12]

P	$G(P)$
Cube	.0833333...
Hexagonal Prism	.0812227...
Rhombic Dodecahedron	.0787451...
Truncated Octahedron	.0785433...

approach to solving this problem has been to determine the Voronoi regions corresponding to the root lattices in each dimension, calculate $G(P)$, and select the lattice with the smallest $G(P)$ as the best *known* lattice quantizer of dimension N. Conway and Sloane (1982) have carried out this procedure for the lattices $A_N (N \geq 1)$, $A_N^* (N \geq 1)$, $D_N (N \geq 3)$, $D_N^* (N \geq 3)$, E_6, E_7, and $E_8 = E_8^*$. Neither finding the Voronoi regions for a lattice nor evaluating the corresponding $G(P)$ is necessarily simple, and different methods may have to be used for different lattices. For example, the lattice A_N and its dual A_N^* demand quite a separate treatment for $N > 2$. The Voronoi regions and normalized second moment are calculated by Monte Carlo integration for the E_6^* and E_7^* lattices in Conway and Sloane (1984). Table VII lists the best known lattice quantizers in dimensions 1–10 along with the normalized second moment $G(P)$.

Also shown in Table VII is something called the sphere bound. Zador (1982) showed that a lower bound to G_N is $(1/N + 2)V_N^{-2/N}$ for the squared error distortion measure, where V_N is the volume of an N-dimensional sphere as given in Eq. (138). This lower bound is the column labeled "Sphere Bound" in Table VII. Another lower bound to G_N suggested by Conway and Sloane (1985) is presented in the "Proposed Bound" column. While this bound is tighter than the sphere bound, only a plausability argument has been given for its validity.

Another way to find candidates for good vector quantizers in N dimensions is to study lattices which have the densest known sphere packings. Lattices which fall in this category are the Coxeter-Todd lattice K_{12} (Coxeter and Todd, 1953), the Barnes-Wall lattice Λ_{16} (Barnes and Wall, 1959), and the

TABLE VII

BEST KNOWN LATTICE QUANTIZERS AND DIMENSIONLESS SECOND MOMENT $G(P)$ (CONWAY AND SLOANE 1982 AND 1984)[13]

N	Sphere Bound	Proposed Bound (Conway and Sloane, 1985)	Best Lattice	$G(P)$
1	.0833	.0833	$A_1 \cong A_1^*$	.0833
2	.0796	.0802	$A_2 \cong A_2^*$	.0802
3	.0770	.0779	$A_3^* \cong D_3^*$	.0785
4	.0750	.0761	$D_4 \cong D_4^*$	.0766
5	.0735	.0747	D_5^*	.0756
6	.0723	.0735	E_6^*	.0742
7	.0713	.0725	E_7^*	.0731
8	.0704	.0716	$E_8 = E_8^*$	.0717
9	.0697	.0709	D_9^*	.0747
10	.0691	.0703	D_{10}^*	.0747

Leech lattice Λ_{24} (Leech, 1964 and 1967). Conway and Sloane (1984) use Monte Carlo integration to compute the normalized second moment for VQs based upon these lattices as $G(K_{12}) = 0.0701$, $G(\Lambda_{16}) = 0.0683$, and $G(\Lambda_{24}) = 0.0658$. The duals of these lattices are also contained in the original lattice, so $K_{12} \cong K_{12}^*$, $\Lambda_{16} \cong \Lambda_{16}^*$, and $\Lambda_{24} = \Lambda_{24}^*$, and VQs based upon these lattices are the best known quantizers in their respective dimensions. Figure 26 presents the normalized second moment $G(P)$ for several important lattice quantizers, as well as the sphere lower bound, the conjectured lower bound of Conway and Sloane (1985), and Zador's upper bound given by

$$G_N \leq \frac{1}{N\pi}\Gamma\left(\frac{N}{2}+1\right)^{2/N}\Gamma\left(1+\frac{2}{N}\right). \tag{146}$$

In light of the results in Fig. 26, which show that known lattice quantizers are close to the sphere bound and extremely close to the proposed bound, it is natural to inquire as to how close the performance of these quantizers is to $D(R)$. Recall the result of Gish and Pierce (1968) that the optimum entropy constrained scalar quantizer for the MSE distortion measure performs within

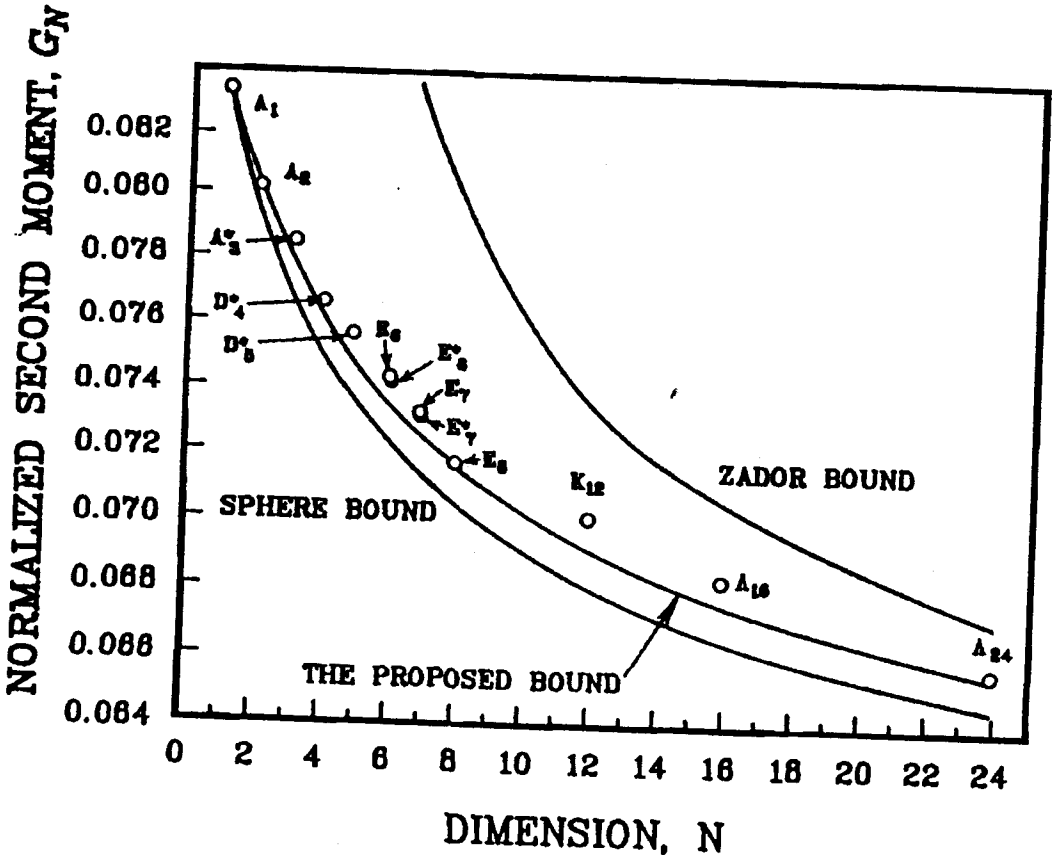

FIG. 26. Performance comparison of several important lattice quantizers (Conway and Sloane, 1982, 1984 and 1985).[14]

1.53 dB of $D(R)$ for a large number of quantization levels L. In comparison, for $N = 8$ and large L, the Gossett lattice (E_8) quantizer with entropy coding reduces this gap to (Makhoul, Roucos and Gish, 1985)

$$1.53 - 10\log_{10}\left[\frac{G(A_1)}{G(E_8)}\right] = 1.53 - 10\log_{10}\left[\frac{0.0833}{0.0717}\right]$$
$$= 0.879dB,$$

where $G(A_1)$ and $G(E_8)$ are taken from Table VII. The reduction in the rate R provided by a higher dimensional quantizer with respect to scalar quantization can be expressed as

$$\Delta R = \left[10\log_{10}\left(\frac{0.0833}{G(P)}\right)\right] \Big/ 6.02 \text{ bits/sample}, \tag{147}$$

and Makhoul, Roucos, and Gish (1985) plot this quantity for many of the best available lattices through dimension $N = 24$. We can check the results obtained in Eqs. (39) and (40) by substituting $G(A_2) \cong 0.0802$ (actually, 0.0801875...) into Eq. (147) to find a rate reduction of 0.028 bits/sample for the hexagonal quantizer over the scalar quantizer.

Lattice quantizer performance results for sources other than those with a uniform distribution are relatively meager. Some performance comparisons for rates less than 2 bits/sample and Gaussian, Laplacian, and Gamma distributed sources are given in Sayood, Gibson and Rost (1984), Rost and Sayood (1984) and Rost (1984).

Of course, the most important question of all is, "How well do these lattice quantizers perform for moderate N and L?" Now we are moving completely out of the realm of the performance analyses presented previously, since the source distribution may no longer be uniform and edge effects at the overload regions may not be negligible. There is a dearth of results for lattice quantizers with finite N and L, with most VQ work in this range emphasizing the LBG algorithm. One particularly striking illustration of the perceptual performance improvement available with lattice VQ is provided by the results in Sayood, Gibson and Rost (1984), where the A_N^* lattices are used to quantize the two-dimensional discrete cosine transform (DCT) coefficients calculated on a monochrome 256 by 256 pixel image at 0.5 bit/pixel for one, four, and eight-dimensional quantization. These results are reproduced here in Fig. 27(a) for scalar quantization, Fig. 27(b) for A_4^* lattice quantization, and Fig. 27(c) for A_8^* lattice quantization. The performance improvement is quite phenomenal. Figure 27(d) is the image that has been reconstructed using the same DCT coefficients as in Figs. 27(a)–(c), but without the coefficients being quantized. Comparing Figs. 27(c) and (d) reveals that the eight-dimensional quantizer is

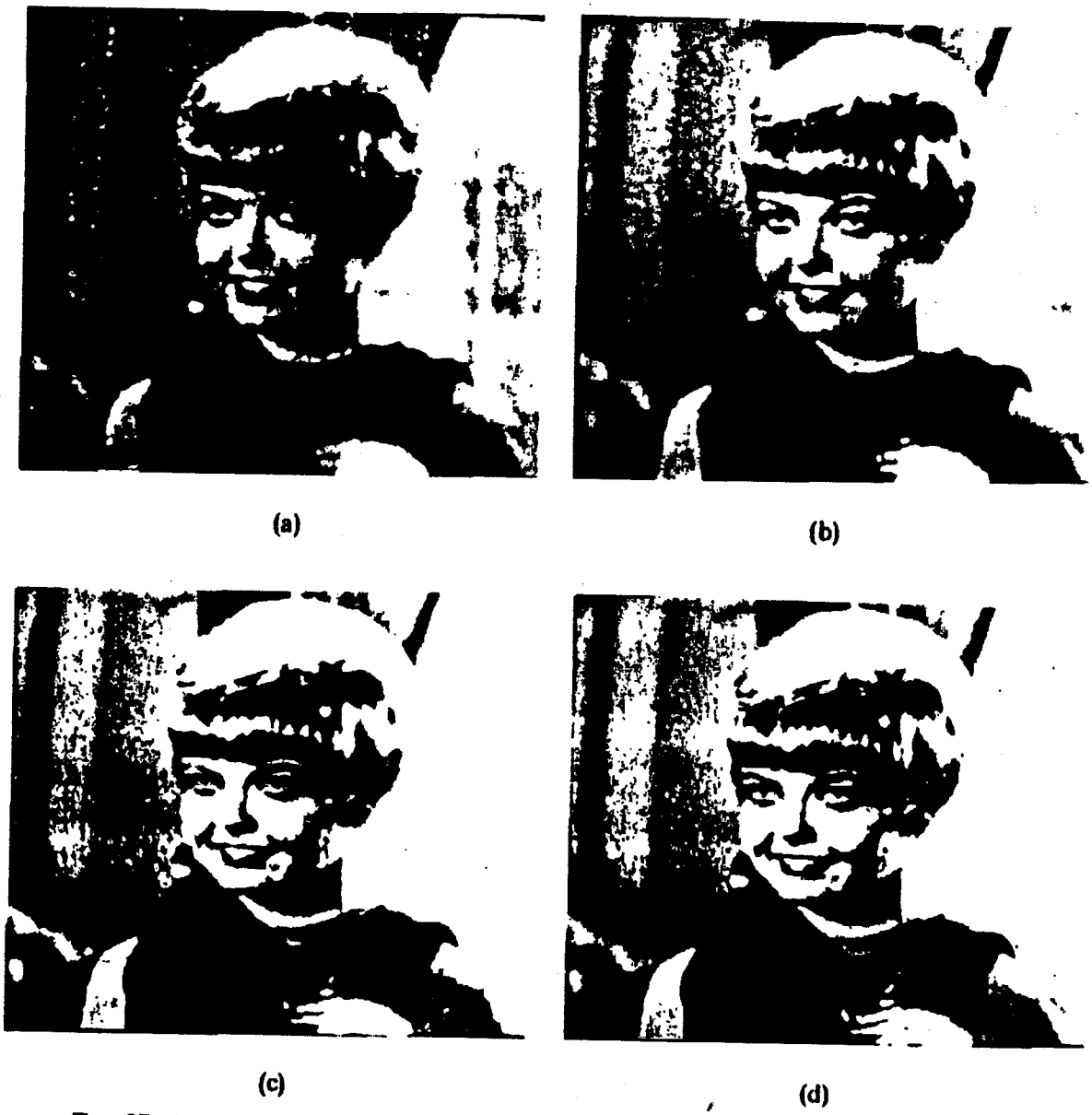

FIG. 27. A comparison of one, four, and eight dimensional lattice quantization of DCT coefficients (Sayood, Gibson and Rost, 1984). (a) Scalar quantizer. (b) Four-dimensional quantizer. (c) Eight-dimensional quantizer, (d) Ideal quantizer.[15]

contributing very little distortion. For more details, the reader is referred to Sayood, Gibson, and Rost (1984). Since A_4^* is not the optimal lattice quantizer (for uniform inputs) in four dimensions and A_8^* is not the optimum lattice quantizer (for uniform inputs) in eight dimensions, it may be possible to improve on the performance shown in Fig. 27. Performance curves for the A_N and $A_N^*(N \geq 1)$ lattices and the D_N and $D_N^*(N \geq 4)$ lattices are shown in Conway and Sloane (1982a; 1984; 1985).

It is interesting to observe, as pointed out by Conway and Sloane, that for all of the results presently available, the optimal or best known lattice quantizer is the *dual* of the densest lattice *packing* (sphere packing whose sphere centers form a lattice). This goes against our intuition since it says that the best N-dimensional VQ is not the same as the best lattice *covering* of N-space. As specific examples, note from Table VII that D_4^* and D_5^* are the best quantizers in four and five dimensions, but it has been shown that A_4^* and A_5^* are the optimal coverings. Furthermore, the best known coverings for dimensions $N \leq 23$ are based on the A_N^* lattices (Ryskov and Baranovskii, 1978; Bambah and Sloane, 1982; Conway and Sloane 1984).

IX. Research Areas and Connections to Other Fields

There is much current research on lattice VQs along the same lines as the work described herein; that is, work is proceeding to find the optimal lattice coverings and packings for all those cases still open, to determine the Voronoi regions of lattices, to compute $G(P)$ and ultimately G_N, and to discover fast encoding and decoding algorithms. Another direction of VQ research based upon lattices is to employ lattices as an encoding tool for non-lattice VQs. One such effort is motivated by the work of Sakrison (1968) alluded to in Sec. III. Sakrison showed that for an N-vector of Gaussian i.i.d. source samples, that as $N \to \infty$, the source vectors fall with high probability on the surface of an N-dimensional sphere. Thus, a good VQ would simply place its representation vectors throughout this high probability region (the sphere surface). This same concept is pursued in Fischer (1986) for a memoryless Laplacian source. In this case, the region of high probability for the source vectors is the surface of an N-dimensional hyperpyramid. Based upon this observation for N large, a finite N VQ is proposed where the output points lie on concentric hyperpyramids, but only those points which lie on the pyramid that are also points of the cubic lattice Z^N are allowable representation points. Thus, the output points lie on a lattice, and a fast encoding procedure is possible (Fischer, 1986). Applications of this pyramid vector quantization approach to speech and images are given in Fischer and Malone (1985) and Tseng and Fischer (1987), respectively. Similar research in spirit has been conducted by Adoul (1986a, b) on what he called spherical vector quantizers. These quantizers are constructed from the points of the Leech lattice (Λ_{24}) that fall on the shells at various radii around the origin. The norm (radius) and the lattice point in a shell are encoded separately, and the main ideas behind this approach are that "sphere hardening" is already taking effect in 24 dimensions and that relatively

fast quantizing algorithms are available because of the lattice structure. Another lattice-based VQ performs the encoding in two steps (Moayeri, Neuhoff and Stark, 1985). First, the source vector is finely quantized using a VQ with a fast encoding algorithm, and second a table look-up finds the codebook output point which is closest to the finely quantized vector.

Yet another research area is that of multidimensional companding. Motivated by the success of logarithmic companding for the scalar quantization of speech signals, investigations are underway to utilize multidimensional companding with lattice based uniform quantizers to produce nonuniform VQs with reduced encoding complexity (Bucklew, 1981 and 1984).

A totally different research area that makes use of lattices is that of coding for reliable transmission of information over communications channels. Pertinent references for this field are Sloane (1984 and 1981), Conway and Sloane (1982b), Leech and Sloane (1971), Forney (to be published) and Forney (1984). The papers by Forney present some very interesting constructions for some of the lattices discussed here which may serve as fast algorithms for vector quantization in the near future.

Further results on these topics are left to the references.

X. Conclusions

An introduction to vector quantization, in general, and lattice quantization, in particular, has been provided. The development presented here shows that lattice-based vector quantizers can perform arbitrarily close to the rate distortion bound as the number of dimensions becomes large and that it may be possible to avoid entropy coding of the quantizer output points with lattice quantizers. Furthermore, fast quantization algorithms are known for many important vector quantizers. On the other hand, the performance improvement provided by lattice quantizers over scalar quantization with entropy coding may be only a few tenths of a dB. What is definitely lacking, however, is enough applications of lattice quantizers to non-i.i.d. sources, such as speech and images, to be able to discern the available subjective performance gains not evident in the mathematical analyses of idealized sources. The few studies available are encouraging, but much work is needed in this area.

The development in this chapter includes enough mathematical detail for the reader to be able to implement lattice quantizers for many applications and to allow judicious tradeoffs among the various lattice-based vector quantizers to be made. A mastery of the material in this chapter is a necessary background for a fruitful investigation of the literature on vector quantization and lattice quantizers; however, this material is not sufficient to conduct

research on many fundamental theoretical issues which remain unresolved in lattice quantization, such as finding the Voronoi regions of certain lattices, and a more detailed examination of the references is required to pursue this goal.

ACKNOWLEDGMENT

The authors are indebted to Dr. Thomas R. Fischer for numerous discussions concerning vector quantization over the past few years.

NOTES

[1] Adapted from Table 8.1 of R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications*, copyright © 1974 John Wiley & Sons, New York. Reprinted by permission of John Wiley & Sons, Inc.

[2] Adapted from Table 1 on p. 59 of J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory*, Springer-Verlag, New York, 1972.

[3] Adapted from the diagram at the top of p. 58 of J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory*, Springer-Verlag, New York, 1972.

[4] Adapted from Table 8.2 of R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications*, copyright © 1974 John Wiley & Sons, New York. Reprinted by permission of John Wiley & Sons, Inc.

[5] Adapted from Fig. 4 of J. H. Conway and N. J. A. Sloane, "On the Voronoi regions of certain lattices," *SIAM J. Algebraic Discrete Methods*, vol. 5, pp. 294–305, 1984, copyright © 1984 Society for Industrial and Applied Mathematics, Philadelphia, PA.

[6] Adapted from Fig. 4 of N. J. A. Sloane, "Tables of sphere packings and spherical codes," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 327–338, May 1981. Copyright © 1981 IEEE.

[7] Adapted from Fig. 5 of N. J. A. Sloane, "Tables of sphere packings and spherical codes," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 327–338, May 1981. Copyright © 1981 IEEE.

[8] Adapted from the figure on p. 122 of N. J. A. Sloane, "The packing of spheres," *Scientific American*, pp. 116–125, Jan. 1984. Copyright © 1984 by Scientific American, Inc. All rights reserved.

[9] Adapted from Fig. 2 of N. J. A. Sloane, "Tables of sphere packings and spherical codes," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 327–338, May 1981. Copyright © 1981 IEEE.

[10] Adapted from Fig. 2 of 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. Copyright © 1983 IEEE.

[11] Adapted from the figure on p. 118 of N. J. A. Sloane, "The packing of spheres," *Scientific American*, pp. 116–125, Jan. 1984. Copyright © 1984 by Scientific American, Inc. All rights reserved.

[12] Adapted from Table I of J. H. Conway and N. J. A. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 211–226, March 1982. Copyright © 1982 IEEE.

[13] Adapted from Table V of J. H. Conway and N. J. A. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 211–226, March 1982. Copyright © 1982 IEEE.

[14] Adapted from Fig. 1 of J. H. Conway and N. J. A. Sloane, "A lower bound on the average error of vector quantizers," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 106–109, Jan. 1985. Copyright © 1985 IEEE.

[15] Adapted from Figs. 3–6 of K. Sayood, J. D. Gibson, and M. C. Rost, "An algorithm for uniform vector quantizer design," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 805–814, Nov. 1984. Copyright © 1984 IEEE.

REFERENCES

Adams, Jr., W. C., and Geisler, C. E. (1978). "Quantizing characteristics for signals having Laplacian amplitude probability function," *IEEE Trans. Commun.* COM-26, 1295–1297.

Adoul, J.-P. (1986a). "La quantification vectorielle des signaux: Approache algébrique," *Ann. Télécommun.* 41.

Adoul, J.-P. (1986b). "Decoding algorithm for spherical codes from the Leech lattice," submitted for publication.

Adoul, J.-P. (1987). "Speech-Coding Algorithms and Vector Quantization," in *Advanced Digital Communications*, K. Feher, ed., Prentice-Hall, Inc. Englewood Cliffs, NJ, pp. 133–181.

Bambah, R. P. and Sloane, N. J. A. (1982). "On a problem of Ryskov concerning lattice coverings," *Acta Arithmetica* 42, 107–109.

Barnes, E. S. and Sloane, N. J. A. (1983). "The optimal lattice quantizer in three dimensions," *SIAM J. Algebraic Discrete Methods* 4, 30–41.

Barnes, E. S. and Wall, G. E. (1959). "Some extreme forms defined in terms of Abelian groups," *J. Australian Math. Soc.* 1, 47–63.

Berger, T. (1971). *Rate Distortion Theory*, Prentice-Hall, Inc. Englewood Cliffs, NJ.

Bucklew, J. A. (1981). "Companding and random quantization in several dimensions," *IEEE Trans. Inform. Theory* IT-27, pp. 207–211.

Bucklew, J. A. (1984) "Two results on the asymptotic performance of quantizers," *IEEE Trans. Inform. Theory* IT-30, 341–348.

Cohn, H. (1962). *A Second Course in Number Theory*, John Wiley & Sons, New York.

Conway, J. H. and Sloane, N. J. A. (1982a). "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans. Inform. Theory* IT-28, 211–226.

Conway, J. H. and Sloane, N. J. A. (1982b). "Fast quantizing and decoding algorithms for lattice quantizers and codes," *IEEE Trans. Inform. Theory* IT-28, 227–232.

Conway, J. H. and Sloane, N. J. A. (1982c). "Laminated lattices," *Annals of Mathematics* 116, 593–620.

Conway, J. H. and Sloane, N. J. A. (1983). "A fast encoding method for lattice codes and quantizers," *IEEE Trans. Inform. Theory* IT-29, 820–824.

Conway, J. H. and Sloane, N. J. A. (1984). "On the Voronoi regions of certain lattices," *SIAM J. Algebraic Discrete Methods* 5, 294–305.

Conway, J. H. and Sloane, N. J. A. (1985). "A lower bound on the average error of vector quantizers," *IEEE Trans. Inform. Theory* IT-31, 106–109.

Coxeter, H. S. M. (1961). *Introduction to Geometry*, John Wiley & Sons, New York.

Coxeter, H. S. M. and Todd, J. A. (1953). "An extreme duodenary form," *Canad. J. Math.* 5, 384–392.

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

Fischer, T. R. (1986). "A pyramid vector quantizer," *IEEE Trans. Inform. Theory* IT-32, 568–583.

Fischer, T. R. and Malone, K. T. (1985). "Transform Coding of Speech with Pyramid Vector Quantization," *Conf. Rec.*, MILCOM '85, 620–623.

Forney, Jr., G. David. "Coset codes I: Geometry and classification," *IEEE Trans. Inform. Theory*, to appear.

Forney, Jr., G. David. "Coset codes II: Binary lattices and related codes," *IEEE Trans. Inform. Theory*, to appear.

Forney, Jr., G. David, *et al.* (1984). "Efficient modulation for bandlimited channels," *IEEE J. Selected Areas Commun.* SAC-2, 632–647.

Gallager, R. G. (1968). *Information Theory and Reliable Communication*, John Wiley and Sons, Inc., New York.

Gameckii, A. F. (1962). "On the theory of covering n-dimensional space with equal spheres," *Soviet Math.* 3, 1410–1414.

Gersho, A. (1979). "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory* IT-25, 373–380.

Gersho, A. (1982). "On the structure of vector quantizers," *IEEE Trans. Inform. Theory* IT-28, 157–166.

Gersho, A. (1986). "Vector Quantization: A New Direction in Source Coding," *Digital Communications*, E. Biglieri and G. Prati, eds., North-Holland, Amsterdam, 267–281.

Gersho, A. and Cuperman, V. (1983). "Vector quantization: A pattern-matching technique for speech coding," *IEEE Communications Magazine* 21, 15–21.

Gilmore, R. (1974). *Lie Groups, Lie Algebras, and Some of Their Applications*, John Wiley & Sons, New York.

Gish, H. and Pierce, J. N. (1968). "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory* IT-14, 676–683.

Gray, R. M. (1984). "Vector quantization," *IEEE ASSP Magazine* 1, 4–29.

Gray, R. M. and Davisson, L. D. (1974). "A Mathematical Theory of Data Compression?," in *Proc. 1974 Int. Conf. Commun.*, 40A-1–40A-5.

Grove, L. C. and Benson, C. T. (1985). *Finite Reflection Groups*, Second edition, Springer-Verlag, New York.

Humphreys, J. E. (1972). *Introduction to Lie Algebras and Representation Theory*, Springer-Verlag, New York.

Jayant, N. S. and Noll, P. (1984). *Digital Coding of Waveforms*, Prentice-Hall, Inc., Englewood Cliffs, NJ.

Joshi, A. W. (1977). *Elements of Group Theory for Physicists*, Second ed., John Wiley & Sons, New York.

Leech, J. (1967). "Notes on sphere packings," *Canad. J. Math.* 19, 251–267.

Leech, J. (1964). "Some sphere packings in higher space," *Canad. J. Math.* 16, 657–682.

Leech, J. and Sloane, N. J. A. (1971). "Sphere packings and error-correcting codes," *Can. J. Math.* 23, 718–745.

Lekkerkerker, C. G. (1969). *Geometry of Numbers*, John Wiley & Sons, New York.

Linde, Y., Buzo, A. and Gray, R. M. (1980). "An algorithm for vector quantizer design," *IEEE Trans. Commun* COM-28, 84–95.

Lyusternik, L. A. (1963). *Convex Figures and Polyhedra*, Dover, New York.

MacQueen, J. (1967). "Some Methods for Classification and Analysis of Multivariate Observations." in *Proc. 5th Berkeley Symp. on Math Statist., and Prob.*, Berkeley, CA: Univ. of Calif. Press, 281–297

MacWilliams, F. J. and Sloane, N. J. A. (1977). *The Theory of Error-Correcting Codes*, North-Holland, Amsterdam.

Max, J. (1960). "Quantizing for minimum distortion," *IRE Trans. Inform. Theory* IT-6, 7–12.

Makhoul, J., Roucos, S. and Gish, H. (1985). "Vector quantization in speech coding," *Proc. IEEE* 73, 1551–1588

Moayeri, N., Neuhoff, D. L. and Stark, W. E. (1985). "Fast Vector Quantizers," in *Proc. of the 23rd Annual Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, 347–353.

Newman, D. J. (1982). "The hexagon theorem," *IEEE Trans. Inform. Theory* IT-28, 137–139.

Rost, M. C. (1984). "Lattice Quantization," M.S. Thesis, Dept. of Electrical Eng., University of Nebraska, Lincoln, NE.

Rost, M. C. and Sayood, K. (1984). "Investigation of Lattice Vector Quantizer," *Proc. of the Twenty-Seventh Midwest Symp. on Circuits and Systems*, Morgantown, West Va., 149–152.

Ryskov, S. S. and Baranovskii, E. P. "C-Types of n-dimensional lattices and 5-dimensional primitive parallelohedra (with application to the theory of coverings)" (in Russian), *Trudy Mat. Inst. Steklov.*, 137, 1976. English translation in *Proc. Steklov. Inst. Math.*, Issue 4, 1978.

Sakrison, D. J. (1968). "A geometric treatment of the source encoding of a Gaussian random variable," *IEEE Trans. Inform. Theory* IT-14, 481–486.

Sakrison, D. J. (1979). "Image Coding Applications of Vision Models," in *Image Transmission Techniques*, W. K. Pratt, ed., Academic Press, New York, 21–51.

Sayood, K. Gibson, J. D. and Rost, M. C. (1984). "An algorithm for uniform vector quantizer design," *IEEE Trans. Inform. Theory* IT-30, 805–814.

Shannon, C. E. (1948). "A mathematical theory of communication," *Bell Syst. Tech. J.* 27, 379–423, 623–656.

Shannon, C. E. (1959). "Coding Theorems for a Discrete Source with a Fidelity Criterion," in *IRE Nat. Conv. Rec.*, Pt. 4, 142–163.

Sloane, N. J. A. (1981). "Tables of sphere packings and spherical codes," *IEEE Trans. Inform. Theory* IT-27, 327–338.

Sloane, N. J. A. (1984). "The packing of spheres," *Scientific American*, 116–125.

Swaszek, P. F. (1986). "Vector Quantization," in *Communications and Networks*, I. F. Blake and H. V. Poor, eds., Springer-Verlag, New York, 362–389.

Tseng, H.-C. and Fischer, T. R. (1987). "Transform and hybrid transform/DPCM coding of images using pyramid vector quantization," *IEEE Trans. Commun.* COM-35, 79–86.

Zador, P. (1982). "Asymptotic quantization error of continuous signals and their quantization dimension," *IEEE Trans. Inform. Theory* IT-28, 139–149.