AN ABSTRACT OF THE THESIS OF

VISHNU BALCHAND JUMANI for the DOCTOR OF PHILOSOPHY
(Name) (Degree)

in APPLIED MATHEMATICS presented on March 28, 1973
(Major) (Date)

Title: A STUDY OF THE QUANTIZATION PROCEDURE: HIGHER ORDER INFORMATION AND THE CONDITIONS FOR MINIMUM ERROR

Signature redacted for privacy.

Abstract approved: _____
(William M. Stone)

Quantization is a non-linear operation of converting a continuous signal into a discrete one, assuming a finite number of levels N. A study is made of the quantization procedure, starting from the year 1898 to the present time. Conditions for minimum error are derived with consideration of quantization in magnitude and time. An extension of the Mehler and Carlitz formulas involving Hermitian polynomials (quadrilinear case) has been created. Further, investigation is conducted toward obtaining an autocorrelation function of the output of the quantizer for Gaussian input. The method calls for the use of two different forms of the Euler-Maclaurin sum formulas and results are derived for a hard limiter, linear detector, clipper, and a smooth limiter. The method lends itself to the extension to the non-uniform case.

A Study of the Quantization Procedure: Higher Order Information and the Conditions for Minimum Error.

by

Vishnu Balchand Jumani

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1973

APPROVED:

Signature redacted for privacy.

Professor of Mathematics

In Charge of Major

Signature redacted for privacy.

Chairman of Department of Mathematics

Signature redacted for privacy.

Dean of Graduate School

Date thesis is presented March 28, 1973

Typed by W. M. Stone for Vishnu Balchand Jumani

ACKNOWLEDGEMENT

I am thankful and will always be very grateful to my major professor, William M. Stone, for the kindness, patient help, encouragement and guidance that I received from him during this research and my studies in general.

I am thankful to my wife Debborah for her understanding and untiring encouragement.

I would also like to extend my appreciation to Professor Arvid T. Lonseth, Department of Mathematics, Oregon State University, for his kindness and his extension of a part of the research grant under the Atomic Energy Commission for the completion of this thesis.

I also wish to extend my appreciation to Ms. Jolan Eross, secretary in the Mathematics Department, for her assistance in this entire project.

TABLE OF CONTENTS

A STUDY OF N-LEVEL QUANTIZATION: HIGHER ORDER INFORMATION AND CONDITIONS FOR MINIMAL ERROR

I. QUANTIZATION AND ITS HISTORY

The quantization process is very widely used today in the field od communication systems. However, its origin dates back to Sheppard (1898), who derived what is known today as Sheppard's correction formula. He used the idea of breaking up the domain of the frequency function $p(x)$--which is assumed to be single-valued and continuous--into equal intervals of length ω: that is,

$$\omega = x_{i+1} - x_i, \quad i = 0, \pm 1, \pm 2, \cdots \cdots \quad . \tag{1.1}$$

He defined

$$A_i = \int_{x_i - \frac{\omega}{2}}^{x_i + \frac{\omega}{2}} p(x)\, dx = \int_{-\frac{\omega}{2}}^{\frac{\omega}{2}} p(x_i + y)\, dy \tag{1.2}$$

and by means of the Euler-MacLaurin identity arrived at his formula. Statistical data, divided into uniform intervals of the domain, thus may show its effect on variance. In the communication field, quantization is described as a nonlinear operation, converting a continuous incoming signal into an outgoing signal that can take on a finite number of levels. This operation, which is essentially an analog to digital conversion, intoduces an error. Our basic aim is to faithfully reproduce the quantizer input signal at the system output

terminal. To achieve this one has to minimize the error between the quantizer input and output. Figure I illustrates the input-output characteristic of a quantizer where $x_1, x_2, \cdots\cdots, x_N$ are the points which subdivide the input signal amplitude range into N nonoverlapping intervals and $y_1, y_2, \cdots\cdots, y_N$ indicate the outputs corresponding to the respective input subintervals.

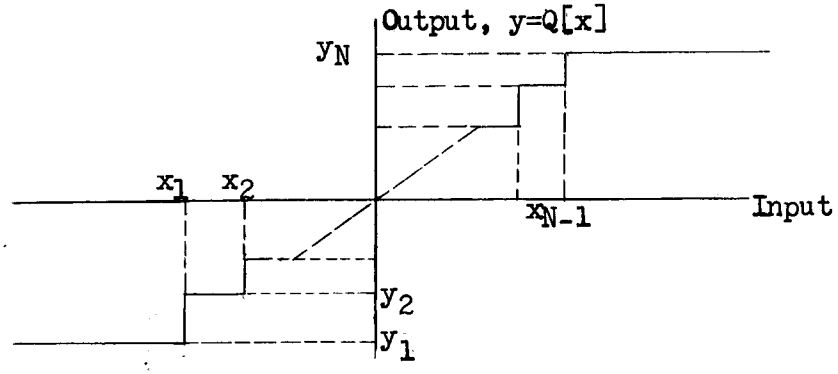

Figure 1.1

This results in N levels and thus the name N-level quantizer. If $x_{i+1} - x_i = \omega$, a constant for all i, then we have a uniform quantizer, otherwise we have a nonuniform quantizer. Goodall and Reeves (1947) were the first to implement this idea. They guided the construction of an 8-channel transmission system. In a subsequent paper published by Black and Edson (1947) quantization, along with the number and size of levels or steps, was investigated.

Widrow (1956) showed in analog to digital conversion that if the

probability density of the quantizer input signal is zero outside some bandwidth, then the amplitude density of the error signal, the difference between the input and output signal, is given by

$$p(x) = \begin{cases} \frac{1}{\omega}, & -\frac{\omega}{2} \leq x \leq \frac{\omega}{2}, \\ 0, & \text{elsewhere}, \end{cases} \tag{1.3}$$

where ω is the stepsize.

Studies were conducted concerning application of this pulse code modulation scheme involving sampling and quantization in transmission of telephone signals. Foremost in this field was Bennett (1948). His method called for quantization of the magnitude of speech signals. The selection was made, not from a continuous range of amplitudes but only from discrete ranges. The speech signal is replaced by a wave constructed of quantized values, the selection made on the basis of minimum distortion. The quantized signal is then transmitted and recovered at the receiver end, then restored to give the original message, provided the interference does not exceed half the difference between adjacent steps. He considered quantization of magnitude and time, whereby it was made possible to encode speech signals and to transmit a discrete set of magnitudes for each distinct time interval. Consider voltage quantization in time, depicted by Figure (1.2):

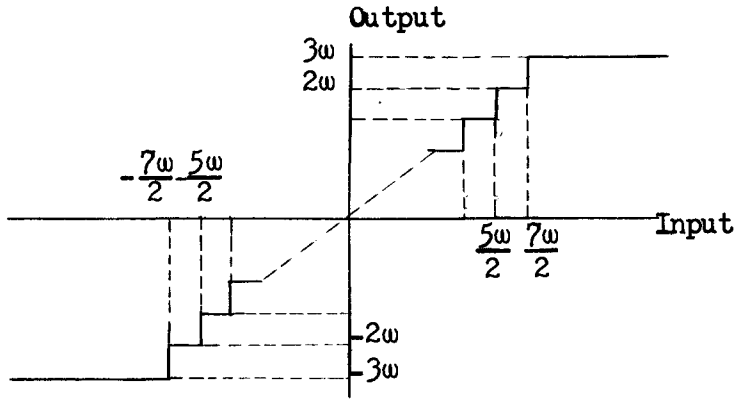

Figure 1.2

The distortion resulting from the quantization process--that is, the difference between input and outout--is shown in Figure (1.3), a saw-tooth function.

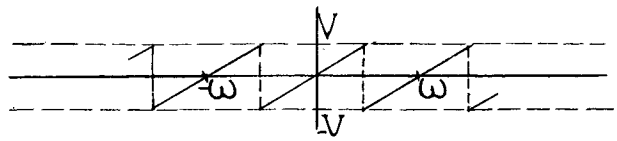

Figure 1.3

If V is the voltage corresponding to any one step and m the slope, then the error can be expressed as

$$e(t) = mt, \; -\frac{V}{2m} \leq t < \frac{V}{2m} , \tag{1.4}$$

and the mean square error is

$$E = \overline{e^2} = \frac{m}{V} \int_{-V/2m}^{V/2m} e^2(t)\, dt = \frac{V^2}{12} , \tag{1.5}$$

the well-known Sheppard's correction.

As mentioned by Bennett (1948), not all distortions of the original signal fall within the signal band. Higher order modulations may have frequencies quite different from those in the original signal,

which can be eliminated by a sharply defined filter. It thus becomes important to calculate the spectrum of the error wave, which is possible by using the theory of correlation. This is based on the fact that the power spectrum of the wave is the Fourier cosine transform of the correlation function. At this stage we shall introduce a notation that will help us review the work of several authors since the earliest of times. As described above, quantization is the nonlinear operation of converting a continuous signal into a discrete signal that assumes a finite number N levels. A typical input-output relationship is exhibited in Figure (1.1). The output is denoted by y_k when the input signal x lies in the range $x_{k-1} \leq x < x_k$. In most communication systems the main problem is to reproduce the original input signal at the receiver output. The quantization process introduces a certain amount of error which is denoted by

$$e(t) = x(t) - Q[x(t)] , \tag{1.6}$$

where $x(t)$ is the input signal and $Q[x(t)]$ is the characteristic of the quantizer. The continuous signal of Equation (1.6) is written as

$$e = x - Q[x] \tag{1.7}$$

for notational convenience. The mean value of e in Equation (1.6) measures the efficiency of the quantizer, defined as

$$E = \int_{-\infty}^{\infty} f(x - Q[x])\, p(x)\, dx , \tag{1.8}$$

where $p(x)$ is the amplitude probability density of the input signal $x(t)$, and $f(e) = f(x - Q[x])$ is the error function. The $f(e)$ is assumed to be a nonnegative function of its argument since it is not desired to have positive and negative instantaneous values to cancel each other. Considering an N-level quantizer, the domain of definition is broken up into N nonoverlapping subintervals. Equation (1.8) can be written in the form

$$E = \sum_{k=o}^{N-1} \int_{x_k}^{x_{k+1}} f(x - Q[x])\, p(x)\, dx , \tag{1.9}$$

where $x_o = -\infty$ and $x_N = \infty$. If an explicit characteristic of the quantizer is defined such that

$$Q[x] = y_k , \quad x_{k-1} \leq x < x_k , \quad k = 1, 2, \cdots\cdots , N, \tag{1.10}$$

then Equation (1.9) can be written as

$$E = \sum_{k=o}^{N-1} \int_{x_k}^{x_{k+1}} f(x - y_{k+1})\, p(x)\, dx. \tag{1.11}$$

With the measure of error as indicated by Equation (1.11) Bennett (1948), Painter and Dite (1951), and Smith (1957) investigated the mean square error criterion with their error relationshio given by

$$E_2 = \sum_{k=o}^{N-1} \int_{x_k}^{x_{k+1}} (x - y_{k+1})^2\, p(x)\, dx . \tag{1.12}$$

It is noted that Equation (1.12) is a special case of Equation (1.11),

where E is replaced by E_2 for notational convenience and the error function $f(x) = x^2$. For best results it is necessary to minimize E_2 which depends on $2N-1$ quantizer parameters,

$$E_2 = E_2(x_1, x_2, \cdots\cdots, x_{N-1}, y_1, y_2, \cdots\cdots, y_N). \tag{1.13}$$

The work of the three authors mentioned above dealt with the minimization of E_2 for large values of N. Bennett (1948) further discusses the fact that in the case of speech it is advantageous to taper the steps of the quantizer in such a way that finer steps would be available for weak signals. Tapered quantization is equivalent to inserting complementary nonlinear, zero-memory transducers in the signal path before and after the analog-to-digital converter, as shown in Figure (1.4).

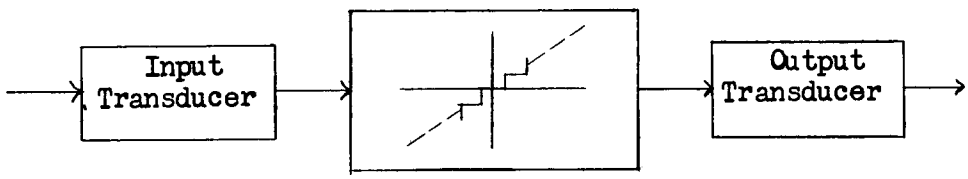

Figure-1.4

Smith (1957) calls the same process companding. Others who have studied this area of optimum and nonoptimum transducers include Lozovoy (1961), Davis (1962), Wiggins and Branham (1963), Mann, Straube and Villars (1962). Max (1960) works the expression for distortion as given by Equation (1.11) and has derived conditions for the minimum of E in Equation (1.11) for fixed N. He shows that for

minimum error,

$$f(x_j - y_j) = f(x_j - y_{j+1}), \quad j = 1, 2, \cdots, N-1 \tag{1.14a}$$

and

$$\int_{x_{j-1}}^{x_j} f'(x - y_j)\, p(x)\, dx = 0, \quad j = 1, 2, \cdots, N-1. \tag{1.14b}$$

His special application of Equation (1.14) to the mean-square error, as expressed by Equation (1.13), yields

$$x_j = \frac{y_{j+1} + y_j}{2}, \quad j = 1, 2, \cdots, N-1 \tag{1.15a}$$

and

$$y_j = \frac{\int_{x_{j-1}}^{x_j} x\, p(x)\, dx}{\int_{x_{j-1}}^{x_j} p(x)\, dx}, \quad j = 1, 2, \cdots, N. \tag{1.15b}$$

He further considers the case of input signal of normal amplitude distribution and derives expressions for the meab-square error. He also looks into equally spaced input-output relationshio which referred to as uniform quantization. Other authors who have studied the mean-square error in the quantization process include Panter and Dite (1951), Algazi (1966), Bluestein (1964) and Wood (1969). It should be noted here that the results derived by Max (1960) were first derived by Panter and Dite (1951), using a slightly different approach.

It is also to be noted that the procedure used to derive conditions for the minimum, as done by Max (1960), was first suggested by Garmash (1957). He concentrated his effort in minimizing the mean-square error as expressed by

$$E_2 = \sum_{k=o}^{N-1} \int_{x_k}^{x_{k+1}} (x - x_k)^2 \, p(x) \, dx \, . \qquad (1.15c)$$

In Equation (1.15c) the quantizer output $Q[x]$ is given by

$$Q[x] = x_k, \quad x_k \leq x \leq x_{k+1} \, , \quad k = 1, 2, \cdots, N-2 \, . \qquad (1.15d)$$

Differentiating Equation (1.15c) with respect to x_k he was able to generate an expression

$$\frac{\Delta_{k-1}}{\Delta_k} = \frac{1}{2} \sqrt{\frac{p(x_{k+1})}{p(x_k)} + 1} \, , \qquad (1.15e)$$

where $\Delta_{k-1} = x_k - x_{k-1}$.

Roe (1964) deals with a special case of Equation (1.11), the mth power distortion. In Equation (1.11) he sets $f(x - y_k) = |x - y_k|^m$, giving

$$E_m = \sum_{k=o}^{N-1} \int_{x_k}^{x_{k+1}} |x - y|^m \, p(x) \, dx \, , \; m = 1, 2, \cdots \, . \qquad (1.16)$$

Using Max's (1960) conditions in Equation (1.14) he was able to show that for minimum E_m the breakup of the input points x_k for a continuously differentiable density function $p(x)$ must satisfy approximately the relationship

$$\int_0^{x_n} [p(x)]^{\frac{1}{m+1}} dx = 2C_1 n + C_2 , \quad n = 1, 2, \cdots, N, \qquad (1.17)$$

where C_1 and C_2 are constants which can be adjusted to give the required fit at the extreme points x_0 and x_N. Max's (1960) tabulated results for x_k can be derived from Equation (1.17). Furthermore, Roe (1964) shows that $m = 2$, a normally distributed input, the x_n must satisfy the relationship

$$x_n = \sqrt{6}\ \mathrm{erf}^{-1}\left(\frac{2n-N}{N+\alpha}\right) , \qquad (1.18)$$

where $\alpha = 0.8532\cdots$ and the error function is well-tabulated in the Annals of the Harvard Computation Laboratory, No. 23 (1952).

Using Roe's (1964) result, Wood (1969) showed that E_2, the mean-square error as expressed by Max (1960), can be approximated as a function of N by a relationship

$$E_2 = \frac{2.73\ N}{(N + 0.8532)^3} . \qquad (1.19)$$

Zador (1964) generalized the Panter and Dite (1951) and the Roe (1964) results by considering multivariate distributions. Let $X = [x(1), x(2), \cdots, x(k)]$ be a random, vector-valued variable with probability measure ρ defined on Lebesgue-measurable subsets of k-dimensional Euclidean space E^k with absolutely continuous distribution function. Let $p(*)$ denote the density of X. Let $\{R_i\}$, $1 \leq i \leq N$, be a set of N Lebesgue measurable disjoint subsets of

E^k, with

$$\sum_{i=1}^{N} \rho(R_i) = 1 . \tag{1.20}$$

In this case the N-region quantizer with quantization regions R_i may be regarded as a function mapping the portion of E^k covered by the union of the R_i onto the integers 1 to N given by

$$Q(x) = i , \quad x \in R_i . \tag{1.21}$$

This maps each x into integer index i, $1 \leq i \leq N$, which labels the region R_i into which x falls. Quantization introduces an error in representation of x, since x must be estimated by some function f of $Q(x) = i$. Zador (1964) first showed that for $k = 1$, absolutely continuous and bounded p, and $0 < r < \infty$, the mean rth error is defined by

$$E_{min} = \frac{C_r}{N^r}\left[\int p^{\frac{1}{1+r}} d\lambda \right]^{1+r} , \tag{1.22}$$

where λ is the Lebesgue measure, C_r is a known constant. The Panter and Dite (1951) result is a special case of Equation (1.22) with $r = 2$. As seen there, $C_r = C_2 = 2/3$. Zador (1964) also considered $k > 1$ and derives the general result

$$E^{k}_{min} \approx \frac{C_{kr}}{N^{r/k}} \left[\int_{E^k} p^{\frac{k}{k+r}} \, d\lambda \right]^{\frac{k+r}{k}} . \qquad (1.23)$$

Here p is absolutely continuous and bounded and λ is the Lebesgue measure on E^k. The constants C_{kr} are not known for $k > 1$.

Elias (1970) introduces a different aporoach in measuring the performance of a given N-level quantizer. He defines a quantizer that divides the interval $[0,1]$ of a random variable x into a set of N quantizing intervals, of which the i^{th} has length Δx_i. He measures his quantization for particular values of x as the length of the quantizing interval in which x finds itself and measures the quantizer performance by the r^{th} mean value of the quantizer interval length, averaged with respect to the distribution function of the random variable x. That is,

$$M_r(q) = \overline{\Delta x^r}^{\frac{1}{r}} . \qquad (1.24)$$

Work done by other authors describes the quantization error, shown by Equations (1.12) and (1.16), in terms of the absolute value of the difference between x, the random variable being quantized, and some representative point $y(x)$ lying in the quantizing interval and not as the size of the quantizing interval. The performance of the quantizer, as described by these authors, is measured by E_r, the

mean r^{th} power of the difference as given by Equation (1.16). Over the interval $\Delta x = x_{k+1} - x_k$ the error is defined by

$$E_{kr} = \frac{\int_{x_k}^{x_{k+1}} |x - y_k|^r p(x)\, dx}{\int_{x_k}^{x_{k+1}} p(x)\, dx} . \tag{1.25}$$

Let $p(x)$ be approximated by a straight line between x_k and x_{k+1},

$$p(x) \approx p(y_k) + p'(y_k)(x - y_k), \tag{1.26}$$

so

$$E_{kr} \approx \frac{p(y_k) \int_{x_k}^{x_{k+1}} |x - y_k|^r\, dx + p'(y_k) \int_{x_k}^{x_{k+1}} (x - y_k)|x - y|^r}{p(y_k) \int_{x_k}^{x_{k+1}} dx + p'(y_k) \int_{x_k}^{x_{k+1}} (x - y_k)\, dx} . \tag{1.27}$$

Let $\Delta x = x_{k+1} - x_k = \omega$ be constant so $y_k = x_k + \frac{\omega}{2}$; also observe that

$$\int_{x_k}^{x_{k+1}} (x - y_k)|x - y_k|^r\, dx = 0 \tag{1.28}$$

for even distortion measure, and, of course, the second term in the denominator of Equation (1.27) has value zero. So, the straight line

approximation to the density function leads finally to

$$E_{kr} = \frac{[M_r(q)]^r}{2^r(r+1)} \tag{1.29}$$

The two measures related by Equations (1.24) and (1.29), fixed r and $F(x)$, the distribution function associated with x, have approximately the same optimum quantizing intervals. They become better for smooth $p(x)$ and increasing values of N. Elias (1970) works with $M_r(q)$ and it is possible, by imposing smoothness conditions on F and p to extract results about E_{kr} from Equation (1.29). But for arbitrary F the generality and exactness of the results available for $M_r(q)$ seem unlikely to hold for E_{kr}. Comments about $M_r(q)$ and E_{kr} will be made later.

To discuss details of the work conducted by Elias (1970) it is necessary to review some concepts, especially those relating to weighted means as discussed by Hardy, Littlewood and Polya (1934), and those defining what we call asymptotic optimum quantizers. The underlying definition of asymptotic optimum quantizer will be given first. Two of the first authors to discuss them were Panter and Dite (1951), who showed that the minimum mean square error attainable is asymptotic in N to

$$E_2 \sim \frac{C_2}{N^2} \left[\int_{E^1} p^{1/3} \, dx \right]^3 . \tag{1.30}$$

Bluestein (1964) investigated asymptotically optimum quantizers with noisy, continuously varying input signal. His aim was to find such a quantizer that the performance would approach that of the zero memory as the number of levels N is increased. This implies that for any $\epsilon > 0$ the error due to quamtization can be made less than ϵ by making the number of quantizing levels appropriately large. Any quantizer designed to fit with the above is an asymptotically optimum quantizer.

Some definitions from page 12, Hardy, Littlewood and Polya (1934) are in order for subsequent use. Define a set of nonnegative numbers, $a = a_s$, $s = 1, 2, \cdots, n$, and, for convenience, let the summations and the products with respect to s be understood to range from $s = 1$ to $s = n$. Primary definitions are then

$$M_r(a) = \begin{cases} \left[\frac{1}{n} \sum a_s^r \right]^{1/r}, & r > 0, \\ 0 \text{ if } r < 0 \text{ and if at least one of the } a_s \text{ is } 0, \end{cases} \tag{1.31}$$

and

$$G(a) = \left[\prod a_s \right]^{1/n}. \tag{1.32}$$

Clearly, $A(a) = M_1(a)$, $H(a) = M_{-1}(a)$, and $G(a)$ are the ordinary arithmetic, harmonic and geometric means of the set a. The case $r = 0$ is excluded but it can be shown that $M_0(a)$ is interpreted as the geometric mean. A more general system of mean values of the set a

is in terms of a set of weights ω_s , $s = 1, 2, \cdots, n$, $\omega_s > 0$,

$$M_r(a,\omega) = \begin{cases} \left(\dfrac{\sum \omega_s a_s^r}{\sum \omega_s} \right)^{1/r} , & r > 0 , \\ 0 \quad \text{if both } r < 0 \text{ and at least one } a_s \\ \quad \text{is zero,} \end{cases} \tag{1.33}$$

and

$$G(a,\omega) = \left[\prod a_s^{\omega_s} \right]^{1/\sum \omega_s} \tag{1.34}$$

Replace ω_s by p_s with the requirement

$$\sum p_s = 1 \tag{1.35}$$

so that, finally,

$$M_r(a,p) = \begin{cases} \left[\sum p_s a_s^r \right]^{1/r} , & r > 0 , \\ 0 \text{ if both } r < 0 \text{ and at least one } a_s \text{ is 0,} \end{cases} \tag{1.35a}$$

$$A(a,p) = M_1(a,p) = \sum p_s a_s , \tag{1.35b}$$

$$H(a,p) = M_{-1}(a,p) = \left[\sum p_s a_s^{-1} \right]^{-1} , \tag{1.35c}$$

$$G(a,p) = \prod a_s^{p_s} . \tag{1.35d}$$

Additional identities, such as

$$M_r(a,p) = [A(a^r,p)]^{1/r} , \tag{1.37a}$$

and

$$G(a,p) = \exp[A(a,p)\log a] , \tag{1.37b}$$

etc., but the principal interest lies in Equation (1.35a).

Returning to Elias's (1970) work, the discrete set of inputs x_i are mapped into corresponding output symbols y_i , $1 \le i \le N$. He defines his quantizer as $q = \{x_i, p_i\}$ where x as a real number finds itself in the unit interval $[0,1]$ with distribution $p = F(x)$. So $q = \{x_i, p_i\}$ is a set of $(N+1)$ points in the unit square with

$$\begin{aligned} &x_{i-1} \le x_i , \quad p_{i-1} \le p_i , \quad 1 \le i \le N, \\ &x_0 = p_0 = 0 , \quad x_N = p_N = 1 . \end{aligned} \tag{1.38}$$

We call the distribution F compatible with q if the graph of $p = F(x)$ passes through the $(N+1)$ points of q. The $\Delta x_i = x_i - x_{i-1}$ are the quantization levels and $\Delta p_i = p_i - p_{i-1}$ correspond to the probability that x falls into the quantizing interval Δx_i and is thus encoded into y_i. Also,

$$\sum_{i=1}^{N} \Delta x_i = \sum_{i=1}^{N} \Delta p_i = 1 \tag{1.39}$$

and $\Delta x_i + \Delta p_i > 0$ for $i = 1, 2, \cdots , N$.

As mentioned before, the performance of the quantizer is measured

by the r-th mean of the Δx_i. The Δp_i may be treated as the normalized set of weights so the $M_r(a,p)$ of Equation (1.36a) comes into play,

$$M_r(q) = \left[\sum \Delta p_i (\Delta x_i)^r\right]^{1/r} . \tag{1.40}$$

This may be computed from the quantizer itself with no knowledge of of $F(x)$ beyond the one limitation that it has compact support, i.e., x lies within the unit interval, a normalization of the range of x. If x should have a nonfinite range, as in the case of a Gaussian variable, one or two of the quantizing intervals are not finite. For the case of uniform, nonoptimum quantization, $\Delta x_i = \frac{1}{N}$, Equation (1.40) has the form

$$M_r(q) = M_r(\tfrac{1}{N}, \Delta p_i) = \left[\sum \Delta p_i (\tfrac{1}{N})^r\right]^{1/r} = \frac{1}{N} . \tag{1.41}$$

If q_1 represents an optimum quantizer then we see that it can be no worse than in Equation (1.41), or

$$N\, M_r(q_1) \leq 1, \tag{1.42}$$

with equality for all r and N when $F(x) = x$, the uniform quantizer case. To double the number of quantizing intervals reduces the rth mean by half. However, it is usually possible to do better by making the Δx_i small when the Δp_i are large and vice versa. Equation

(1.42) indicates the upper bound on $M_r(q)$. It is possible to derive a lower bound, which is exhibited by Elias (1970) as

$$I = \left[\int_0^1 [f(x)]^p \, dx \right]^{1/q}, \tag{1.43}$$

where $p = \frac{1}{1+r}$, $q = \frac{r}{1+r}$, $f(x) = F'(x)$, the density of the absolutely continuous part of $F(x)$.

Before we go further we note here the definition of $M_r(q_1)$. For given F, N and r, one asks how small the rth mean quantizing interval may be made by adjusting Δx_i and Δp_i. A quantizer q_1, whose rth mean quantizing interval is given by

$$M_r(q_1) = \min_{\{\Delta x_i, \Delta p_i\}} M(\Delta x_i, \Delta p_i), \tag{1.44}$$

subject to conditions stated in Equation (1.39), is defined to be the optimum quantizer for F, N and r. With Equation (1.44) one can for nonnegative r and optimum quantizer q_1, consistent with F, write

$$I_r \leq N M(q_1) \leq 1. \tag{1.45}$$

It should be noted that in the limit $r = 0$ $(q = 0)$ Equation (1.45) bounds the geometric mean $M_0(q_1)$ of the Δx_i as given by Equation (1.35d) while in the limit $r = \infty$ $(q=1)$ Equation (1.45) bounds the maximum value $M_\infty(q_1)$ of the Δx_i. Elias (1970a) does not stop here but goes on to establish the existence of class Q^* quantizers that are asymptotically optimum as $N \to \infty$ and shows that for q Q^*

$$I_r \leq N M_r(q_1) \leq N M_r(q) \leq 1. \tag{1.46}$$

He also proves that for $q \in Q^*$

$$\lim_{N\to\infty} N M(q) = I_r \; . \tag{1.47}$$

After the proof of the existence of this limit he discusses the rate of convergence of the above and shows that if certain conditions of convexity and concavity on F are met then bounds can be found on the rate of convergence. This discussion on the rate of approach of $NM_r(q)$ to I for $q \in Q^*$ is generalized by the consideration of a class C_J of distributions F where the subscript J indicates the composition of F of no fewer than the number J of alternately convex down and convex up pieces. He goes on to show how the convergence is governed by the ratio $n = N/J$, the average number of quantizing intervals per convex domain of F. There is a resulting inequality for $q^* \in Q^*$,

$$I \leq N M(q) \leq I_r^{\frac{n-1}{n}} \exp\left(\frac{1}{q}\,\frac{1+\ln n}{n-1}\right). \tag{1.48}$$

He lists more results which are tighter than Equation (1.48) if f satisfies certain conditions.

Elias (1970a) also derives results concerning the bounds on the optimum quantizer and the convergence of this quantizer, considering a multi-dimensional case. Some of Elias's (1970a) results can be mentioned in terms of Zador's (1964) multi-dimensional results. The measure of performance is the quantization error in the t^{th} coordinate of $x \in R_1$, which is defined as the width $\Delta_i(x_t)$ of R_1 in the t-th coordinate. That is

$$\Delta_i(x_t) = \sup_{x \in R_i} \{x_t\} - \inf_{x \in R_i} \{x_t\}, \quad 1 \leq t \leq k . \tag{1.49}$$

As before, the measure of performance of a quantizer q with respect to the measure ρ is given by $M_r(q)$, the rth mean of errors $\Delta_i(x_t)$, averaging over k coordinates of all R_i. For equal weights and different weights over R_i, $M_r(q)$ in the multi-dimensional cases is given by

$$M_r(q) = \left[\sum_{i=1}^{N} \rho(R_i) \frac{1}{k} \sum_{j=1}^{k} \Delta_j^r(x_t)\right]^{1/r}, \quad 0 < r < \infty . \tag{1.50}$$

A second measure of performance is given by

$$M_r^*(q) = \left[\sum_{i=1}^{N} \rho(R_i)\, \lambda(R_i)^{r/k}\right]^{1/r}, \quad 0 < r < \infty . \tag{1.51}$$

Here $\lambda(R_i)$ is the Lebesgue measure of R_i in the kth dimension and he goes on to show that

$$M_r^*(q) \leq M(q) . \tag{1.52}$$

He has additional results analogous to those of the one-dimensional case which are not mentioned here.

II. MINIMIZING CONDITIONS

A quantization system, or as we say, an N-level quantizer, is described by specifying the endpoints of the N input ranges and the output level y_k, which correspond to each input range. Given the amplitude probability density of the input signal the probability density of the output can be determined as a function of the x_k and y_k. If a symbol D indicates the total distortion in quantization it can be expressed as

$$D = E[f(s_{in} - s_{out})] = \sum_{i=0}^{N} \int_{x_i}^{x_{i+1}} f(x - y_i)\, p(x)\, dx , \qquad (2.1)$$

where x_o and x_{N+1} are arbitrarily large left and right and the error function $f(x)$ is defined on page 6. Indeed in the study of realistic communication problems one is interested in the minimization of D. Max (1960) derives the necessary conditions for the minimum value of D by differentiating Equation (2.1) with respect to the x_j's and the y_j's and setting such derivatives equal to zero:

$$\frac{\partial D}{\partial x_j} = [f(x_j - y_{j-1}) - f(x_j - y_j)]\, p(x_j) = 0, \quad j = 2,3,\cdots,N \qquad (2.2)$$

and

$$\frac{\partial D}{\partial y_j} = -\int_{x_j}^{x_{j+1}} f'(x - y_j)\, p(x)\, dx = 0, \quad j = 1, 2,\cdots,N. \qquad (2.3)$$

These two equations reduce to

$$f(x_j - y_{j-1}) = f(x_j - y_j), \quad j = 2,3,\cdots,N, \qquad (2.4)$$

provided $p(x_j) \neq 0$ and

$$\int_{x_j}^{x_{j+1}} f'(x - y_j)\, p(x)\, dx = 0 , \quad j = 1,2,\cdots\cdots,N. \tag{2.5}$$

As Max (1960) indicates these conditions are also sufficient, a statement proved by Bruce (1965) and Fleischer (1967). Max (1960) goes on to show that for such $f(x) = x^2$ Equations (2.4) and (2.5) reduce to the form

$$x_j = \frac{y_j + y_{j+1}}{2} , \quad j = 2,3,\cdots\cdots,N, \tag{2.6}$$

and

$$\int_{x_{j-1}}^{x_j} (x - y_j)\, p(x)\, dx = 0 , \quad j = 1,2,\cdots\cdots,N. \tag{2.7}$$

Note that Max (1960) is involved with the problem of quantizing magnitude alone. Consider Figure (1.1), which depicts the operations of interest from the initial input to the final output of the quantizer. Thus, from Equation (2.7),

$$y_j = \frac{\int_{x_{j-1}}^{x_j} x\, p(x)\, dx}{\int_{x_{j-1}}^{x_j} p(x)\, dx} , \quad j = 2,3,\cdots\cdots,N. \tag{2.8}$$

If the normal distribution,

$$p(x) = \varphi^{(0)}(x) = \frac{e^{-x^2/2}}{2\pi} , \quad -\infty < x < \infty, \tag{2.9}$$

then the associated functions are defined as

$$\varphi^{(-1)}(x) = \int_0^x \varphi^{(0)}(y)\, dy , \quad -\infty < x < \infty, \tag{2.10}$$

and for values of the index ≥ 1,

$$\varphi^{(n)}(x) = \frac{d^n}{dx^n}\,\varphi^{(0)}(x)\,, \quad -\infty < x < \infty, \tag{2.11}$$

then there is formed a complete set in the usual sense. That is,

$$y_j = \frac{\int_{x_{j-1}}^{x_j} x\,\varphi^{(0)}(x)\,dx}{\int_{x_{j-1}}^{x_j} \varphi^{(0)}(x)\,dx} = \frac{\varphi^{(0)}(x_{j-1}) - \varphi^{(0)}(x_j)}{\varphi^{(-1)}(x_j) - \varphi^{(-1)}(x_{j-1})}\,, \tag{2.12}$$

$$j = 1,2,\cdots,N.$$

Keep in mind that Equation (2.12) yields the quantization levels in terms of amplitude quantization only. Later this result is to be compared with one obtained in terms of quantization in amplitude and time.

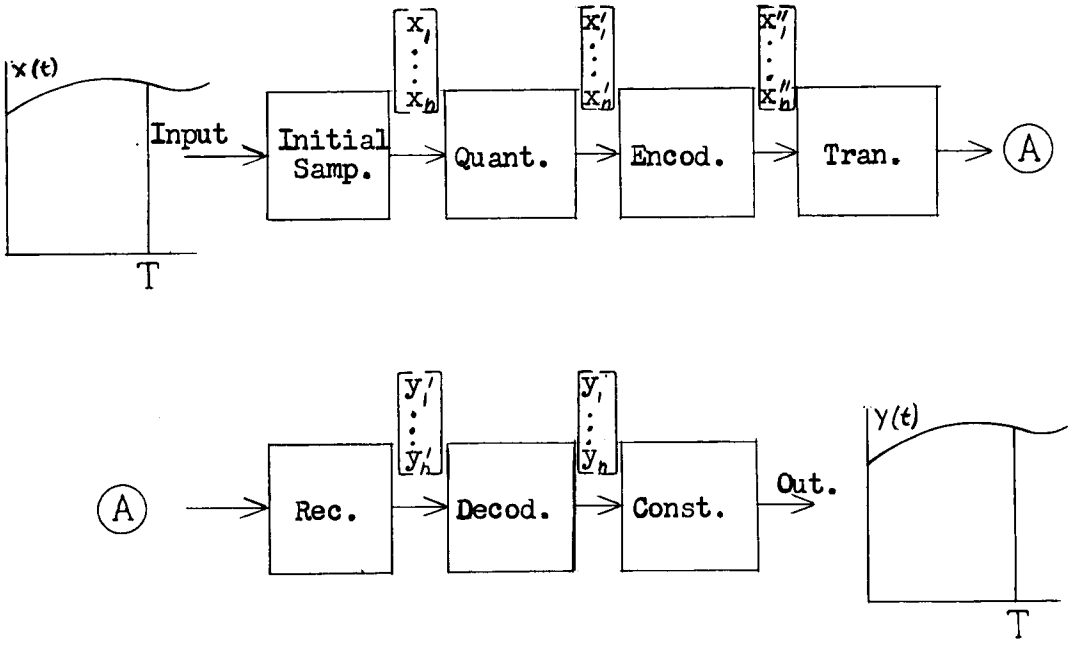

Figure 2.1

This figure shows the stages involved in transmission of data through the channel. It is clear that this involves the operation of encoding, transmission, decoding, etc. These side operations before the final output y_i is obtained involve a certain delay before y_i is reproduced at the output. Denote this amount of delay by tge symbol "d" (see Figure 2.2)

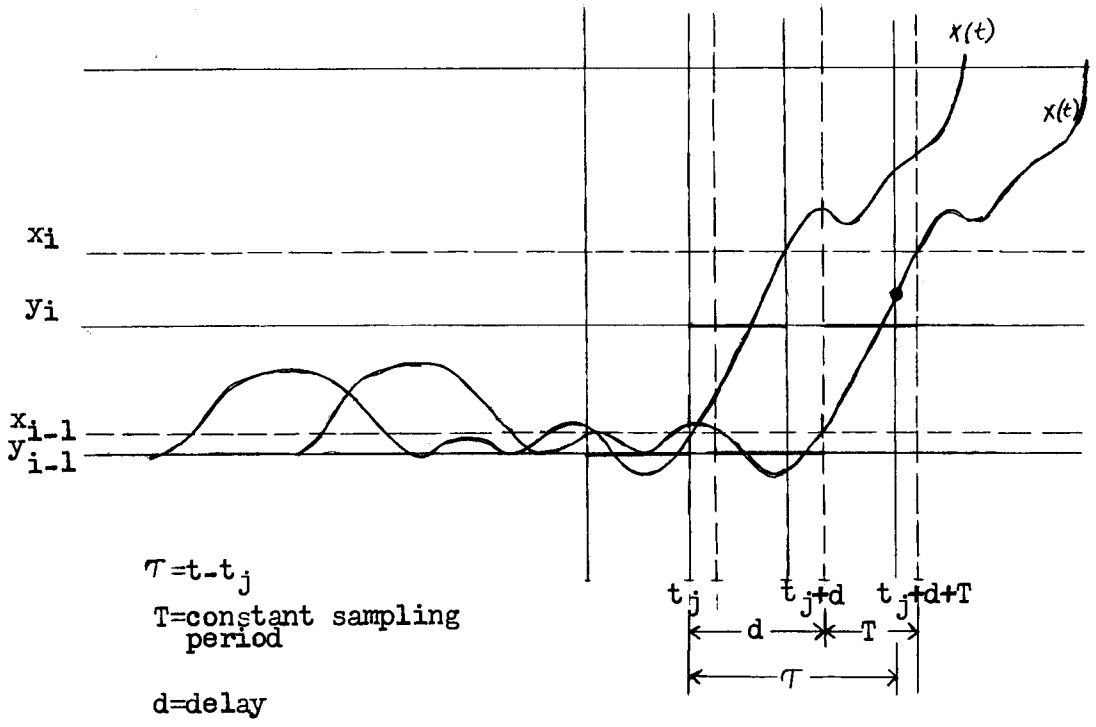

Figure 2.2

In the figure , $x(t)$ indicates the input signal on which N-level quantization is performed for transmission and recovery at the output terminal as $y(t)$. The N-level action calls for breaking up the initial range of $x(t)$ into N intervals (x_{i-1},x_i), $i = 1,2,\cdots,N+1$, with x_o and x_{N+1} arbitrarily large left and right respectively. For $x_{i-1} < x(t_j) < x_i$, $i = 1,2,\cdots,N$, the $x(t_j)$ is obtained by comparison with the endpoints of the ith interval. The quantizer transmits a signal $x_i(t_j)$, where i indicates the level and t_j indicates the instant of time at which $x(t)$ is sampled. This means quantization with respect to both amplitude level and time. The transmitter signal $x_i(t_j)$ is reconstructed at the terminal output to yield y_i , a mapping defined by

$$x_i(t_j) \longrightarrow y_i \ , \quad x_{i-1} < x(t_j) < x_i \ , \ i = 1,2,\cdots,N-1. \tag{2.13}$$

The amount of distortion, or the expected value of the difference between input and output during the ith range on the time scale, is

$$D = E[x(t) - y_i(t)], \tag{2.14}$$

and the mean square distortion in the transmission of $y_i(t)$ is given by

$$d_i = \int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} (x - y_i)^2 \, p[x,x(t_j),\tau \,] \, dx(t_j)dx, \ i = 1,2,\cdot\cdot,N, \tag{2.15}$$

where $\tau = t - t_j$ in the argument of the probability density associated with the input signal. The t is the instant at which the error is analyzed. Averaging Equation (2.15) with respect to all the levels

and the time τ the total distortion--using the mean square criteria--can be expressed as

$$D = \sum_{i=1}^{N} \frac{1}{T} \int_{d}^{d+T} \int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} (x - y_i)^2 p[x, x(t_j), \tau] \, dx(t_j) dx d \quad , \tag{2.16}$$

or as

$$D = \sum_{i=1}^{N} \int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} (x - y_i)^2 \, p^T[x, x(t_j)] \, dx(t_j) dx \quad , \tag{2.17}$$

where

$$p^T[x, x(t_j)] = \frac{1}{T} \int_{d}^{d+T} p[x, x(t_j), \tau] \, d \quad . \tag{2.18}$$

For minimum value of D require that the partial derivatives take zero value,

$$\frac{\partial D}{\partial y_i} = -2 \int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} (x - y_i) \, p^T[x, x(t_j)] \, dx(t_j) dx = 0 \quad , \tag{2.19}$$

and a principal result of this paper is that

$$y_i = \frac{\int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} x \, p^T[x, x(t_j)] \, dx(t_j) dx}{\int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} p^T[x, x(t_j)] \, dx \, (t_j) dx} \quad , \qquad 1 \le i \le N-1. \tag{2.20}$$

Note that Equation (2.17) may be written in the form

$$D = \sum_{i=1}^{N} \left[\int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} x^2 p^T[x, x(t_j)] dx(t_j) dx - y_i^2 \int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} p^T[x, x(t_j) dx(t_j) dx \right]. \tag{2.21}$$

Note that there are two terms in each summation that involve x_i,

i fixed. Thus D_i might be written as

$$D_i = \int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} x^2 p^T[x,x(t_j)]dx(t_j)dx + \int_{-\infty}^{\infty} \int_{x_i}^{x_{i+1}} x^2 p^T[x,x(t_j)]dx(t_j)dx \tag{2.22}$$

$$- y_i^2 \int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} p^T[x,x(t_j)]dx(t_j)dx - y_{i+1}^2 \int_{-\infty}^{\infty} \int_{x_i}^{x_{i+1}} p^T[x,x(t_j)]dx(t_j)dx.$$

Referring to Equation (2.20), the requirement that the partial derivative of D_i with respect to x_i yields

$$2(y_i - y_{i+1}) \int_{-\infty}^{\infty} x\, p^T[x,x_i(t_j)]\, dx = (y_i^2 - y_{i+1}^2) \int_{-\infty}^{\infty} p^T[x,x_i(t_j)]\, dx \tag{2.23}$$

or

$$\frac{y_i + y_{i+1}}{2} = \frac{\int_{-\infty}^{\infty} x\, p^T[x,x_i(t_j)]\, dx}{\int_{-\infty}^{\infty} p^T[x,x_i(t_j)]\, dx}, \tag{2.24}$$

the second important result of this paper.

Another important result is the delay "d" (see Figure 2.2), defined by an extension of Equation (2.17) as

$$D = \sum_{i=1}^{N} \left[\int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} \int_{d}^{d+T} x^2\, p[x,x(t_j),\tau]\, d\tau\, dx(t_j)dx \right] - \sum_{i=1}^{N} \left[y_i^2 \int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} \int_{d}^{d+T} p[x,x(t),\tau]\, d\, dx(t)dx \right]. \tag{2.25}$$

So, a final form goes as

$$\frac{\partial D}{\partial d} = \sum_{i=1}^{N} \int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} (x^2 - y_i^2)\, p[x, x(t_j), d+T] - p[x, x(t_j), d]\, dx(t_j)\, dx. \tag{2.26}$$

To obtain the minimum value of D require that this partial derivative take value zero and this yields

$$p[x, x(t_j), d+T] = p[x, x(t_j), d] \tag{2.27}$$

or

$$|d+T| = |d| , \tag{2.28}$$

and this implies

$$d = -\frac{T}{2} . \tag{2.29}$$

For application let us consider the input signal having a normally distributed amplitude. In terms of the set of functions defined by Equations (2.9), (2.10) and (2.11) above Slepian (1972) has written down the bilinear expansion

$$\begin{aligned} p_{x,y}(\alpha, \beta,) &= \frac{1}{2\pi\sqrt{1-u^2}} \exp\left[-\frac{\alpha^2 + \beta^2 - 2u(\tau)\alpha\beta}{2(1-u^2)}\right] \\ &= \sum_{n=o}^{\infty} \frac{\varphi^{(n)}(\alpha)\, \varphi^{(n)}(\beta)}{n!}\, u^n(\tau), \\ &\qquad -\infty < \alpha, \beta < \infty, \ |u| < 1, \end{aligned} \tag{2.30}$$

where $u(\tau)$ is the correlation function. Therefore

$$\begin{aligned} \int_{x_{i-1}}^{x_i} p^T[x, x(t_j), \ ]\, dx(t_j) &= \sum_{n=o}^{\infty} \int_{x_{i-1}}^{x_i} \frac{\varphi^{(n)}(x)\, \varphi^{(n)}[x(t_j)]}{n}\, u^n dx(t_j) \\ &= \sum_{n=o}^{\infty} \frac{1}{T} \int_{x}^{x_i} \int_{d}^{d+T} \frac{\varphi^{(n)}(x) \varphi^{(n)}[x(t_j)]}{n!}\, u^n d\tau dx(t_j) \end{aligned}$$

or, finally,

$$\int_{x_{i-1}}^{x_i} p^T[x,x(t_j)]\,dx(t_j) = \sum_{n=0}^{\infty} \frac{\varphi^{(n)}(x)[\varphi^{(n)}(x_i) - \varphi^{(n)}(x_{i-1})]}{n!} \cdot \frac{1}{T}\int_d^{d+T} u^n(\tau)d \quad . \tag{2.32}$$

From Equation (2.20)

$$y_i = - \frac{[\varphi^{(0)}(x_i) - \varphi^{(0)}(x_{i-1})]\frac{1}{T}\int_d^{d+T} u(\tau)\,d}{\int_{-\infty}^{\infty}\int_{x_{i-1}}^{x_i} p^T[x,x(t_j)]\,dx} \quad , \tag{2.33}$$

since

$$\int_{-\infty}^{\infty} x\,\varphi^{(n)}(x)\,dx = \begin{cases} -1, & n = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{2.34}$$

By a similar argument

$$\int_{x_{i-1}}^{x_i} p^T[x,x(t_j)]dx(t_j) = \sum_{n=0}^{\infty} \frac{1}{n!}\int_{x_{i-1}}^{x_i} \varphi^{(n)}(x)\varphi^{(n)}[x(t_j)]\,dx(t_j) \cdot \frac{1}{T}\int_d^{d+T} u^n(\tau)\,d \tag{2.35}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!}\varphi^{(n)}(x)[\varphi^{(n-1)}(x_i) - \varphi^{(n-1)}(x_{i-1})]\frac{1}{T}\int_d^{d+T} u^n(\tau)d \quad ,$$

so

$$\int_{-\infty}^{\infty}\int_{x_{i-1}}^{x_i} p\,[x,x(t_j)]dx(t_j)dx = \sum_{n=0}^{\infty} \frac{1}{n!}\int_{-\infty}^{\infty} \varphi^{(n)}(x)[\varphi^{(n-1)}(x_i) - \varphi^{(n-1)}(x_{i-1})]\,\frac{1}{T}\int_d^{d+T} u^n(\tau)dxd \quad . \tag{2.36}$$

Since

$$\int_{-\infty}^{\infty} \varphi^{(n)}(x)\, dx = \begin{cases} 1, & n = 0, \\ 0 & \text{otherwise}, \end{cases} \tag{2.37}$$

Equation (2.36) reduces to

$$\int_{-\infty}^{\infty} \int_{x_{i-1}}^{x_i} p_T[x, x(t_j)]\, dx(t_j) dx = \varphi^{(-1)}(x_i) - \varphi^{(-1)}(x_{i-1}) \tag{2.38}$$

and, finally,

$$y_i = - \frac{\varphi^{(0)}(x_i) - \varphi^{(0)}(x_{i-1})}{\varphi^{(-1)}(x_i) - \varphi^{(-1)}(x_{i-1})} \frac{1}{T} \int_d^{d+T} u(\tau)\, d . \tag{2.39}$$

Compare Equation (2.39) with Equation (2.8). Note that Equation (2.39) involves quantization in amplitude and time while Equation (2.8) implies quantization in amplitude only with no delay present.

Equation (2.24) implies that for the usual expression of the density function,

$$\begin{aligned} \int_{-\infty}^{\infty} x\, p[x, x(t_j), \tau]\, dx &= \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} x\, \frac{\varphi^{(n)}(x)\, \varphi^{(n)}[x(t_j)]}{n!} u^n(\tau)\, dx \\ &= -\varphi^{(1)}[x(t_j)]\, \frac{1}{T} \int_d^{d+T} u(\tau)\, d\tau . \end{aligned} \tag{2.40}$$

Also

$$\int_{-\infty}^{\infty} p_T[x, x(t_j)]\, dx = \varphi^{(0)}[x(t_j)] \tag{2.41}$$

and

$$\frac{y_i + y_{i+1}}{2} = - \frac{\varphi^{(1)}(x_i)}{\varphi^{(0)}(x_i)} \frac{1}{T} \int_d^{d+T} u(\tau)\, d = x_i \frac{1}{T} \int_d^{d+T} u(\tau)\, d\tau , \tag{2.42}$$

or, finally,

$$x_i = \frac{y_i + y_{i+1}}{\frac{2}{T}\int_d^{d+T} u(\tau)d} = \frac{y_i + y_{i+1}}{2u^T(T)} . \qquad (2.43)$$

III. EXTENSION OF MEHLER AND CARLITZ FORMULAS

Much recent progress has been made with the Mehler formula in terms of the bilinear generating functions for the Hermite polynomial set, usually written as

$$\exp[2xt - t^2] = \sum_{n=0}^{\infty} \frac{H_n(x)}{n!} t^n , \quad |t| < 1 . \tag{3.1}$$

Slepian (1972), in a very recent paper, rewrites this in a form of considerable interest to probabilists and statistians,

$$p_{x,y}(\alpha,\beta) = \frac{1}{2\pi\sqrt{1-u^2}} \exp[-\frac{\alpha^2+\beta^2-2u\alpha\beta}{2(1-u^2)}] = \sum_{n=0}^{\infty} \frac{\varphi^{(n)}(\alpha)\varphi^{(n)}(\beta)}{n!} u^n, \tag{3.2}$$

$$|u| < 1,$$

where the set of derivatives of the well-known normal density function have been defined in Equations (2.9), (2.10) and (2.11). They have been well-tabulated by the Harvard Computation Laboratory (1952). Carlitz (1970) and Srivastava and Singhal (1972) extended the theory to the trivariate case. The formula due to the latter authors goes as

$$\sum_{m,n,p=0}^{\infty} H_{n+p}(x)H_{p+m}(y)H_{m+n}(z) \frac{u^m}{m!} \frac{v^n}{n!} \frac{w^p}{p!} \tag{3.3}$$

$$= D^{-\frac{1}{2}} \exp[\sum x^2 - \frac{1}{D}(\sum x^2 - 4\sum u^2x^2 - 4\sum wxy + 8\sum uvxy)] ,$$

where

$$D = 1 - 4u^2 - 4v^2 - 4w^2 + 16uvw \tag{3.4}$$

and $\sum x^2$, $\sum u^2x^2$, $\sum wxy$, and $\sum uvxy$ are symmetric functions in the indicated variables.

The principal purpose of this chapter is to point out that these recent and interesting extensions of the Mehler formula for the Hermite polynomials lack applicability to interesting and real world problems, particularly in the sense of the well-known state of the art paper by Tukey (196). So, consider the transform mate approach as emphasized by Cramer (1946). The moment generating or characteristic function $M_x(s)$ and the probability density function $p_x(\alpha)$ are Fourier transform mates in vector terms. That is, if n Gaussian variates (0,1) are under discussion

$$p_x(\alpha) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \cdots \int e^{is^t\alpha - \frac{1}{2} s^t \Lambda^{-1} s} \, ds_1 \cdots ds_n \, , \tag{3.5}$$

where

$$M_x(s) = e^{-\frac{1}{2} s^t \Lambda s} \, , \tag{3.6}$$

and the variance-covariance or moment matrix

$$\Lambda = [E(x_r x_s)] = [u_{rs}] \, , \; r,s = 1,2,\cdots,n \tag{3.7}$$

and the determinent of the moment matrix is D. These transform mates start out simply but there is an explosive build-up. Thus

$$n = 1, \quad M_x(s) = e^{-\frac{1}{2} s^t I s} \, , \qquad p_x(\alpha) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{is\alpha - \frac{1}{2} s^2} ds$$
$$= \frac{e^{-\frac{\alpha^2}{2}}}{2\pi} \, , \quad -\infty < \alpha < \infty , \tag{3.8}$$

provided a simple shift of the line of integration is carried out.

$$n = 2, \quad \Lambda = \begin{bmatrix} 1 & u \\ u & 1 \end{bmatrix} \qquad P_x(\alpha) = \frac{1}{2\pi\sqrt{1-u^2}} \exp\left[-\frac{\alpha^2+\beta^2-2u\alpha\beta}{2(1-u^2)}\right] , \tag{3.9}$$
$$D = 1 - u^2$$

the same form as in Equation (3.2). The case for $n = 3$ is already quite complex; because of the symmetry moment matrix and determinant

$$\begin{bmatrix} 1 & u_{12} & u_{13} \\ u_{12} & 1 & u_{23} \\ u_{13} & u_{23} & 1 \end{bmatrix} \quad \text{and} \quad D = 1 - u_{12}^2 - u_{13}^2 - u_{23}^2 + 2u_{12}u_{13}u_{23} \tag{3.10}$$

and let the third order density be written in the form

$$\frac{e^{-\frac{1}{2D}\alpha^t \begin{bmatrix} 1-u_{23}^2 & u_{12}u_{13}-u_{23} & u_{12}u_{23}-u_{13} \\ u_{12}u_{13}-u_{23} & 1-u_{13}^2 & u_{23}u_{13}-u_{12} \\ u_{12}u_{23}-u_{13} & u_{23}u_{13}-u_{12} & 1-u_{12}^2 \end{bmatrix} \alpha}}{(2\pi)^{\frac{3}{2}} D^{\frac{1}{2}}} . \tag{3.11}$$

The case $n = 4$ is too complicated to write completely in this form but at least

$$\Lambda = \begin{bmatrix} 1 & u_{12} & u_{13} & u_{14} \\ u_{12} & 1 & u_{23} & u_{24} \\ u_{13} & u_{23} & 1 & u_{34} \\ u_{14} & u_{24} & u_{34} & 1 \end{bmatrix} \tag{3.12}$$

and

$$\begin{aligned} D = 1 &- u_{12}^2 - u_{13}^2 - u_{14}^2 - u_{23}^2 - u_{24}^2 - u_{34}^2 \\ &+ u_{12}^2 u_{34}^2 + u_{13}^2 u_{24}^2 + u_{14}^2 u_{23}^2 \\ &+ 2[u_{12}u_{13}u_{23} + u_{12}u_{14}u_{24} + u_{13}u_{14}u_{34} + u_{23}u_{24}u_{34}] \\ &- 2[u_{12}u_{13}u_{24}u_{34} + u_{12}u_{14}u_{23}u_{34} + u_{13}u_{14}u_{23}u_{24}] . \end{aligned} \tag{3.13}$$

Note that if all u's with 4 in the subscript were set equal to zero this would reduce to the D for the case $n = 3$. A piecemeal

construction of the density function has been made but is relegated to the appendix.

The alternative approach is to arrange that each probability density of increasing order be representable by a sum of products of nth order derivatives of the Gaussian density, each factor being a function of one variate only. Since the moment generating function is constructed directly from the variance-covariance matrix Equation (3.5) is pertinent but the decomposition feature just mentioned requires that if

$$\varphi^{(0)}(x) = \frac{e^{-x^2/2}}{2\pi} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{isx - \frac{s^2}{2}}\, ds \tag{3.14}$$

then

$$\frac{d^n}{dx^n}\varphi^{(0)}(x) = \varphi^{(n)}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} (is)^n e^{isx - \frac{s^2}{2}}\, ds. \tag{3.15}$$

The Mehler type expansion of the bivariate density function in Equation (3.2) is determined from

$$\begin{aligned} P_{x,y}(\alpha,\beta) &= \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int e^{i(s\alpha + t\beta) - \frac{s^2+t^2+2ust}{2}}\, dsdt \\ &= \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int e^{i(s\alpha + t\beta) - \frac{s^2+t^2}{2}} \sum_{n=0}^{\infty} \frac{(i^2ust)^n}{n!}\, dsdt \\ &= \sum_{n=0}^{\infty} \frac{\varphi^{(n)}(\alpha)\varphi^{(n)}(\beta)}{n!}\, u^n, \qquad |u| < 1. \end{aligned} \tag{3.16}$$

As established in the next chapter this form is useful in resolving some interesting questions, particularly in the matter of quantization over a finite range.

The Mehler expansion for order 3 goes as

$$p_{x,y,z}(\alpha,\beta,\gamma) \tag{3.17}$$

$$= \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} \int e^{i(r\alpha+s\beta+t\gamma)-\frac{r^2+s^2+t^2}{2}-(u_{12}\alpha\beta+u_{23}\beta\gamma+u_{13}\alpha\gamma)} drdsdt$$

$$= \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} \int e^{i(r\alpha+s\beta+t\gamma)-\frac{r^2+s^2+t^2}{2}} \sum_{n=o}^{\infty} i^{2n}(u_{12}rs+u_{23}st+u_{13}rt)^n \cdot \frac{drdsdt}{n!} .$$

One may further simplify the notation if the order of factors in each term is preserved such that

$$\varphi^i\varphi^j\varphi^k = \varphi^{(i)}(\alpha)\varphi^{(j)}(\beta)\varphi^{(k)}(\gamma) . \tag{3.18}$$

A few terms written out go as

$$\begin{aligned} p_{x,y,z}(\alpha,\beta,\gamma) &= \varphi^o\varphi^o\varphi^o + u_{12}\varphi^1\varphi^1\varphi^o + u_{23}\varphi^o\varphi^1\varphi^1 + u_{13}\varphi^1\varphi^o\varphi^1 \\ &+ \frac{1}{2!}[u_{12}^2\varphi^2\varphi^2\varphi^o + u_{23}^2\varphi^o\varphi^2\varphi^2 + u_{13}^2\varphi^2\varphi^o\varphi^2 \\ &+ 2u_{12}u_{23}\varphi^1\varphi^2\varphi^1 + 2u_{23}u_{31}\varphi^1\varphi^1\varphi^2 + 2u_{12}u_{13}\varphi^2\varphi^1\varphi^1] \\ &+ \frac{1}{3!}[u_{13}^3\varphi^3\varphi^3\varphi^o + u_{23}^3\varphi^o\varphi^3\varphi^3 + u_{13}^3\varphi^3\varphi^o\varphi^3 + 3u_{12}^2(u_{23}\varphi^2\varphi^3\varphi^1 + u_{31}\varphi^3\varphi^2\varphi^1) \\ &+ 3u_{23}^2(u_{13}\varphi^1\varphi^2\varphi^3 + u_{12}\varphi^1\varphi^3\varphi^2) + 3u_{13}^2(u_{12}\varphi^3\varphi^1\varphi^2 + u_{23}\varphi^2\varphi^1\varphi^3) \\ &+ 6u_{12}^2u_{23}^2u_{31}^2\varphi^2\varphi^2\varphi^2] \\ &+ \frac{1}{4!}[\cdots\cdots] + \cdots\cdots\cdots \end{aligned} \tag{3.19}$$

Note that if u_{13} and $u_{23} = 0$ and $u_{12} = u$ (equivalent to reducing z clear out of consideration) then Equation (3.19) reduces to

Equation (3.16) A more compact form for the third order density is

$$p_{x,y,z}(\alpha,\beta,\gamma) \tag{3.20}$$
$$= \sum_{n=o}^{\infty} \frac{1}{n!} \sum_{i,j,k=o}^{n} \binom{n}{i,j,k} u_{12}^{i} u_{23}^{j} u_{13}^{k} \varphi^{(k+i)}_{(\alpha)} \varphi^{(i+j)}_{(\beta)} \varphi^{(j+k)}_{(\gamma)},$$

where the multinomial is defined as

$$\binom{n}{i,j,k} = \frac{n!}{i!j!k!} \quad , \; i+j+k = n \; . \tag{3.21}$$

A remark similar to the one above is true for the case of removing z from consideration.

To paraphrase David Hilbert, the important step seems to be from order 2 to order 3, hence it is tempting to write down the 4th order case after making a remark or two . For third order case there are $\binom{3}{2}$ correlation functions, for the fourth order case there are $\binom{4}{2}$ correlation functions. Hence the conjecture goes as

$$p_{x,y,z,w}(\alpha,\beta,\gamma,\delta) = \sum_{p=o}^{\infty} \frac{1}{p!} \sum_{i,j,k,l,m,n=O}^{p} u_{12}^{i} u_{13}^{j} u_{14}^{k} u_{23}^{l} u_{24}^{m} u_{34}^{n} \tag{3.22}$$

$$\binom{p}{i,j,k,l,m,n} \varphi^{(i+j+k)}_{(\alpha)} \varphi^{(i+l+m)}_{(\beta)} \varphi^{(j+l+n)}_{(\gamma)} \varphi^{(k+m+n)}_{(\delta)}$$

where

$$\binom{p}{i,j,l,m,n} = \frac{p!}{i!j!k!l!m!n!} \quad , \; i+j+k+l+m+n = p. \tag{3.23}$$

IV. AUTOCORRELATION

Consider the mean square error criterion as defined by Max (1960). Distortion is expressed by

$$\sigma^2 = \sum_{k=1}^{N} \int_{x_{k-1}}^{x_k} (x - y_{k-1})^2 p(x)\, dx, \tag{4.1}$$

where E_2 has been replaced by σ^2 for notational convenience.

Let

$$p_{k-1} = \int_{x_{k-1}}^{x_k} p(x)\, dx\ , \tag{4.2}$$

then

$$y_{k-1} = \frac{1}{p_{k-1}} \int_{x_{k-1}}^{x_k} x\, p(x)\, dx \tag{4.3}$$

and

$$x_k = \frac{y_k + y_{k-1}}{2} \tag{4.4}$$

(see Equation (2.24)). Then Equation (4.1) may be written as

$$\sigma^2 = \sigma_x^2 - 2 \sum_{k=1}^{N} y_{k-1} \int_{x_{k-1}}^{x_k} x\, p(x)\, dx + \sum_{k=1}^{N} y_{k-1}^2 p_k, \tag{4.5}$$

where

$$\sigma_x^2 = \int_{-\infty}^{\infty} x^2 p(x)\, dx \tag{4.6}$$

is the variance of the original signal. From the second of Max's (1960) conditions Equation (4.5) can be written

$$\sigma^2 = \sigma_x^2 - \sum_{k=1}^{N} y_{k-1}^2 p_{k-1}. \tag{4.7}$$

Note that p_{k-1} can be written in an expanded form by assuming a linear relation between the points x_{k-1} and x_k so Equation (4.7) becomes

$$\sigma^2 - \sigma_x^2 = \sum_{k=1}^{N} y_{k-1}^2 [p(y_{k-1}) + (x - y_{k-1})p'(y_{k-1})] . \tag{4.8}$$

If the derivative of the probability function at y_{k-1} may be considered to be negligible

$$\sigma_x^2 - \sigma^2 = \sum_{k=1}^{N} y_{k-1}^2 p(y_{k-1}), \tag{4.9}$$

where, as observed above, σ^2 is the variance of the quantized signal. Wood (1969) expressed this result as

$$\sigma_x^2 - \sigma^2 = \frac{1}{12} \sum_{k=1}^{N} \Delta_{k-1}^2 \, p(x_{k-1}), \tag{4.10}$$

where

$$\Delta_{k-1} = x_k - x_{k-1} . \tag{4.11}$$

Note that Equations (4.9) and (4.10) express the difference in the variances in terms of nonuniform quantization, different from the uniform case in which we can assume

$$\Delta_{k-1} = \omega, \text{ a constant, } k = 1,2,\cdots,N, \tag{4.12}$$

and the result corresponding to Equation (4.10) is

$$\sigma_x^2 - \sigma^2 = \frac{\omega^2}{12} , \tag{4.13}$$

which is the "Sheppard's Correction", repeatedly proved by many authors using different approaches. One such approach is that of

Banta (1965) who considers a general theory of the autocorrelation of quantizer output of a certain signal. He shows how the autocorrelation error can be isolated and reduced to

$$R_{\epsilon}(0) = \frac{\omega^2}{12}, \tag{4.14}$$

where $R_{\epsilon}(0)$ in turn is derived from the definition of the autocorrelation function

$$R_f(\tau) = E[f(t)f(t+\tau)], \tag{4.15}$$

where $f(t)$ is random. Naturally

$$R_f(0) = E[f^2(t)], \tag{4.16}$$

the variance of $f(t)$ in the usual statistical terms. A flow diagram of Banta's (1965) analysis is depicted in Figure (4.9).

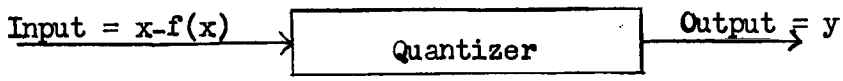

Figure 4.1

Herewith the Fourier Series representation of the input quantized signal, uniform quantization,

$$f(x) = \sum_{k=1}^{\infty} \frac{(-1)^k \omega}{k\pi} \sin \frac{2\pi k x}{\omega}. \tag{4.17}$$

The input signal is contaminated by noise,

$$x(t) = m(t) + n(t), \tag{4.18}$$

and signal and noise are assumed to be statistically independent. The input-output quantizer and the noise characteristics are plotted in Figures (4.2) and (4.3) respectively.

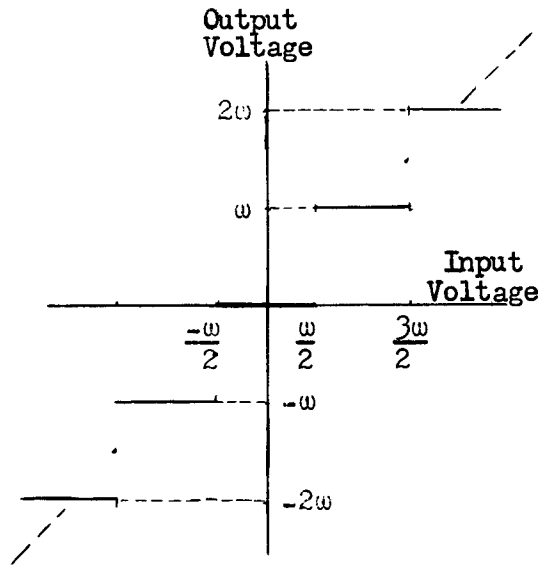

Figure 4.2

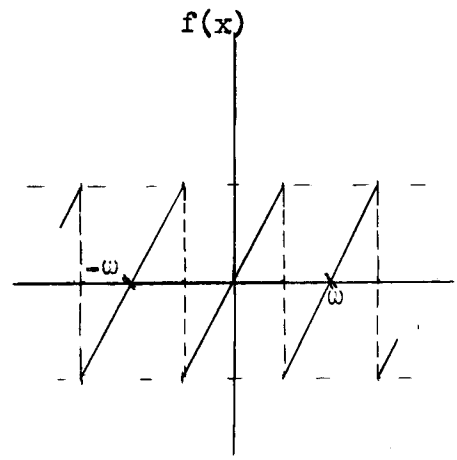

Figure 4.3

Considering a determinant signal and Gaussian noise input Banta (1965) derived a generalized expression for the autocorrelation function of the quantized output

$$R_o(\tau) = R_{mm}(\tau) + R_{nn}(\tau) + R_{\epsilon}(\tau) \; ; \qquad (4.19)$$

the terms on the right represent the separate autocorrelation functions of the signal, the noise and the quantizer effect respectively. Certain assumptions on the magnitude of the signal voltage with respect to RMS noise voltage were made to derive Equation (4.19). He established Equation (4.14) again, and $R_{\epsilon}(0) = \omega^2/12$ can be described as the "quantizing power". Some people who have worked in this field are Rice (1945), Robin (1952), Bennett (1956), Widrow (1956) Baum (1957), Trofimov (1958), Kosygin (1961), Velechkin (1962) and Hurd (1967), to name a few.

In order to clarify what we intend to discuss in this chapter we make some remarks on what each author has done. Bennett (1956) was the first to derive the expression for the autocorrelation function for the quantizer power and this was also obtained by Kosygin (1961), using a method of characteristic functions. Widrow (1956) introduced the characteristic function theory in analyzing the quantization process, later developed by Kosygin (1961). Hurd (1967) worked with the autocorrelation of the output where the input to the quantizer is the sum of a sine wave and zero mean, stationary Gaussian noise. Price (1958) laid down his theory for the autocorrelation function of the output for strictly stationary Gaussian inputs and derived a relationship expressing the partial derivatives of the output autocorrelation in terms of the input correlation coefficients. Several authors later improved on his basic theorem and he himself checked out earlier results with his method, even though it involved differential equations with almost impossible boundary conditions. Extension of Equation (4.15) to n random variables,

$$R(\tau) = E \prod_{i=1}^{n} f_i(x_i), \tag{4.20}$$

$f_i(x)$ the n zero-memory nonlinear devices specifying the input-output relationships, allowed Price (1958) to consider the clipper, the hard limiter and the smooth limiter. A unified approach to problems of this type is to resort to the bilinear expansion of the joint Gaussian density function in terms of the derivatives of the error function, as given by Equation (3.2). The immediate question is about the $u(\tau)$

function. Consider a linear filter K(f) to be opened over the frequency axis so that the normalized autocorrelation u( ), associated with sampled values of signal and noise input, is sufficient to determine the joint density function. The process is described in Figures (4.4) and (4.5):

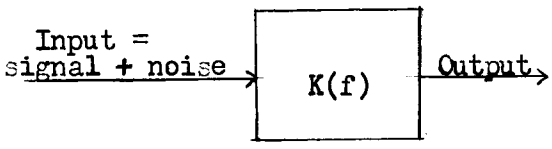

Figure 4.4

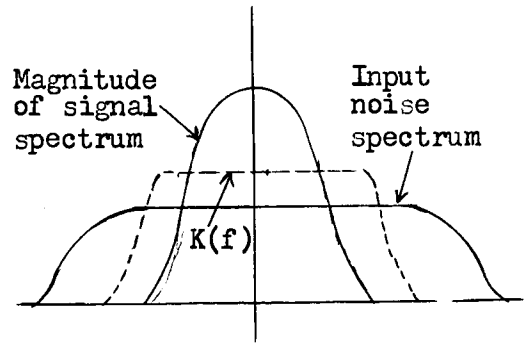

Figure 4.5

If ω_1 is the half power bandwidth of the input noise power spectrum, ω_s is the half power bandwidth of the signal spectrum, and A is defined by

$$A - 1 = \frac{\psi_s}{\psi_o} , \tag{4.21}$$

the ratio of input noise power level to input signal power level, then

$$u(\tau,\delta) = \frac{(\delta^2 - A^2)e^{-\omega_1} + \delta(A^2 - 1)e^{-\omega_s}}{(\delta - 1)(\delta + A^2)} , \tag{4.22}$$

where $\delta = \omega_1/w_s > 1$. If there is no distributed signal power present then for **A** = 1 the $u(\tau,\delta)$ quickly reduces to

$$u(\tau) = e^{-\omega_1} \quad . \tag{4.23}$$

Let **x** and **y** be two normally distributed (0,1) variates; the joint density is that of Equation (3.2) with either of the two designated values of $u(\tau)$.

An important special case is to set $u(\tau) = 1$. That is,

$$\lim_{u\to 1_-} p_{x,y}(\alpha.\beta) = \delta(\alpha-\beta)\ \varphi^{(0)}(\alpha), \tag{4.24}$$

where $\varphi^{(0)}(\alpha)$ is defined by Equation (2.9). An interesting effect of the Dirac delta function is developed by Papoulis (1962).

$$\lim_{c\to 0} \frac{1}{c} \int_0^{\infty} \varphi^{(0)}\left(\frac{\alpha}{c}\right) \varphi^{(2n)}(\alpha)d\alpha = \int_0^{\infty} \delta(\alpha)\ \varphi^{(2n)}(\alpha)\ d\alpha$$
$$= \frac{1}{2}\varphi^{(2n)}(0) = \frac{(-1)^n(2n)!}{\sqrt{2\pi}\ 2^{n+1}n!} \quad . \tag{4.25}$$

Consider a finite form of the input-output relationship of Banta (1965):

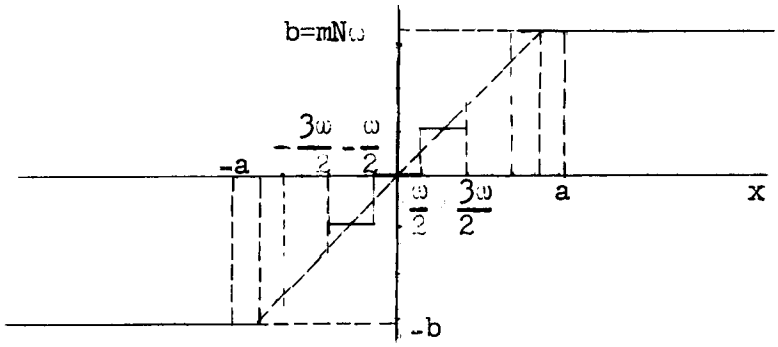

Figure 4.6

Quantization takes place into 2N-1 different levels, equally spaced, if input voltage lies in the range $-a(1-\frac{1}{2N}) \leq v_i(t) < a(1-\frac{1}{2N})$; if if input lies outside this range the output is taken to be either ma or -ma, as the case may be. Let the interval be defined as

$$\omega = \frac{a}{N} \quad ; \tag{4.26}$$

then the input-output relationship may be defined in terms of the greatest integer notation as

$$f_x(\alpha) = \begin{cases} b\ , & (N-\frac{1}{2})\omega \leq \alpha\ , \\ m\omega[\frac{\alpha}{\omega} + \frac{1}{2}]\ , & -(N-\frac{1}{2})\omega \leq \alpha < (N-\frac{1}{2})\omega, \\ -b\ , & \alpha < -(N-\frac{1}{2}), \end{cases} \tag{4.27}$$

$$m = \frac{b}{a}\ ,\ \omega = \frac{a}{N}\ .$$

This staircase function of finite scope meets the conditions outlined by Price (1958), so autocorrelation of output has the form

$$R(u,a,\omega) = \int_{-\infty}^{\infty} f_x(\alpha)\, f_y(\beta) \sum_{n=o}^{\infty} \frac{\varphi^{(n)}(\alpha)\, \varphi^{(n)}(\beta)}{n!}\, u^n(\tau)\, d\alpha d\beta \tag{4.28}$$

Only the odd order terms of the resulting series are not zero, a point which is established by the identities

$$\varphi^{(n)}(\alpha) - \varphi^{(n)}(-\alpha) = \begin{cases} 2\varphi^{(n)}(\alpha), & n \text{ odd}, \\ 0, & n \text{ even}\ . \end{cases} \tag{4.29}$$

So

$$R(u,a,\omega) = 4m^2\omega^2 \sum_{n=o}^{\infty} \frac{u^{2n+1}}{(2n+1)!} \Bigg\{ \sum_{k=1}^{N-1} \sum_{j=1}^{N-1} jk \int_{(j-\frac{1}{2})\omega}^{(j+\frac{1}{2})\omega} \varphi^{(2n+1)}(\alpha)\, d\alpha \int_{(j-\frac{1}{2})\omega}^{(k+\frac{1}{2})\omega} \varphi^{(2n+1)}(\beta)\, d\beta \tag{4.30}$$

$$+ 2N \sum_{j=1}^{N-1} j \int_{(j-\frac{1}{2})\omega}^{(j+\frac{1}{2})\omega} \varphi^{(2n+1)}(\alpha)\, d\alpha \int_{(N-\frac{1}{2})\omega}^{\infty} \varphi^{(2n+1)}(\beta)\, d\beta + N \int_{(N-\frac{1}{2})\omega}^{\infty} \varphi^{(2n+1)}(\alpha)\, d\alpha \int_{(N-\frac{1}{2})\omega}^{\infty} \varphi^{(2n+1)}(\beta)\, d\beta \Bigg\}$$

$$= 4m^2\omega^2 \sum_{n=o}^{\infty} \frac{u^{2n+1}}{(2n+1)!} \left\{ \sum_{j=1}^{N-1} j[\varphi^{(2n)}[(j+\tfrac{1}{2})\omega] - \varphi^{(2n)}[(j-\tfrac{1}{2})\omega]] - N\,\varphi^{(2n)}[(N-\tfrac{1}{2})\omega] \right\}^2$$

$$= 4m^2\omega^2 \sum_{n=o}^{\infty} \frac{u^{2n+1}}{(2n+1)!} \left\{ \sum_{j=1}^{N} \varphi^{(2n)}[(j-\tfrac{1}{2})\omega] \right\}^2, |u| \leq 1 .$$

The extreme value $u(\tau) = 1$ is of value as $R(1,a,\omega)$ is an elementary function. From Equation (4.24)

$$\lim_{u\to 1_-} R(u,a,\omega) = \int_{-\infty}^{\infty} f_x^2(\alpha)\, \varphi^{(0)}(\alpha)\, d\alpha \qquad (4.31)$$

$$= 2m^2\omega^2 \sum_{j=1}^{N-1} j^2 \int_{(j-\frac{1}{2})\omega}^{(j+\frac{1}{2})\omega} \varphi^{(0)}(\alpha)\, d\alpha + N^2 \int_{(N-\frac{1}{2})\omega}^{\infty} \varphi^{(0)}(\alpha)\, d\alpha$$

$$= 2m^2\omega^2 \sum_{j=1}^{N-1} j^2 [\varphi^{(-1)}[(j+\tfrac{1}{2})\omega] - \varphi^{(-1)}[(j-\tfrac{1}{2})]] + N^2[\tfrac{1}{2} - \varphi^{(-1)}[(N-\tfrac{1}{2})\omega]] .$$

$$= m^2[a^2 - 4\omega^2 \sum_{j=1}^{N} (j-\tfrac{1}{2})\, \varphi^{(-1)}[(j-\tfrac{1}{2})\omega]].$$

A smooth quantizer in the sense of Baum (1957) would be specified by an input-output relationship

$$f_x(\alpha) = 2b\, \varphi^{(-1)}([\alpha + \tfrac{1}{2}]\tfrac{\omega}{c}) , \qquad (4.32)$$

where the argument implies an integer multiple of $\frac{\omega}{c}$, ω the length of all the subintervals on the entire axis, c an arbitrary parameter controlling the curvature.

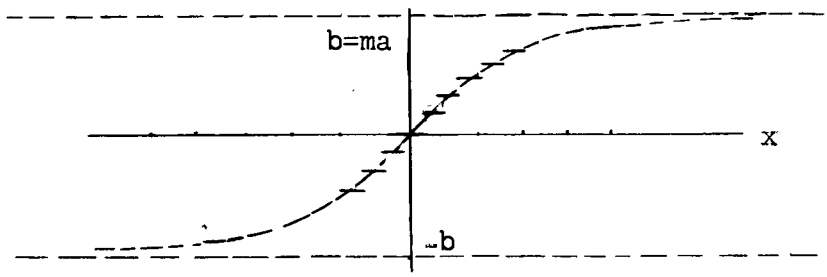

Figure 4.7

In this case Equation (4.28) takes the form

$$R(u,\omega,c) = (4b)^2 \sum_{n=0}^{\infty} \frac{u^{2n+1}}{(2n+1)!} \left\{ \sum_{j=1}^{\infty} \varphi^{(-1)}\left(\frac{j\omega}{c}\right)\left[\varphi^{(2n)}\left[\left(j+\tfrac{1}{2}\right)\omega\right] - \varphi^{(2n)}\left[\left(j-\tfrac{1}{2}\right)\omega\right]\right] \right\}^2 \tag{4.33}$$

$$= (4b)^2 \omega \sum_{n=0}^{\infty} \frac{u^{2n+1}}{(2n+1)!} \left\{ \sum_{j=1}^{\infty} \varphi^{(-1)}\left(\frac{j\omega}{c}\right) \sum_{k=0}^{\infty} \frac{\left(\frac{\omega}{2}\right)^{2k}}{(2k+1)!} \varphi^{(2n+1+2k)}(j\omega) \right\}^2$$

$$= (4b)^2 \omega^2 \sum_{n=0}^{\infty} \frac{u^{2n+1}}{(2n+1)!} \left\{ \sum_{k=0}^{\infty} \frac{\left(\frac{\omega}{2}\right)^{2k}}{(2k+1)!} \sum_{j=1}^{\infty} \varphi^{(-1)}\left(\frac{j\omega}{c}\right) \varphi^{(2n+1+2k)}(j\omega) \right\}^2 .$$

There is a simple application of the Taylor series formula here. From Equation (4.21)

$$\lim_{u\to 1_-} R(u,\omega,c) = 2(2b)^2 \sum_{j=1}^{\infty} \left[\varphi^{(-1)}\left(\frac{j\omega}{c}\right)\right]^2 \left[\varphi^{(-1)}\left[\left(j+\tfrac{1}{2}\right)\omega\right] - \varphi^{(-1)}\left[\left(j-\tfrac{1}{2}\right)\omega\right]\right] \tag{4.34}$$

$$= 8b^2\omega \sum_{j=1}^{\infty} \left[\varphi^{(-1)}\left(\frac{j\omega}{c}\right)\right]^2 \sum_{k=0}^{\infty} \frac{\left(\frac{\omega}{2}\right)^{2k}}{(2k+1)!} \varphi^{(2k)}(j\omega)$$

$$= 8b^2\omega \sum_{k=0}^{\infty} \frac{\left(\frac{\omega}{2}\right)^{2k}}{(2k+1)!} \sum_{j=1}^{\infty} \left[\varphi^{(-1)}\left(\frac{j\omega}{c}\right)\right]^2 \varphi^{(2k)}(j\omega) .$$

These formulations appear to be of questionable value but simplifications are possible by means of the two Euler-MacLaurin summation formulas. That is, the summation with respect to the index j may

be evaluated in terms of elementary functions. Hildebrand (1956) and Gould and Squire (1963) have discussed the two slightly different forms of the Euler-MacLaurin sum formula, both involving the Bernoulli numbers defined by the identity

$$\frac{1+e^{z}}{1-e^{z}} = 2\sum_{k=0}^{\infty}\frac{B_{2k}}{(2k)!}z^{2k-1}, \quad |z|<2\pi, \tag{4.35}$$

$$B_0 = 1,\ B_2 = \frac{1}{6},\ B_4 = -\frac{1}{30},\ B_6 = \frac{1}{42},\ B_8 = -\frac{1}{30},\ B_{10} = \frac{5}{66}, \cdots .$$

The second form may be written out as (see Equation (5.8.18), Hildebrand (1956))

$$\begin{aligned}\sum_{k=1}^{N} f[k-\tfrac{1}{2})\omega] &= \frac{1}{\omega}\int_0^a f(\alpha)d\alpha - \sum_{k=1}^{\infty}\frac{B_{2k}}{(2k)!}\left(1-\frac{1}{2^{2k-1}}\right)\omega^{2k-1} \\ &\quad\cdot[f^{(2k-1)}(a) - f^{(2k-1)}(0)] \\ &= \frac{1}{\omega}\int_0^a f(\alpha)\,d\alpha - \frac{\omega}{24}[f'(a)-f'(0)] + \frac{7\omega^3}{5760}[f'''(a)-f'''(0)] \\ &\quad - \frac{31\omega^5}{967{,}680}[f^{V}(a)-f^{V}(0)] + \cdots\cdots .\end{aligned} \tag{4.36}$$

This may be applied at once to the finite sums occurring in the study of the quantizer of finite scope. Equation (4.30) becomes

$$\begin{aligned}R(u,a,\omega) = 4m^2\sum_{n=0}^{\infty}\frac{u^{2n+1}}{(2n+1)!}\Bigg\{&\int_0^a \varphi^{(2n)}(\alpha)\,d\alpha - \frac{\omega^2}{24}\varphi^{(2n+1)}(a) \\ &+ \frac{7\omega^4}{5760}\varphi^{(2n+3)}(a) - \cdots\cdots\Bigg\}^2\end{aligned} \tag{4.37}$$

$$= (2m)^2 \sum_{n=0}^{\infty} \frac{u^{2n+1}}{(2n+1)!} \left\{ [\varphi^{(2n-1)}(a)]^2 - \frac{\omega^2}{12} \varphi^{(2n-1)}(a)\varphi^{(2n-)}(a) \right.$$

$$\left. + \frac{\omega^4}{576} [\varphi^{(2n+1)}(a)]^2 + \frac{7\omega^4}{2880} \varphi^{(2n-1)}(a)\, \varphi^{(2n+3)}(a) - \cdots\cdots \right\}.$$

The first term of this series would represent the autocorrelation of output of a finite clipper, discussed by Price (1958) and tabulated by Laning and Battin (1956). Tabulation by this means would seem to be easier and more precise than by the numerical solution of the second order differential equation with singularities. Note that if a is taken to be zero then ω is also, and the hard limiter result is

$$\lim_{a,\omega \to 0} R(u,a,\omega) = \lim_{a \to 0} \left(\frac{2b}{a}\right) \sum_{n=0}^{\infty} \frac{u^{2n+1}}{(2n+1)!} [\varphi^{(2n-1)}(a)]^2 \tag{4.38}$$

$$= 2b^2 \sum_{n=0}^{\infty} \frac{u^{2n+1}}{(2n+1)!} [\varphi^{2n)}(0)]^2 = \frac{(2b)^2}{2\pi} \sum_{n=0}^{\infty} \frac{u^{2n+1}}{(2n+1)!} \frac{[(2n)!]^2}{2^{2n}(n!)^2}$$

$$= \frac{2b^2}{\pi} \sin^{-1}[u(\tau)],$$

as given by Price (1958). If N, the number of steps on each side of zero, is allowed to become arbitrarily large, then the parameter a is allowed to do likewise, the input-output relationship becomes trivial and the autocorrelation of output would be simply $m^2 u$.

In similar fashion Equation (4.31) becomes

$$R(1,a,\omega) = m^2 \left[a^2[1-2\varphi^{(-1)}(a)] + 2\varphi^{(-1)}(a) - 2a\,\varphi^{(0)}(a) \right. \tag{4.39}$$
$$\left. + \frac{\omega^2}{6}[\varphi^{(-1)}(a) + a\,\varphi^{(0)}(a)] - \frac{7\omega^4}{1440}[3\varphi^{(1)}(a) + a\varphi^{(2)}(a)] + \cdots \right].$$

$R(1,a,0)$ was discussed and tabulated by Laning and Battin (1956) with m taken to be unity. It is clear that

$$\lim_{a\to\infty} R(1,a,0) = m^2 . \tag{4.40}$$

For further study of the smooth limiter the first Euler-MacLaurin sum formula (see Equation (5.8.12), Hildebrand (1956)) leads to a series expansion for $R(u,\omega,c)$ in our Equation (4.33) and

$$\lim_{\omega\to 0} R(u,\omega,c) = \frac{2b^2}{\pi} \sin^{-1} \frac{u(\tau)}{1+c^2} , \tag{4.41}$$

the inverse sine function obtained by Baum (1957).

For the different forms of the autocorrelation functions of the output of the several quantizers discussed above we were concerned with uniform quantizers. Initially the domain of definition was broken up into intervals, each of size ω. For the non-uniform case one might set up a method of analysis to encompass non-uniformity, which results from use of the least mean square criterion. This brings in Max's (1960) minimizing conditions.

Consider the second order case and the corresponding autocorrelation function of the output, as expressed by Equations (4.31) and (4.32). The function $f_x(\alpha)$ does not follow the step-wise characteristic as shown in Figure (4.6). The step characteristic will be

entirely dependent upon the minimizing conditions as given by Equations (4.3) and (4.4). Consider a generalized step-wise behavior given by $y = g(x)$, where x takes on discrete values dictated by the conditions of Equations (4.3) and (4.4). Taking input as Gaussian (0,1) the autocorrelation function of output can be written as

$$R(u,a,g) = \int_{-\infty}^{\infty}\int g(\alpha)g(\beta) \sum_{n=0}^{\infty} \frac{\varphi^{(n)}(\alpha)\,\varphi^{(n)}(\beta)}{n!} u^n(\tau)\, d\alpha d\beta \,, \qquad |u(\tau)| < 1, \tag{4.42}$$

or

$$R(u,a,g) = \sum_{n=0}^{\infty} \frac{u^n}{n!} \int_{-\infty}^{\infty} g(\alpha)\, \varphi^{(n)}(\alpha)\, d\alpha \int_{-\infty}^{\infty} g(\beta)\, \varphi^{(n)}(\beta)\, d\beta. \tag{4.43}$$

Integration by parts leads to

$$R(u,a,g) = \sum_{n=0}^{\infty} \frac{u^n}{n!} \int_{-\infty}^{\infty} g'(\alpha)\, \varphi^{(n-1)}(\alpha) d\alpha \int_{-\infty}^{\infty} g'(\beta)\, \varphi^{(n-1)}(\beta)\, d\beta. \tag{4.44}$$

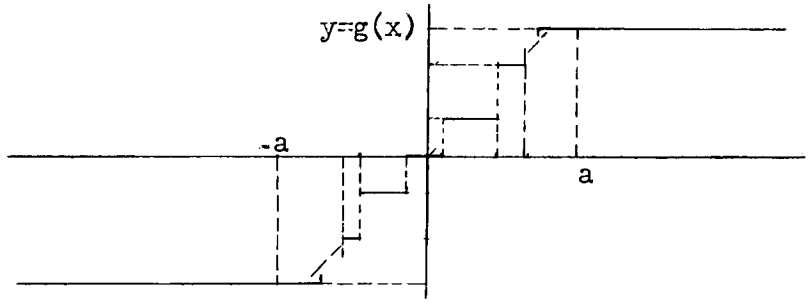

Figure 4.8

In Figure (4.8) the quantizer levels may be described by a relationship

$$y = g(x) = y_o + \sum_{k=1}^{N-1} \int_{-\infty}^{x} \Delta_k\, \delta(s - x_k)\, ds \,, \tag{4.45}$$

where $\Delta_k = y_k - y_{k-1}$ and the Dirac function serves to pinpoint the

jump steps in the quantization procedure. The y_o in Equation (4.45) is the initial quantizer level fixed by choosing x_o. Differentiation of Equation (4.45) yields

$$g'(x) = \sum_{j=1}^{N-1} \Delta_k \, \delta(x - x_k) \tag{4.46}$$

Putting Equation (4.46) into Equation (4.44) leads to

$$\begin{aligned} R(u,a,g) = \sum_{n=0}^{\infty} \frac{u^n}{n!} \Bigg\{ \sum_{k=1}^{N-1} \Delta_k \int_{-\infty}^{\infty} \delta(\alpha - x_k) \, \varphi^{(n-1)}(\alpha) \, d\alpha \\ \cdot \sum_{j=1}^{N-1} \Delta_j \int_{-\infty}^{\infty} \delta(\beta - x_j) \, \varphi^{(n-1)}(\beta) \, d\beta \Bigg\} . \\ = \sum_{n=0}^{\infty} \frac{u^n}{n!} \left\{ \sum_{k=1}^{N-1} \Delta_k \, \varphi^{(n-1)}(x_k) \right\}^2 . \end{aligned} \tag{4.47}$$

Compare Equation (4.47) with Equation (4.30) and note the effect of a non-uniform quantizer on the structure of the resulting equations. If in Equation (4.47) the $u(\ )$ takes value unity for $= 0$ then

$$R(1,a,g) = \sum_{n=0}^{\infty} \frac{1}{n!} \left\{ \sum_{k=1}^{N-1} \Delta_k \, \varphi^{(n-1)}(x_k) \right\}^2 . \tag{4.48}$$

BIBLIOGRAPHY

1. Algazi, V. R. Useful Approximations to Optimum Quantization. IEEE Trans. Communication Technology, vol. COM-14, pp. 297-301, 1966.

2. Banta, E. E. On the autocorrelation function of the quantized signal plus noise. IEEE Trans. Information Theory, vol. IT-11, pp. 114-118. 1965.

3. Baum, R. F. The correlation function of smoothly limited Gaussian noise. IEEE Trans. Information Theory, vol. IT-3, pp. 193-197. 1957.

4. Bennett, W. R. Spectra of Quantized signals. Bell Sys. Tech. J., vol. 27, pp. 446-472. 1948.

5. Black, H. S. and J. O. Edson. Pulse code modulation. AIEE Trans. vol. 66, pp. 495-499. 1947.

6. Bluestein, L. I. Asymptotically optimum quantizers and optimum analog to digital converters. IEEE Trans. Information Theory (Correspondence), vol. IT-10, pp. 242-246. 1964.

7. Bluestein, L. I. and R. J. Schwarz. Optimum zero-memory filters. IRE Trans., vol. IT-8, pp. 337-342. 1962.

8. Bruce, J. D. Optimum quantization. Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, Mass. Tech. Rept. 429, March, 1965.

9. Carlitz, L. An extension of Mehler's formula. Boll.Un. Math. 21, Ital. (4) 3, pp. 43-46, 1970.

10. _____. Some extensions of Mehler formula. Collect. Math. pp. 117-130. 1970

11. Cramer, H. Mathematical Methods of Statistics. Princeton University Press, 1946. 575 p.

12. Davis, C. G. An experimental pulse-code modulation system for short-haul systems. Bell. Sys. Tech. J., vol. 41, pp. 1-24.1962.

13. Elias, P. Bounds on performance of optimum quantizers. IEEE Trans Information Theory, vol. IT-16, pp. 172-184. 1970.

14. _____. Bounds and asymototes for the performance of multivariate quantizers. Annals of Mathematical Statistics, vol. 41, part II, pp. 1249-1259. 1970.

15. Fleischer, P. E. Sufficient conditions for achieving minimum distortion in a quantizer. IEEE International Convention Record, Part I, pp. 104-111. 1964.

16. Garmash, V. A. The quantization of signals with non-uniform steps. Elektrosvyaz, vol. 10, pp. 10-12. 1957.

17. Gish, H. and J. N. Pierce. Asymptotically efficient quantizing. IEEE Trans. Information Theory, vol. IT-14, pp. 676-683. September, 1968.

18. Goodall, W. M. Telephony by pulse-code modulation. Bell Sys. Tech J., vol. 26, pp. 395-409. 1947.

19. Goulds, H. W. and W. Squire. Maclaurin's second formula and its generalizations. American Mathematical Monthly, vol. 70, pp. 44-52, January, 1963.

20. Hammerle, K. J. On the effect of noisy smeared signal, first order-first order receiver system. Analysis Group Report 3 . (Boeing Company).

21. Hardy, G. H. and J. E. Littlewood and G. Polya. Inequalities. London: Cambridge University Press, 1934. 314 p.

22. Harvard Computation Laboratory. Tables of the Error Function and Its First Twenty Derivatives, vol. 23, Annals of Harvard Computation Laboratory, Harvard University Press, 1946.

23. Hildebrand, F. B. Introduction to Numerical Analysis. New York: McGraw-Hill, 1956. 511 p.

24. Hurd, W. J. Correlation function of quantized sine wave plus Gaussian noise. IEEE Trans. Information Theory, vol. IT-13, pp. 65-68, January 1967.

25. Jackson, D. Fourier Series and Orthogonal Polynomials. The Carus Mathematical Monographs, No. 6, The Mathematical Association of America, Oberlin, Ohio, 1941. 234 p.

26. Kendall, M. G. Proof of relations connected with the tetrachoric series and its generalization. Biometrika, vol. 40, pp. 196-198. 1953.

27. Kosyakin, A. A. The statistical theory of amplitude quantization. Automatika Telemekh, vol. 22, pp. 624-630. June, 1961.

28. Laning, J. H. and R. H. Battin. Random Processes in Automatic Control. New York, McGraw-Hill, 1956. 434 p.

29. Lozovoy, I. A. Regarding the computation of the characteristics of compression in systems with pulse-code modulation. Telecommunications, No. 10, pp. 18-25.

30. Mann, H. H. M. Straube, and C. P. Villars. A companded coder for an experimental PCM terminal. Bell Sys. Tech. J., vol. 41, pp. 173-226. 1962.

31. Max, J. Quantizing for minimum distortion. IEEE Trans. Information Theory, vol. IT-6, pp. 7-12. 1960.

32. Messerschmitt, D. G. Quantizing for maximum output entropy. IEEE Trans. Information Theory, vol. IT-17, pp. 612-615. September, 1971.

33. Munroe, M. E. An Introduction to Measure and Integration. Cambridge, Mass., Addison-Wesley, 1953. 310 p.

34. Panter, P. F. and W. Dite. Quantization distortion in pulse count modulation with non-uniform spacing of levels. Proc. IRE, vol. 39, pp. 44-48. January, 1951.

35. Papoulis, A. The Fourier Integral and its applications. New York, McGraw-Hill, 1962. 318 p.

36. Price, R. A useful theorem for non-linear devices having Gaussian inputs. IREE Trans. Information Theory, vol. IT-4, pp. 69-72. 1958.

37. Purton, R. F. A survey of telephone speech signal statistics and their significance in the choice of a PCM companding law. Proc. Inst. Elec. Engrs., (London), vol. 109B, pp. 60-66. 1962.

38. Rainville, E. D. Special Functions. 3rd ed., New York, MacMillan, 1960. 365 p.

39. Rice, S. O. Mathematical Analysis of Random Noise. Bell Sys. Tech J. vol. 23, pp. 282-332 and vol. 24, pp. 46-156.

40. Robin, L. The autocorrelation function and power spectrum of clipped noise. Ann. Telecomm., vol. 7, pp. 375-387. 1952.

41. Roe, G. M. Quantizing for minimum distortion. IEEE Trans. Information Theory (Correspondence), vol. IT-10, pp. 384-385. 1964.

42. Schwartz, S. C. Estimation of probability density by an orthogonal series. The Annals of Mathematical Statistics, vol. 38, Part 2, pp. 1261-1265. 1967.

43. Schwartz, S. C. and W. L. Root. On dominating an average associated with dependent Gaussian vectors. The Annals of Mathematical Statistics, vol. 39, pp. 1844-1848. 1968.

44. Shannon, C. E. A Mathematical Theory of communication. Bell Sys. Tech. J., vol. 27, pp. 379-423, and pp. 623-656. July, 1948.

45. Sheppard, W. F. On the calculation of the most probable values of frequency constants, for data arranged according to equidistant divisions of a scale. London Mathematical Society, 29, pp. 353-357. 1898.

46. Sherman, R. J. On output statistics of non-linear devices: 1. Third and higher order information, 2. Quadriphase carrier reconstruction, 3. Analysis of point processes. Ph.D. dissertation at Oregon State University, Corvallis, Oregon. 1969.

47. Shtein, V. M. On group signal transmission with frequency division of channels by the pulse-code modulation method. Elektrosvyaz, No. 2, pp. 43-54. 1959.

48. Slepian, David. On the symmetrized Kronecker power of a matrix and extension of Mehler's formula for Hermite polynomials. SIAM Journal on Mathematical Analysis, vol. 3, No. 4, Nov., 1972.

49. Smith, B. Instantaneous companding of quantized signals. Bell Sys. Tech. J., vol. 36, pp. 653-709. May, 1957.

50. Srivastava, H. M. and J. P. Singhal. Some extensions of the Mehler formula. Proceedings of the American Mathematical Society, vol. 31, pp. 135-141. January, 1972.

51. Trofimov, B. E. Quantization noises in the coding of signals of uniform spectral density. Elektrosvyaz, No. 7, pp. 3-12. 1960.

52. Tukey, J. W. Data analysis and frontiers of geophysics. Science, vol. 148, part 2, pp. 1283-1289. 1965.

53. Velichkin, A. I. Correlation function and spectral density of a quantized process. Telecommunications and Radio Engineering, part II, Radio Engineering, pp. 70-77. July, 1962.

54. _____, Optimum characteristics of quantizer. Telecommunications and Radio Engineering, Part II, Radio Engineering, vol. 18, pp. 1-7. February, 1963.

55. Widrow, B. A study of rough amolitude quantization by means of Nyquist sampling theory. IRE Tran., Circuit Theory, vol. 1-3, pp. 266-276. December, 1956.

56. Wiggins, M. J. and R. A. Branham. Reduction in quantizing levels for digital voice transmission. IEEE International Convention Record, Part 8, pp. 282-288. 1963.

57. Wood, R. C. On optimum quantization. IEEE Trans. Information Theory, vol. IT-15, pp. 248-252. March, 1969.

58. Wainstein, L. A. and V. D. Zubahov. Extraction of Signals from Noise. Prentice-Hall, 1962. 382 p.

59. Wozencraft, J. M. and I. M. Jacobs. Principles of Communication Engineering. John Wiley, New York, 1965. 720 p.

60. Zador, P. Development and evaluation of procedures for quantizing multivariate distributions. Ph.D. dissertation, Stanford University, Stanford, California. 1964.

APPENDIX

An explanation of the many forms of $u(\tau)$--the τ represents the time interval between sample instants-- for the several forms of higher order density functions goes as follows. Consider that a filter is opened in the presence of the noise of the universe, the so-called white noise, in such a manner as to capture the signal energy, also assumed to be distributed in frequency and amplitude. This situation is crudely depicted in Figure (4.5). A pseudo-filter concept, originally due to K. J. Hammerle (1959) of the Boeing Company, seems to handle the situation. Define a pseudo-filter characteristic as

$$F_p(\omega) = G(\frac{\omega-\omega_o}{\omega_s})F(\frac{\omega-\omega_o}{\omega_1}) + G(\frac{\omega+\omega_o}{\omega_s})F(\frac{\omega+\omega_o}{\omega_1}) , \qquad (A.1)$$

where ω_s is the half power bandwidth associated with the signal energy distributed about $\pm\ \omega_o$. Think of signal energy being admitted by opening a filter defined by $G(\omega)$ in the presence of white noise of power level ψ_s. Then signal power admitted by the physical filter, defined by $F(\omega)$, is

$$N_s = \psi_s \int_{-\infty}^{\infty} g^2(\alpha)\, d\alpha , \qquad (A.2)$$

where $g(t)$ is the Fourier transform of $G(\omega)$. Similarly, noise power is defined by

$$N_o = \psi_o \int_{-\infty}^{\infty} f^2(\alpha)\, d\alpha , \qquad (A.3)$$

where $f(t)$ is the Fourier transform of $F(\omega)$. It seems proper to choose the most simple form of $G(\omega)$ to avoid undue complexity. That is, let $G(\omega)$ be the characteristic of a first order filter,

$$G(\omega) = \frac{1}{1+i\omega}, \tag{A.4}$$

so

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\left[\frac{1}{1+i\frac{\omega-\omega_o}{\omega_s}} + \frac{1}{1+i\frac{\omega-\omega_o}{\omega_s}}\right] d\omega \tag{A.5}$$

$$= \frac{\omega_s \cos\,\omega_o t}{\pi}\int_{-\infty}^{\infty} \frac{e^{ix\omega_s t}}{1+ix}\,dx = \frac{\omega_s \cos\,\omega_o t}{\pi}\int_{-\infty}^{\infty} e^{ix\omega_s t}\,\frac{1-ix}{1+x^2}\,dx.$$

Because of the well-known integrals,

$$\int_{-\infty}^{\infty} \frac{\cos\,ax}{1+x^2}\,dx = \int_{-\infty}^{\infty} \frac{x\sin\,ax}{1+x^2}\,dx = \pi e^{-a}, \tag{A.6}$$

$$a > 0,$$

the final form of this Fourier transform is

$$g(t) = 2\omega_s e^{-\omega_s t}\cos\,\omega_o t\; U(t), \tag{A.7}$$

where the unit step function has the usual definition. The one-sided nature of this transform has the physical implication "realizable". Thus signal power present is

$$N_s = 4\omega_s^2\psi_s\int_0^{\infty} e^{-2\omega_s\alpha}\cos^2\omega_o\alpha\,d\alpha = \omega_s\psi_s\left[\frac{1}{\omega_s} + \frac{\omega_s^2}{\omega_o^2+\omega_s^2}\right] \tag{A.8}$$

$$= \omega_s\psi_s\,\frac{\omega_o^2+2\omega_s^2}{\omega_o^2+\omega_s^2} \approx \omega_s\psi_s\,, \quad \omega_o >> \omega_s\,.$$

The effect of opening a fictitious (or pseudo-) filter in the presence of pure noise at level ψ_o is equivalent to opening a physical filter, defined by $F(\omega)$, in the presence of pure noise and normally distributed signal energy with given parameters ψ_s and ω_s, such that all the signal energy is admitted. The pseudo-filter has a

characteristic of the form

$$F_p(\omega) = \left[\frac{\psi_s/\psi_o}{1+i\frac{\omega-\omega_o}{\omega_s}} + 1\right] F\left(\frac{\omega-\omega_o}{\omega_1}\right) + \left[\frac{\psi_s/\psi_o}{1+i\frac{\omega+\omega_o}{\omega_s}} + 1\right] F\left(\frac{\omega+\omega_o}{\omega_1}\right) . \qquad (A.9)$$

It is convenient to define signal to noise ratio as

$$z = \frac{N_s}{N_o} = \frac{\psi_s \int_0^\infty g^2(\alpha)\, d\alpha}{\psi_o \int_0^\infty f^2(\alpha)\, d\alpha} \qquad (A.10)$$

and the bandwidth ratio

$$\delta = \frac{\omega_1}{\omega_s} > 1 , \qquad (A.11)$$

so that one might say that essentially all signal energy is admitted by the physical filter. To begin with, let the physical filter also be of first order; the Fouier transform is at once the same as in Equation (A.7),

$$f(t) = 2\omega_1 e^{-\omega_1 t} \cos \omega_o t\, U(t), \qquad (A.12)$$

and the signal to noise ratio is

$$\frac{\omega_s \psi_s}{\omega_1 \psi_o} \frac{\omega_o^2+2\omega_s^2}{\omega_o^2+\omega_s^2} \frac{\omega_o^2+\omega_1^2}{\omega_o^2+2\omega_1^2} \approx \frac{\omega_s \psi_s}{\omega_1 \psi_o} , \quad \omega_o >> \omega_1 > \omega_s . \qquad (A.13)$$

It is convenient to set

$$\frac{\psi_s}{\psi_o} + 1 = A , \qquad (A.14)$$

so

$$f_p(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t} F_p(\omega)\, d\omega$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t}\left[\frac{A + i\frac{\omega-\omega_o}{\omega_s}}{1+i\frac{\omega-\omega_o}{\omega_s}}\;\frac{1}{1+\frac{\omega-\omega_o}{\omega_1}} + \frac{A+i\frac{\omega+\omega_o}{\omega_s}}{1+i\frac{\omega+\omega_o}{\omega_s}}\;\frac{1}{1+i\frac{\omega+\omega_o}{\omega_1}}\right] d\omega \qquad (A.15)$$

$$= \frac{\omega_1 \cos\,\omega_o t}{\pi}\int_{-\infty}^{\infty} e^{ix\omega_1 t}\;\frac{A + i\delta x}{1 + i\delta x}\;\frac{1}{1+ix}\, dx \;.$$

This integrand should be arranged into even and odd terms;

$$f_p(t) = \frac{\omega_1 \cos\,\omega_o t}{\pi(\delta-1)}\int_{-\infty}^{\infty}\left[\cos x\omega_1 t\left[\frac{\delta-A}{1+x^2} + \frac{\delta(A-1)}{1+\delta^2x^2}\right]\right. \qquad (A.16)$$

$$\left. + x \sin x\omega t\left[\frac{\delta-A}{1+x^2} + \frac{\delta^2(A-1)}{1+\delta^2x^2}\right]\right] dx.$$

A slight rearrangement of the definite integrals of Equation (A.6) is

$$\int_{-\infty}^{\infty}\frac{\cos ax}{1+b^2x^2}dx = \int_{-\infty}^{\infty}\frac{bx \sin ax}{1+b^2x^2}\, dx = \frac{\pi}{b}e^{-a/b}\,, \qquad (A.17)$$

$$a,b > 0\,,$$

and it is clear that $f_p(t)$ is zero for negative values of t. Finally

$$f_p(t) = 2\omega_1 \cos\,\omega_o t\,\frac{(\delta-A)e^{-\omega_1 t} + (A-1)e^{-\omega_s t}}{\delta-1}\,U(t)\,. \qquad (A.18)$$

Note that for $A = 1$, no signal energy present, this reduces to $f(t)$ for the first order physical filter.

With this model of Gaussian noise plus Gaussian signal in a

sufficiently wide first order filter the normalized autocorrelation function becomes a simple sum of exponentials multiplied by the sinusoidal factor. To avoid undue complexity the Riemann theorem

$$\lim_{\omega_o \to \infty} \int_a^b f(x) \cos \omega_o x \, dx = 0, \quad f(x) \in C , \tag{A.19}$$

may be employed. This applies under the restriction $\omega_o >> \omega_1 > \omega_s$. thus the exact expression for the pertinent integral is

$$\int_0^\infty f_p(\alpha) f_p(\alpha+\tau) \, d\alpha = \left(\frac{2\omega_1}{\delta-1}\right)^2 \int_0^\infty \cos \omega_o \alpha [(\delta-A)e^{-\omega_1 \alpha} - (A-1)e^{-\omega_s \alpha}] \cdot [(\delta-A)e^{-\omega_1(\alpha+\tau)} + (A-1)e^{-\omega_s(\alpha+\tau)}] \cos \omega_o(\alpha+\tau) \, d\alpha, \tag{A.20}$$

and if the Riemann theorem is applied,

$$\begin{aligned}
&\int_{-\infty}^\infty f_p(\alpha) f_p(\alpha+\tau) \, d\alpha \\
&= \left(\frac{2\omega_1}{\delta-1}\right)^2 e^{-\omega_1 \tau} \cos \omega_o \tau \int_0^\infty \tfrac{1}{2}(\delta-A)[(\delta-A)e^{-2\omega_1 \alpha} + (A-1)e^{-(\omega_1+\omega_s)\alpha}] \, d\alpha \\
&+ \left(\frac{2\omega_1}{\delta-1}\right)^2 e^{-\omega_s \tau} \cos \omega_o \tau \int_0^\infty \tfrac{1}{2}(A-1)[(\delta-A)e^{-(\omega_1+\omega_s)\alpha} + (A-1)e^{-2\omega_s \alpha}] \, d\alpha \\
&= \frac{2\omega_1^2 \cos \omega_o \tau}{(\delta-1)^2} \; e^{-\omega_1 \tau}(\delta-A)\left[\frac{\delta-A}{2\omega_1} + \frac{A-1}{\omega_1+\omega_s}\right] + e^{-\omega_s \tau}(A-1)\left[\frac{\delta-A}{\omega_1+\omega_s} + \frac{A-1}{2\omega_s}\right] \\
&= \omega_1 \cos \omega_o \tau \, \frac{(\delta^2 - A^2) e^{-\omega_1 \tau} + \delta(A^2-1) e^{-\omega_s \tau}}{\delta^2 - 1} .
\end{aligned} \tag{A.21}$$

The normalized autocorrelation function of interest in implementing the developments above is, therefore,

$$u(\tau,\delta) = \frac{(\delta^2 - A^2)\, e^{-\omega_1 \tau} + \delta(A^2 - 1)\, e^{-\omega_s \tau}}{(\delta - 1)(\delta + A^2)} \cos \omega_o \tau . \qquad (A.22)$$

Note that for $A = 1$, no signal energy at all, this reduces to

$$u(\tau) = e^{-\omega_1 \tau} \cos \omega_o \tau ,$$

the first order physical filter case. Similar calculations can be made for filters of other than first order but there is considerable effort involved.