63-3-6

RADC-TDR-63-136

14 March 1963

Final Report

STATISTICAL INSTRUMENTATION STUDY

A. W. Crooke
W. B. Floyd
A. H. Nuttall

LITTON SYSTEMS, INC.
Data Systems Division
Communication Sciences Laboratory
221 Crescent Street
Waltham 54, Massachusetts

Contract AF30(602)-2663
Project 4519, Task 451903
Electronic Systems Division (RADC)

Prepared for
Rome Air Development Center
Air Force Systems Command
United States Air Force

Griffiss Air Force Base
New York

RADC-TDR-63-136

14 March 1963

Copy No. 34

Final Report

Statistical Instrumentation Study

A. W. Crooke
W. B. Floyd
A. H. Nuttall

Litton Systems, Inc.
Communication Sciences Laboratory
Data Systems Division
Waltham 54, Massachusetts

Contract Number AF30(602)-2663

Project 4519, Task 451903
Electronic Systems Division (RADC)

Prepared for
Rome Air Development Center
Air Force Systems Command
United States Air Force
Griffiss Air Force Base
New York

Statistical Instrumentation Study

Contract AF30(602)-2663

14 March 1963

Prepared by

A. W. Crooke
W. B. Floyd
A. H. Nuttall

Approved by

George Sebestyen
Technical Director

John Gerdes
Assistant Manager

LITTON SYSTEMS, INC.
DATA SYSTEMS DIVISION

Foreword

This report has been prepared by the Communication Sciences Laboratory within the Data Systems Division of Litton Systems, Inc., a division of Litton Industries. The work reported here has been performed over a period of 12 months, under Contract Number AF30(602)-2663, as Task Number 451903 of Project Number 4519, entitled "Statistical Instrumentation Study". This project has been completed under the direction of the Communications Directorate within the Rome Air Development Center.

Several individuals within the Communication Sciences Laboratory have made major contributions to the study and development of high order statistical estimation techniques reported here. The Probability Analyzer breadboard device has been designed and constructed under the guidance of Mr. Arthur Crooke; most of the theoretical studies have been conducted by Mr. Thomas Crystal, Mr. William Floyd and Dr. Albert Nuttall; and Dr. George Sebestyen has provided technical guidance for the entire program.

ABSTRACT

Several techniques for estimating n-th order statistics of signals are investigated, including curve fitting methods involving estimation of average values of functions of signal amplitudes, and success counting methods for which probabilities are estimated as the percentage time that a specified condition exists.

One of the success counting methods is selected for implementation, and a breadboard model constructed. This device will calculate fourth (and lower) order joint and conditional probability density functions and distribution functions for signals with bandwidth less than 10 Kcps. Provisions are incorporated in the breadboard for calculating probability of any fourth order event in signal space, through simple and inexpensive modification of one unit in the device.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

Problems in which high-order statistical characterizations of signals are needed arise in many contexts, but the difficulties of measuring and presenting appropriate data for such characterizations are often effective barriers to their use. The absence of a device which would provide a complete characterization of second order statistics of signals, for instance, has led to the study of correlation between such quantities as estimates of target range and azimuth estimates provided at the output of a radar. But the correlation between two statistically fluctuating quantities may be not only relatively uninformative for characterizing their joint behavior, but also misleading if too much reliance is placed on the results obtained.

Moreover, for some computations there is no short cut for bypassing calculation of the joint n-th order statistics of a set of signals for n fairly large. Determination of the boundary in any multi-dimensional space which results from a threshold applied to the likelihood ratio* is an example.

A variety of techniques for estimating first order statistics of signals** have been implemented over the years, and in at least one case***, a device has been built to estimate up to third order statistics. In Section 2, several of these techniques are described and evaluated for audio frequency bandwidth signals. One of these techniques has been selected for implementation, and a breadboard has been constructed. The capabilities and details of construction of this device are described in Section 3.

Conclusions regarding the utility of the techniques studied, and recommendations for utilizing the capabilities of the device constructed, are presented in Section 4.

---

*[7]

**Including those described in [3], [8], and [11].

***[5]

2. METHODS OF ESTIMATING N-TH ORDER STATISTICS OF SIGNALS

2.1 STATISTICAL DESCRIPTION OF SIGNALS

Due either to a lack of knowledge, inability to discover the basic mechanism, or the genuinely random character of a signal or process, an exact deterministic description of a future event from past history is often impossible. For example, it is virtually impossible to predict the outcome of a thoroughly shaken die; yet theoretically, given the initial position of the die and the movements of the shaking element, the next outcome of the throw is determined. However, since the determination of the next outcome would involve an exhorbitant amount of detailed computations, we often choose to say that the outcome is random with the probability of any one particular face being 1/6. Furthermore, we say that each outcome is independent of previous results. In this manner, we obtain a complete (albeit approximate) statistical description of the random process called die throwing. Whether or not this is an adequate description depends on the symmetry of a particular die, and the amount of shaking before throwing.

Similarly, with voltages which vary as functions of time, when there appears (through a limited investigation) to be no underlying deterministic behavior, we characterize such processes by probabilistic statements. It is possible and customary to define a hierarchy of probabilistic rules, each of which is more general than the previous one, and the limit of which is defined as a complete statistical description of the process.

2.1.1 First Order Descriptors

The first rule is the probability that at a time t, a voltage value s(t) will be less than or equal to a given value x. This function, denoted $P_1(x, t)$, is called the first order probability distribution, and is the most general first-order statistic there is. From this quantity may be found other more simple first order statistics such as the mean, variance, or ν-th moment of the voltage, all at time t. Entirely equivalent to P_1 is the first order probability density function (p. d. f.), which is the derivative with respect to x of P_1, and the first order characteristic function, which is the Fourier transform of the p. d. f.

2.1.2 Second Order Descriptors

The second rule is the joint probability that at time t_1, the voltage is less than x_1, while at time t_2, the voltage is less than x_2. This function, $P_2(x_1, t_1; x_2, t_2)$, is the second order probability distribution, and is quite an information bearing quantity. For example, by letting x_2 equal infinity, we realize the first order distribution, but in addition, from P_2 we may calculate

various cross moments such as $E[s(t_1)s(t_2)]$ or $E[s^2(t_1)s^2(t_2)]$. The first of these two averages is the correlation function of the process $\{x(t)\}$ and has proved to be very useful in filtering and prediction.* For instance, if the correlation function depends only on $t_2 - t_1$, the Fourier transform of the correlation function indicates directly where in frequency the power of the process is located. The latter function is, of course, the power density function. Notice that although P_2 gives the correlation function and/or the power density spectrum, P_2 cannot in general be found from these latter quantities. That is, P_2 is a much more general second order descriptor. Again, by differentiation or Fourier transformation, respectively, two equivalent descriptors obtained are the second order p.d.f. and characteristic function.

2.1.3 Higher Order Descriptors

The first and second order descriptors can be extended to the n-th order, where one asks for the probability that at time t_k, the voltage is less than x_k, $k = 1, 2, \ldots, n$. This is the n-th order probability distribution: $P_n(x_1, t_1; \ldots; x_n, t_n) \equiv P_n(\underset{\sim}{x}, \underset{\sim}{t})$. By letting some of the $\{x_k\}$ equal infinity, lower order distributions are obtained. The larger n is made, the more complete becomes the description of the random process.

In many problems one is interested in describing the statistical behavior of several processes or signals which are defined in terms of a common parameter. (Usually, time is the parameter.) For instance, it may be desirable to know the joint (second order) probability that voltage s_1 is less than x_1 at time t_1 and voltage s_2 is less than x_2 at time t_2. In general, the function $P_n(\underset{\sim}{x}, \underset{\sim}{t})$ can be interpreted as a joint probability distribution of n signals each of which may be observed at an independently specified time. By considering the n signals to be delayed versions of a single waveform, the n-th order probability distribution of the single waveform is seen to be a special case of the more general interpretation of $P_n(\underset{\sim}{x}, \underset{\sim}{t})$.

In summary, an n-th order statistical description of signals is provided by the probability that at time t_k voltage s_k is less than x_k, $k = 1, 2, \ldots, n$, designated $P_n(\underset{\sim}{x}, \underset{\sim}{t})$. The voltages $\{s_k\}$, values $\{x_k\}$, and times $\{t_k\}$, may or may not all be distinct.

*See [9], for instance.

2.1.4 Stationarity

When the n-th order probability distribution is a function only of time differences, and not absolute time, for all n, the process is called strict sense stationary. Such processes are often met in practice and possess the helpful feature of remaining "statistically constant" as time progresses. Thus, measurements at one time are as good as, and equivalent to, those made at another time. Naturally this property can be, and has been, utilized to simplify data collection and processing.

When the process is not necessarily strict sense stationary, but possesses a mean which is independent of time, and a correlation function which is dependent only on time differences, then the process is called wide-sense stationary. The effects of nonstationarity or estimation accuracy are discussed in Section 2.3.

2.2 METHODS OF ESTIMATING N-TH ORDER STATISTICS

In this section, several methods will be discussed by which higher order statistical signal descriptors can be estimated. Before going into any details, it is appropriate to point out an important qualitative aspect of the problem of estimating statistical characteristics of signals. If a reasonably large amount of time is available for processing the signals, then questions of optimal efficiency of estimation methods in making use of a given sample size are not crucial. Only if the time required to obtain a large amount of useful history of the signals is high, do these questions deserve close scrutiny. With this in mind, we have considered the selection of estimation methods to be based primarily on (a) the versatility of the technique, and (b) simplicity and cost of construction and operation of equipment which is required to implement any method.

There are essentially two basic approaches to the problem of estimating an n-th order distribution function $P_n(\underset{\sim}{x}, \underset{\sim}{t})$. The first approach involves the measurement of average values of quantities which are related to the signal amplitudes, and will be called the moment estimation approach. The second approach involves calculation of the percentage of time that the signal amplitudes satisfy a specified set of conditions, and will be called the success counting approach. Each of several ways of exploiting these two approaches to the problem of estimating $P_n(\underset{\sim}{x}, \underset{\sim}{t})$ will be described in this section, indicating the versatility, relative accuracy and simplicity of equipment associated with the technique.

The basic restrictions on signal sets which we shall assume for the techniques to be considered are that (a) the amplitude of each signal is bounded (the j-th amplitude, s_j, lies in the range $L_{sj} \leq s_j \leq L_{mj}$; $R_j = L_{mj} - L_{sj}$) and

(b) the bandwidth, W_j, of each signal is less than a specified quantity W.

2.2.1 Moment Estimation Methods

It is well known that the n-th order probability density function,* $p_n(\underset{\sim}{x})$, can be obtained from a knowledge of all moments of the signal amplitudes, i.e., the average values of the products of powers of the signal amplitudes. Specifically, $p_n(\underset{\sim}{x})$ is the Fourier transform of the characteristic function, $F_n(\xi_1, \xi_2, \ldots, \xi_n) = F_n(\underset{\sim}{\xi})$, and $F_n(\underset{\sim}{\xi})$ can be expressed in terms of the moments by a power series:

$$F_n(\underset{\sim}{\xi}) = \sum_{k=0}^{\infty} \sum_{\{k_m\}}^{(k)} \left(\overline{\prod_{m=1}^{n} x_m^{k_m}} \right) \prod_{m=1}^{n} \left(\frac{i\xi_m}{k_m!} \right)^{k_m} \tag{1}$$

where $\sum_{\{k\}}^{(k)} = \sum_{\text{All } \{k_m\} \text{ such that } \sum_{m=1}^{n} k_m = k}$

and $\overline{\prod_{m=1}^{n} x_m^{k_m}}$ = one of the $\binom{n+k-1}{k}$ different moments of degree k, $\sum_{m=1}^{n} k_m = k$.

This characterization is possible for many processes with finite moments. Thus, conceivably a procedure for estimating the moments of a process could be set up, and the probability density function could be obtained by transforming the resulting estimate of the characteristic function for the process. In practice, however, only a finite number of moments can be estimated. The question then arises: how is $p_n(\underset{\sim}{x})$ related (even approximately) to an incomplete power series representation of $F_n(\underset{\sim}{\xi})$? Unfortunately, except in cases for which the function is already known (notably the Gaussian p.d.f.), the answer to this question is in

*For convenience, we shall write $p_n(\underset{\sim}{x})$ and $P_n(\underset{\sim}{x})$ for $p_n(\underset{\sim}{x}, \underset{\sim}{t})$ and $P_n(\underset{\sim}{x}, \underset{\sim}{t})$, respectively, wherever it is not necessary to specify explicitly the times, $\underset{\sim}{t}$.

general unknown. Therefore, if moments of a process are to serve as the initial information bearing quantities from which an estimate of $p_n(\underset{\sim}{x})$ is to be derived, some other rationale for utilizing a finite number of moments must be developed.

2.2.1.1 Power Series Approximation

One method of utilizing estimates of a finite number of moments is to attempt to approximate $p_n(\underset{\sim}{x})$ with a polynomial of the form

$$p_n(\underset{\sim}{x}) \cong p_{nq}(\underset{\sim}{x}) \equiv \sum_{j=o}^{q} \sum_{k=1}^{A_{nj}} a_{jk}\, x_1^{i_1}\, x_2^{i_2} \cdots x_n^{i_n} \tag{2}$$

where $\sum_{m=1}^{n} i_m = j$ for all values of j and k, and k takes on $A_{nj} = \binom{n+j-1}{j}$ values,

each of which corresponds to some ordering of the possible values for the n-tuples $\{i_1, i_2, \ldots, i_n\}$, and q is called the degree of the polynomial. With this approach, there are two parts in the problem of estimating a probability density function:

1) Finding the coefficients, $\{a_{jk}\}$, which provide a good fit of a q-th degree polynomial to the true probability density function, and

2) Obtaining estimates $\{\hat{a}_{jk}\}$ of these coefficients.

The result is an estimate, $\hat{p}_{nq}(\underset{\sim}{x})$, of the probability density function:

$$\hat{p}_{nq}(\underset{\sim}{x}) = \sum_{j=o}^{q} \sum_{k=1}^{A_{nj}} \hat{a}_{jk}\, x_1^{i_1}\, x_2^{i_2} \cdots x_n^{i_n} \tag{3}$$

To solve the first part of this problem, we need a device for measuring the error in some sense between the polynomial $p_{nq}(\underset{\sim}{x})$ and $p_n(\underset{\sim}{x})$. One such measure is the integrated square of the difference between $p_n(\underset{\sim}{x})$ and $p_{nq}(\underset{\sim}{x})$, where the

averaging is carried out with uniform weighting over the region for which $p_n(\underset{\sim}{x})$ is non-zero. Denoting this quantity by ϵ_q^2 we have

$$\epsilon_q^2 = \int_X \{p_n(\underset{\sim}{x}) - p_{nq}(\underset{\sim}{x})\}^2 \, d\underset{\sim}{x} \tag{4}$$

where X denotes all values of $\underset{\sim}{x}$ for which $p_n(\underset{\sim}{x})$ is non zero. If it is desired that $p_n(\underset{\sim}{x})$ be approximated with high percentage accuracy over all of the region X, then perhaps a better criterion is the integrated square of the relative error between $p_n(\underset{\sim}{x})$ and $p_{nq_i}(\underset{\sim}{x})$. This quantity, denoted by ϵ_{qr}^2, can be written

$$\epsilon_{qr}^2 = \int_X \left\{ 1 - \frac{p_{nq}(\underset{\sim}{x})}{p_n(\underset{\sim}{x})} \right\}^2 d\underset{\sim}{x}. \tag{5}$$

While there are other means of determining goodness of fit we shall limit our discussion to these two.

No matter which method is employed, the general procedure for determining the form of the $\{a_{jk}\}$ is the same. To illustrate this procedure, it will be carried out in detail using the measure ϵ_q^2. Here, the problem is to choose the $\{a_{jk}\}$ such that ϵ_q^2 is minimized, where

$$\epsilon_q^2 = \int_X \left\{ p_n(\underset{\sim}{x}) - \sum_{j=o}^{q} \sum_{k=1}^{A_{nj}} a_{jk} x_1^{i_1} \cdots x_n^{i_n} \right\}^2 d\underset{\sim}{x} \tag{6}$$

(Weierstrass' theorem tells us that we can choose the $\{a_{jk}\}$ to make $\lim_{q \to \infty} \epsilon_q^2 = 0$ if the region over which $p_n(\underset{\sim}{x})$ is non-zero is closed, and if $p_n(\underset{\sim}{x})$ is reasonably well-behaved.)

Setting the first derivative of ϵ_q^2 with respect to a_{rm} (for which $i_j = m_j$, $j = 1, 2, \ldots, n$) equal to zero, for all $r = 0, 1, \ldots, q$, and $m = 1, 2, \ldots A_{nr}$, provides $N_{qn} = \sum_{j=o}^{q} A_{nj} = \binom{n+q}{q}$ equations in N_{qn} unknowns:

$$\sum_{j=o}^{q} \sum_{k=1}^{A_{nj}} C_{jk}^{(m)} \; a_{jk} = d_m, \; m = 1, 2, \ldots, N_{qn}, \tag{7}$$

where

$$C_{jk}^{(m)} \equiv \int_X x_1^{m_1+k_1} \; x_2^{m_2+k_2} \ldots x_n^{m_n+k_n} \; d\underset{\sim}{x}$$

$$d_m \equiv \int_X x_1^{m_1} \; x_2^{m_2} \ldots x_n^{m_n} \; p_n(\underset{\sim}{x}) \; d\underset{\sim}{x}$$

Or, rewriting these equations in matrix notation,

$$\underset{\sim}{C} \; \underset{\sim}{a} = \underset{\sim}{d} \tag{8}$$

where

$$\underset{\sim}{C} \equiv \begin{bmatrix} C_{01}^{(1)} & C_{11}^{(1)} & C_{12}^{(1)} \ldots C_{1n}^{(1)} & C_{21}^{(1)} & C_{22}^{(1)} \ldots & C_{q, A_{qn}}^{(1)} \\ C_{01}^{(2)} & C_{11}^{(2)} & C_{12}^{(2)} \ldots C_{1n}^{(2)} & C_{21}^{(2)} & C_{22}^{(2)} \ldots & C_{q, A_{qn}}^{(2)} \\ C_{01}^{(N_{qn})} & C_{11}^{(N_{qn})} \; \vdots & & & & C_{q}^{(N_{qn})} \end{bmatrix}$$

and

$$\underset{\sim}{a} \equiv \begin{bmatrix} a_{01} \\ a_{11} \\ a_{12} \\ \cdot \\ \cdot \\ \cdot \\ a_{1n} \\ a_{21} \\ a_{22} \\ \cdot \\ \cdot \\ \cdot \\ a_{q,\, A_{nq}} \end{bmatrix} \qquad \text{and } \underset{\sim}{d} = \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ \cdot \\ \cdot \\ d_{N_{qn}} \end{bmatrix}$$

To solve for the $\{a_{jk}\}$ it is sufficient to invert the matrix $\underset{\sim}{C}$, to obtain

$$\underset{\sim}{a} = \underset{\sim}{C}^{-1} \underset{\sim}{d} \tag{9}$$

Thus, if $p_n(\underset{\sim}{x})$ is known, a q-th degree polynomial can best be fitted to $p_n(\underset{\sim}{x})$ (in the sense of minimizing integrated squared error) by solving (9). However, we do not know $p_n(\underset{\sim}{x})$, and in fact are concerned with the problem of estimating this function. This brings up the second part of the problem which is the estimation of the vector $\underset{\sim}{a}$.

Each of the $\{x_i\}$ is assumed to lie in a finite restricted range, and the elements of $\underset{\sim}{C}$ are independent of $p_n(\underset{\sim}{x})$ and can be calculated:

$$C_{jk}^{(m)} = \prod_{\nu=1}^{n} \frac{L_{m\nu}^{m_\nu+k_\nu+1} - L_{s\nu}^{m_\nu+k_\nu+1}}{m_\nu+k_\nu+1} \tag{10}$$

The elements of $\underset{\sim}{d}$ are moments of the amplitudes of the signals:

$$d_m = \overline{\prod_{\nu=1}^{n} x_\nu^{m_\nu}} \tag{11}$$

To estimate the $\{a_{jk}\} = \underset{\sim}{a}$ it is only necessary to estimate the $\{d_m\}$, and use

$$\hat{\underset{\sim}{a}} = \underset{\sim}{C}^{-1} \hat{\underset{\sim}{d}} \tag{12}$$

Thus, to obtain an estimate of $p_n(\underset{\sim}{x})$ using this method requires that a device for estimating all q-th and lower order moments of the signal amplitudes be developed, and the inversion of an $\binom{n+q}{q} \times \binom{n+q}{q}$ matrix. For $n > 1$, the latter operation will require the use of a general purpose computer. The main storage capacity of most computers would serve to preclude this calculation for values of n and q greater than 4.

The block diagram of a device which would provide estimates of moments of stationary signals is shown in Figure 1. Each moment is estimated by a time average over a finite interval, T, of the product of the powers of the $\{s_i\}$ involved.* Specifically,

$$\hat{d}_m = \frac{1}{T} \int_0^T \prod_{\nu=1}^{n} s_\nu^{m_\nu}(t)\, dt \tag{13}$$

As indicated in Figure 1, a programmer could be incorporated in the moment estimating device to automatically step the powers of the $\{s_i\}$ through all values less than a specified degree, q.

*This method of estimating average values of quantities is shown in Section 2.2.1.3 to be near-optimum.

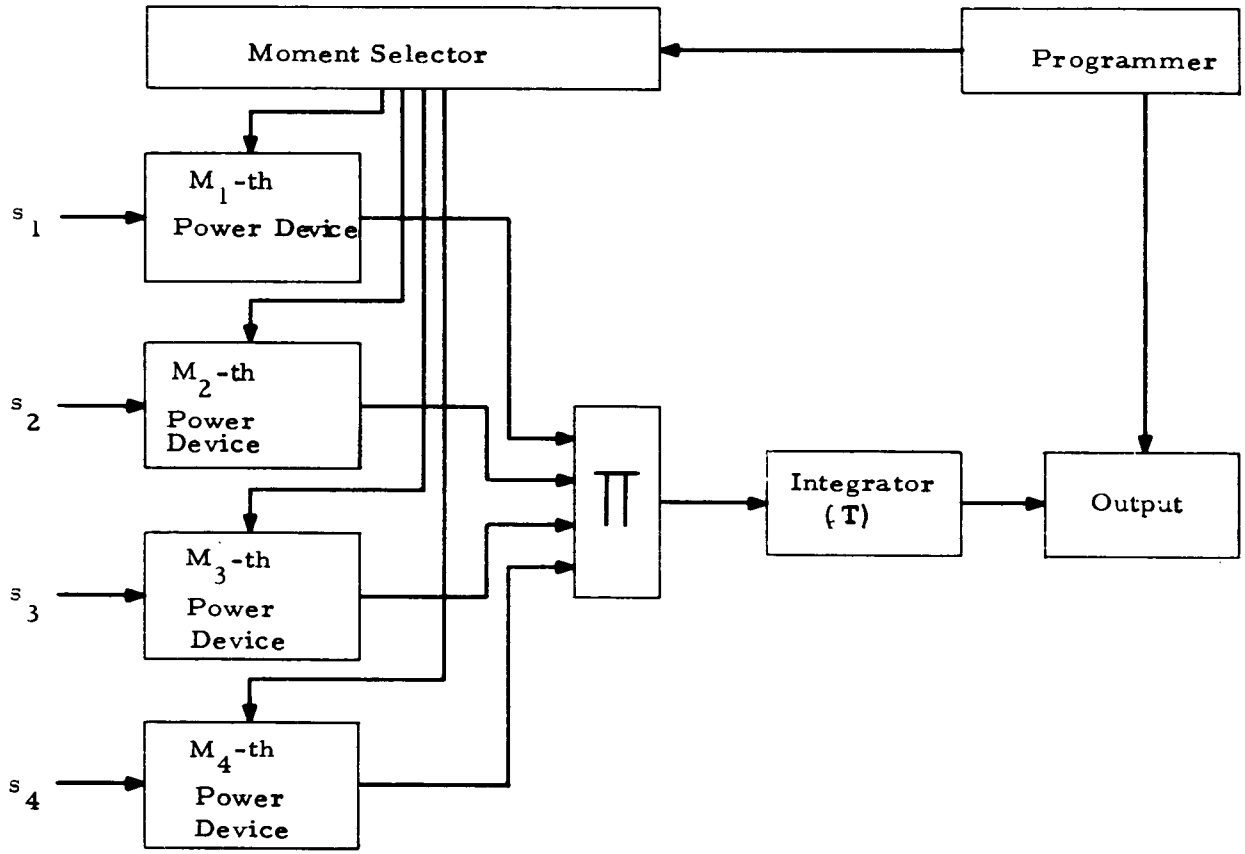


Figure 1. Block Diagram of a Fourth Order Moment Estimation Device

In reviewing the power series approximation method of estimating $p_n(\underset{\sim}{x})$ (or $P_n(\underset{\sim}{x})$), there appear to be three major objections to its use for $n > 1$. The first (and perhaps most important) objection is that preliminary or partial results cannot be obtained; i.e., no part of the probability density or distribution functions can be examined without going through a complete set of possibly expensive and time consuming calculations for $\hat{p}_{nq}(\underset{\sim}{x})$.

The second objection is that for values of n and q greater than one, a computer must be utilized to invert a matrix to obtain the desired coefficients. This feature degrades the potential utility of a probability density estimation device for performing a quick analysis of signals, unless a general purpose computer is immediately available to the user.

The other major objection to this technique is that the construction of an accurate power device and multiplier with wide dynamic range is a costly undertaking. In addition to the method outlined above, there are a variety of

techniques for designing these devices, but as the order n becomes larger, the accuracy requirements become so stringent that the equipment feasibility of calculating higher order moments becomes doubtful.

There are several other factors which bear upon the accuracy attainable with this approach to the problem (for instance, the effect of the limited processing time, T) some of which are common to all methods of estimation. An examination of moment estimation accuracy is reported in subsection 2.2.1.3.

If another method of measuring closeness of approximation is used in the power series method, then similar, but possibly less convenient results are obtained. For instance, by minimizing the integrated relative squared error, ε_{qr}^2, instead of ε_q^2, the solution for the $\{a_{jk}\}$ is again given by (9), but with

$$C_{jk}^{(m)} \equiv \int_X \frac{x_1^{m_1+k_1} x_2^{m_2+k_2} \cdots x_n^{m_n+k_n}}{[p_n(\underset{\sim}{x})]^2} \, d\underset{\sim}{x} \qquad (14)$$

and

$$d_m \equiv \int_X \frac{x_1^{m_1} x_2^{m_2} \cdots x_n^{m_n}}{p_n(\underset{\sim}{x})} \, d\underset{\sim}{x} \, .$$

Apparently, there is no direct way in which either the $\{C_{jk}^{(m)}\}$ or the $\{d_m\}$ can be estimated when $p_n(\underset{\sim}{x})$ is unknown. Therefore, in addition to the lengthy matrix calculations and cumbersome operations associated with an attempt to approximate $p_n(\underset{\sim}{x})$ with a power series, one must be content to attempt to minimize a particular measure of closeness of approximation to $p_n(\underset{\sim}{x})$.

2.2.1.2 General Series Approximation

A natural generalization of the power series approximation is the representation of $p_n(\underset{\sim}{x})$ with a series of the form

$$p_n(\underset{\sim}{x}) = p_{nM}(\underset{\sim}{x}) \equiv \sum_{m=o}^{M} a_m f_m(\underset{\sim}{x}) \tag{15}$$

where the $\{f_m(\underset{\sim}{x})\}$ are known functions defined over the region X. Following the same procedure as with the power series approximation, we may choose the coefficients $\{a_m\}$ so that the integrated squared error

$$\varepsilon_M^2 = \int_X \{p_n(\underset{\sim}{x}) - p_{nM}(\underset{\sim}{x})\}^2 \, d\underset{\sim}{x} \tag{16}$$

is minimized, These coefficients are determined by

$$\underset{\sim}{a} = \underset{\sim}{C}^{-1} \underset{\sim}{d} \tag{17}$$

where

$$C_{jk} \equiv \int f_j(\underset{\sim}{x}) f_k(\underset{\sim}{x}) \, d\underset{\sim}{x}$$

and $$d_j \equiv \int f_j(\underset{\sim}{x}) p_n(\underset{\sim}{x}) \, d\underset{\sim}{x} = \overline{f_j(\underset{\sim}{x})}$$

for $$j, k = 1, 2, \ldots, M.$$

By choosing the $\{f_m(\underset{\sim}{x})\}$ to be orthonormal over X, this solution is simplified to

$$a_m = \overline{f_m(\underset{\sim}{x})}, \tag{18}$$

and the problem of estimating $p_n(\underset{\sim}{x})$ with $p_{nM}(\underset{\sim}{x})$ evolves into that of estimating a_m with $\hat{a}_m$. Then

$$\hat{p}_{nM}(\underset{\sim}{x}) = \sum_{m=o}^{M} \hat{a}_m f_m(\underset{\sim}{x}) \tag{19}$$

While the matrix inversion problem involved in the power series approximation has been obviated, this approach still requires the construction of nonlinear devices for determining $\hat{p}_{nM}(\underset{\sim}{x})$.

One way to implement this technique is to choose the $\{f_m(\underset{\sim}{x})\}$ to be separable functions of $\underset{\sim}{x}$; i.e.,

$$f_m(\underset{\sim}{x}) = \prod_{k=1}^{n} g_{m_k}(x_k) \tag{20}$$

where a unique index m is assigned to each n-tuple $(m_1, m_2, \ldots, m_n)$. If the first order density function of each variate is to be approximated by q terms of a series expansion in the g-functions, then an estimate of the n-th order process can include up to $M = q^n$ terms. Estimates of the coefficients are now provided by

$$\hat{a}_m = \frac{1}{T} \int_T \prod_{k=1}^{n} g_{m_k}[s_k(t)]\, dt \tag{21}$$

The estimates $\{\hat{a}_m\}$ can be obtained by constructing a device as indicated in Figure 2. As indicated, if all q^n terms are to be included in the approximation,

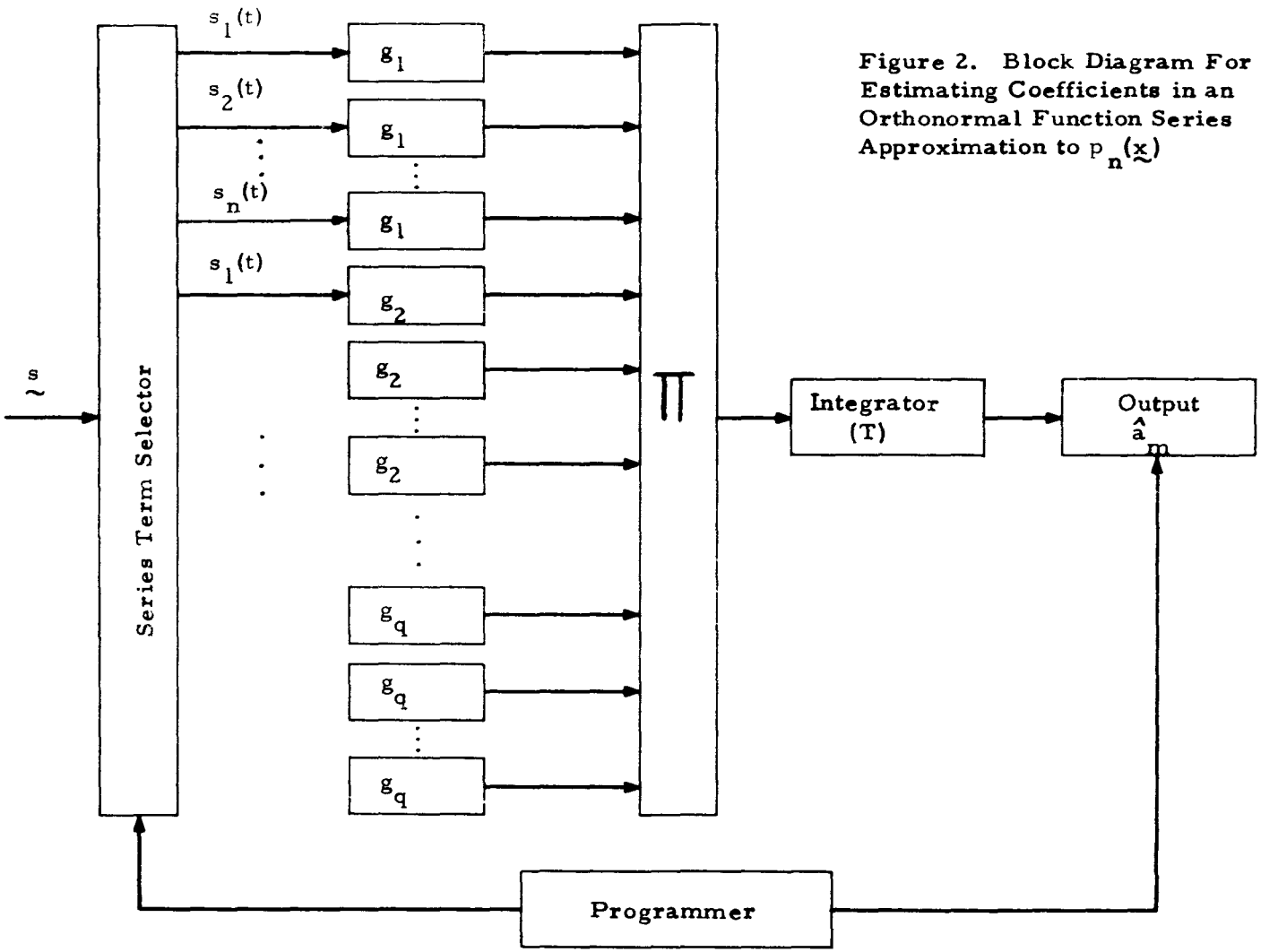


Figure 2. Block Diagram For Estimating Coefficients in an Orthonormal Function Series Approximation to $p_n(\underset{\sim}{x})$

then (nq) different g-function devices are required, since the series includes terms of the form $\prod_{k=1}^{n} g_{m_k}(x_k)$. It is conceivable that less than q^n terms could be used for the approximation to reduce equipment construction requirements. However, this would definitely degrade severely the accuracy with which some probability density functions could be estimated.

The major objections to the use of this method of estimation are (a) the requirement to perform all of the calculations involved in an estimate of the entire probability density function before any results can be obtained, (b) the high accuracy requirements in constructing the g-function devices, and (c) the large number of such devices required for accurately estimating higher order functions. The absence of a matrix inversion calculation makes this method much more attractive than the power series approximation.

We have considered the above objections to the orthonormal function series approximation method to be sufficient to remove it from consideration for implementation on this project. However, for display purposes this method does possess a novel feature which merits description.

Consider for the moment the problem of displaying the first order probability density function, $p_1(x)$. If $q = M$ linear filters are constructed so that the k-th filter has an impulse response, $g_k(t)$, then the estimate $\hat{p}_1(t)$ can be obtained as a function of time by summing the weighted outputs of these filters when their common input is an impulse. As indicated in Figure 3, the weighting coefficients are the $\{\hat{a}_k\}$. Thus, $\hat{p}_1(t)$ is computed by

$$p_1(t) = \sum_{k=1}^{m} \hat{a}_k \int \delta(t-\tau)\, g_k(\tau)\, d\tau \qquad (22)$$

$$= \sum_{k=1}^{m} \left\langle g_k[s(t)] \right\rangle_T$$

where the bracket symbols $\langle\ \rangle_T$ denote a finite time average of duration T.

This function of time can be displayed on an oscilloscope or in any other conveniently readable way.

Although this display feature has attracted experimenters to the orthonormal function series approximation technique for estimating first order probability density functions*, its practical utility for use with higher order statistics would be highly limited. First, only two-dimensional cross sections of $\hat{p}_n(\underset{\sim}{x})$ could be displayed, and these would have the abscissa defined by $x_j = x_1 + \delta_j$, $j = 1, \ldots, n$, where δ_j is a constant of a given display. A more important difficulty associated with this display method for use with higher order statistics is the complexity of equipment required. The form of the display device for n-th order functions is shown in Figure 4. Equipment requirements include q linear filters, nq adjustable delays, and q^n product devices. The output of each linear filter is delayed n (possibly all different) times and these n waveforms are routed to q^{n-1} different product devices. In the special case $n=1$, no product devices are required; however, for higher order statistics the utility of this display technique is tremendously outweighed by the equipment complexity involved.

*See, for instance, [10].

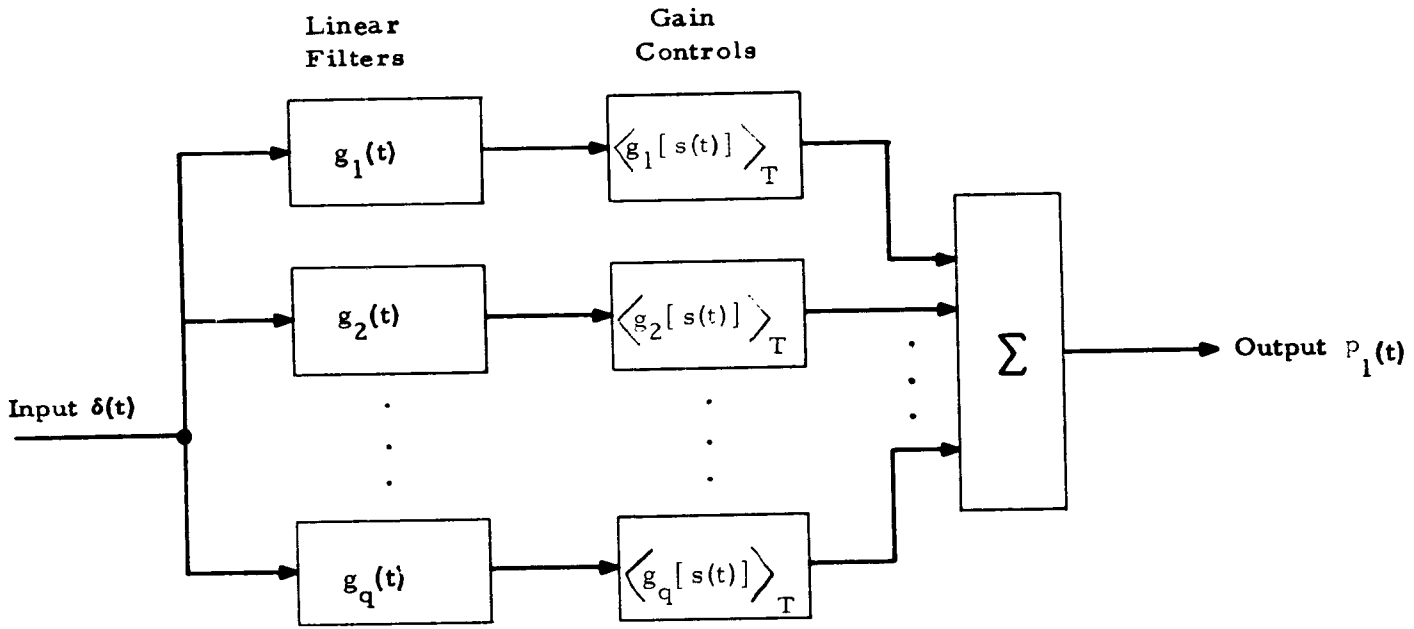


Figure 3. A Device for Displaying $p_1(x)$

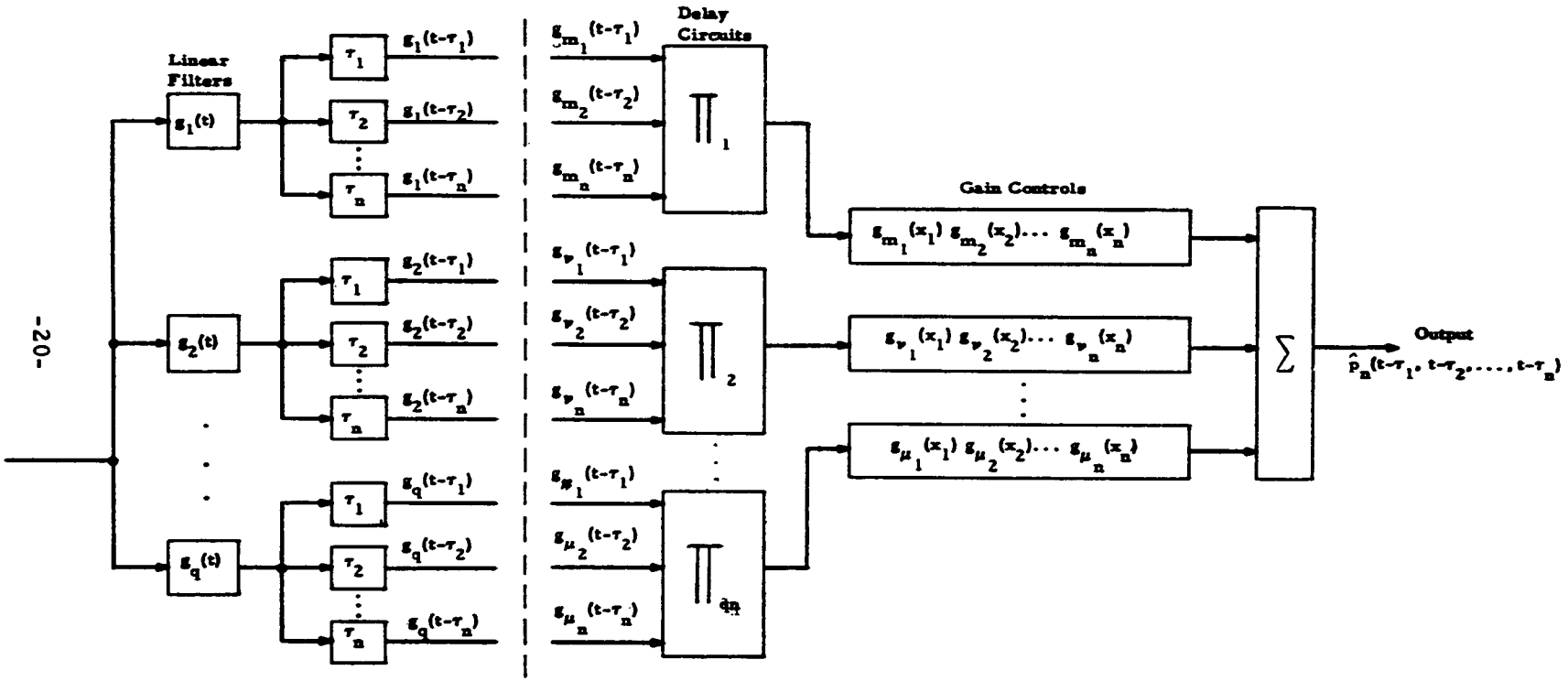


Figure 4. A Device for Displaying Two-Dimensional Cross Sections of $\hat{p}_n(\underset{\sim}{x})$

2.2.1.3 Methods of Estimating Averages of Functions of Signals

To clarify and justify the method we have proposed in the two preceding subsections for estimating averages of functions of stationary processes, we now consider a rather general question regarding the estimation of $\overline{f(\underset{\sim}{x})}$. Specifically, suppose we observe a single member of a stationary random process $\{s(t)\}$ for a limited observation time T. What is the best method of processing this data to obtain an estimate of $\overline{f(x)}$? Clearly the answer to this question will depend on the criterion of "best" and on the class of allowable operators on the process $\{s(t)\}$. We will consider the class of linear operators, and utilize the mean square error as an indication of estimation error.

As an example, suppose we are trying to estimate the mean of the process $\{s(t)\}$, according to the linear operator

$$\hat{m} = \int_T h(t)\ s(t)\ dt. \tag{23}$$

(Notice that as a special case, if h(t) were a comb of impulses, $\hat{m}$ would be formed as a sum of samples; thus, sampling is a subclass of the operators we are considering.) In order that $\hat{m}$ be an unbiased estimate, we should have (since $\{s(t)\}$ is a stationary process)

$$\overline{\hat{m}} = \int_T h(t)\ \overline{s(t)}\ dt = \int_T h(t)\ m dt = m \tag{24}$$

giving

$$\int_T h(t)\ dt = 1. \tag{25}$$

The problem then becomes that of minimizing the variance of $\hat{m}$ by choice of h(t) subject to the constraint above. A calculus of variations technique yields an integral equation for the optimum h(t):

$$\int_T h(\tau)\ R(t-\tau)\ d\tau = C,\ t\ \varepsilon\ T \tag{26}$$

where $R(\tau)$ is the autocorrelation function of $s(t)$ and C is a constant. Now $R(\tau)$ decays to a steady value for values of τ approximately $\frac{2}{W}$ or greater, where W is the bandwidth of the process $\{s(t)\}$. Then if $T >> \frac{2}{W}$, an approximate but very good solution, except for negligible end effects, is

$$h(\tau) = \frac{1}{T}, \quad \tau \in T. \tag{27}$$

This is not the exact solution; however for $TW >> 1$, it performs just about as well.* This solution indicates that one should not sample the waveform $s(t)$ at all, but use all of it according to

$$\hat{m} = \frac{1}{T} \int_T s(t)\, dt. \tag{28}$$

It may then be shown that the estimation error is approximated by

$$\sigma^2(\hat{m}) \cong \frac{1}{T} \int [R(\tau) - m^2]\, d\tau. \tag{29}$$

The same result is obtained for the problem of estimating the average value of any function of $\underset{\sim}{s}(t)$, if $R(t)$ and W are interpreted as the correlation function and bandwidth, respectively, of the function of $\underset{\sim}{s}(t)$. Thus the optimum linear method of estimating $\overline{f[\underset{\sim}{s}(t)]}$ is to integrate the function over the available time T, and divide by T.

To obtain some idea of the processing time, T, required to obtain an accurate estimation with this method, we now consider an alternate method which provides similar (but less) accuracy. Specifically, suppose that the n-th moment of a stationary process, μ_n, is estimated by sampling the n-th power of the process periodically and averaging the sample values. This is a linear estimation procedure with $h(t) = \frac{1}{T} \sum_{k=1}^{N} \delta(t-k\tau)$, i.e.,

*See [1] and [4]

$$\hat{\mu}_n = \frac{1}{T} \sum_{k=1}^{N} s_k^n \tag{30}$$

where $\{s_k\}$ is the sequence of N samples. This estimate is unbiased, and if the N samples are all statistically independent, the variance of the estimate is given by

$$\sigma^2(\hat{\mu}_n) = \frac{1}{N}(\mu_{2n} - \mu_n^2) \tag{31}$$

Now we do not know the $\{\mu_n\}$; in fact, we are trying to approximate them. However we can get a rough idea of how the variances $\sigma^2(\hat{\mu}_n)$ depend on n and N by assuming specific forms for $p_1(x)$ and calculating the dependence. In this manner, by combining the results, a rough quantitative estimate of the number of samples N necessary to evaluate the various moments will be obtained. We shall here assume only one particular form, Gaussian.

For a Gaussian distribution (zero mean)

$$\mu_n = \begin{cases} 0, & n \text{ odd} \\ \sigma^n (n-1)(n-3)\ldots(1), & n \text{ even} \end{cases} \tag{32}$$

where σ^2 is the variance of the process. The variance of the estimate is then obtained by substituting in the formula above. The relative accuracy is, for n even,

$$\frac{\sigma(\hat{\mu}_n)}{\mu_n} = \frac{1}{\sqrt{N}} \sqrt{\frac{\mu_{2n}}{\mu_n^2} - 1} = \frac{1}{\sqrt{N}} \left\{ \frac{\binom{2n}{n}}{\binom{n}{\frac{n}{2}}} - 1 \right\}^{1/2} \approx \frac{2^{\left(\frac{n}{2} - \frac{1}{4}\right)}}{\sqrt{N}} \tag{33}$$

Thus the number of terms, N, necessary to obtain accurate estimates of μ_n increases as n increases. We can, from the above formula, however, determine just how many samples are necessary to include higher order terms. For example, to include the sixth order term with a relative accuracy of 5 percent, 18,000 terms are necessary; of course, this number of terms gives better accuracy for lower order terms. (For n odd, values of N intermediate between values for neighboring even n values will be sufficient.)

For other forms of density functions, the number of terms, N, necessary for a prescribed accuracy will differ from those above, although not widely. For example, assumption of an exponential p. d. f. leads to values of N different by roughly a factor of two for low n.

An alternate method of establishing when a particular sample size N is adequate, which makes no presumptions about the form of the p. d. f. being estimated, is now discussed. This method has not been studied extensively, but it is a powerful and important technique which merits consideration. Suppose we are attempting to approximate μ_2. The variance of the estimate is then

$$\sigma^2 (\hat{\mu}_2) = \frac{1}{N} (\mu_4 - \mu_2^2) \tag{34}$$

Now we may put a bound on this quantity as follows: for a p. d. f. with a limited dynamic range and prescribed second moment, we must have

$$\mu_4 \leq \mu_2 B^2 \tag{35}$$

where B is the maximum value of $|x|$. This may be seen by noticing that

$$\mu_4 = \int_{-B}^{B} x^4 p_1(x)\, dx \leq \int_{-B}^{B} B^2 x^2 p_1(x)\, dx = B^2 \mu_2, \tag{36}$$

and in fact may be realized by

$$p_1^*(x) \left(1 - \frac{\sigma^2}{B^2}\right) \delta(x) + \frac{\sigma^2}{B^2} \delta(x - B). \tag{37}$$

Therefore

$$\sigma^2(\hat{\mu}_2) \leq \frac{1}{N} \mu_2 (B^2 - \mu_2), \tag{38}$$

and a bound on the variance of the estimate is obtained as a function of the statistic itself. Therefore, as N is increased, and $\hat{\mu}_2$ begins to stabilize, this value may be substituted in the right-hand side of the above equation to

also place a bound on the variance. Thus, estimation of μ_2 carries along with it an estimate of the variance of estimation! Notice that only a very crude pre-estimate of μ_2 need by evaluated this way. For example, we might take $N_1 = 100$, obtaining $\hat{\mu}_2^{(1)}$. For a prescribed $\sigma^2(\hat{\mu}_2)$, the above equation might then indicate $N_2 = 500$. As a check on this sample size, we can lastly compute, for $N_2 = 500$, $\sigma^2(\hat{\mu}_2^{(2)})$ and see if this is satisfactory. Notice that this method requires no presumption about the p.d.f. form, except for B, the dynamic range, which can be quickly and easily evaluated.

This method of bounding the variance of an estimate, by using approximate values of the statistic itself, can be extended to higher order moments. It is expected that the bounds obtained become relatively weaker as the order increases.

As an indication of the number of statistically independent samples which may be obtained from a process in time T, we may take the Nyquist rate: $\tau = \frac{1}{2W}$. Thus, the relative error in estimation of the n-th moment of a process $\{s(t)\}$ with bandwidth W, is approximately

$$\frac{\sigma(\hat{\mu}_n)}{\mu_n} \cong \frac{2^{\frac{n}{2}} - \frac{3}{4}}{\sqrt{TW}} \tag{39}$$

The inclusion of higher moments in an approximation to $p_1(x)$ is thus seen to impose a longer processing time for each moment in the estimation procedure. For instance, (39) indicates that to estimate second and fourth order moments of a 10 kc bandwidth process with a one percent relative error requires $T = 0.7$ second and $T = 5.7$ second, respectively. For a 100 cps process, these times are approximately 1 and 9 minutes, respectively.

The difference between the optimum and sampling estimation methods is probably not very great. For instance, for $R(\tau)$ parabolic between the origin and $\tau = \frac{1}{2W}$, it can readily be shown that

$$\frac{\sigma^2(\hat{\mu}_1)\ \text{(sampling method)}}{\sigma^2(\hat{\mu}_1)\ \text{(optimum method)}} = \frac{3}{2} \tag{40}$$

For higher order moments the difference may be larger, but it is expected that (40) provides a reasonable estimate of the relative error associated with an estimate of the n-th order moment of a process.

2.2.2 Success Counting Methods

Another approach to the problem of estimating either $p_n(\underset{\sim}{x})$ or $P_n(\underset{\sim}{x})$ consists of simply counting the number of "successes" in a number of trials of an experiment. For any process(es) it would be possible to obtain an n-tuple sample, $\underset{\sim}{s} = (s_1, s_2, \ldots, s_n)$, by simply initiating the process and recording the sample values taken at the appropriate times (s_1 at time t_1, etc.). In general, N statistically independent n-tuple samples of the process $\underset{\sim}{s} = (s_{k1}, s_{k2}, \ldots, s_{kn})$, $k = 1, 2, \ldots, N$, could be obtained by reinitiating the process after each sample. In this manner, it would be possible to perform an estimation of $p_n(\underset{\sim}{x})$ based on an examination of N members of the ensemble of waveforms constituting the process. Specifically, let

$$y_k = \begin{cases} 1 \text{ if } \underset{\sim}{s}_k \text{ falls in } R_{\underset{\sim}{x}, \Delta} \\ 0 \text{ if } \underset{\sim}{s}_k \text{ falls outside } R_{\underset{\sim}{x}, \Delta} \end{cases} \tag{41}$$

where the region $R_{\underset{\sim}{x}, \Delta}$ contains all values of $\underset{\sim}{x}'$ such that

$$\max_{1 \le \ell \le n} |x'_\ell - x_\ell| \le \frac{\Delta}{2} \tag{42}$$

and Δ is a (small) positive number. An estimate of $p_n(\underset{\sim}{x})$ is provided by

$$\hat{p}_n(\underset{\sim}{x}) \equiv \frac{1}{\Delta^n N} \sum_{k=1}^{N} y_k, \tag{43}$$

This estimate is the number of samples which fall in $R_{\underset{\sim}{x}, \Delta}$, divided by the total number of samples (properly normalized for the given Δ). One justification for this method is provided by the law of large numbers*, which states that

*[2], Chapter X.

$$\Pr\left\{\left|\sum_{k=1}^{N} y_k - \int_{R_{\underset{\sim}{x},\Delta}} p_n(\underset{\sim}{x})\, d\underset{\sim}{x}\right| > \varepsilon\right\} \longrightarrow 0 \tag{44}$$

as $N \to \infty$, where ε is any positive number,
Now this method of estimating the value of $p_n(\underset{\sim}{x})$ at any given point $\underset{\sim}{x}$ (as contrasted with earlier methods of fitting a curve to $p_n(\underset{\sim}{x})$ over all $\underset{\sim}{x}$) can be applied to all processes for which*

$$\frac{1}{N}\sum_{k=1}^{N} \overline{(y_k - \overline{y_k})^2} \longrightarrow 0 \tag{45}$$

as $N \longrightarrow \infty$.

A possible drawback to this brute force approach is that considerable time may be required to generate each sample, and therefore an exhorbitant delay may be encountered in completing the estimation of $p_n(\underset{\sim}{x})$ at many points $\underset{\sim}{x}$. If the process is nonstationary, then there may be no substitute for this direct approach. However, if the process is stationary, then instead of dealing with samples from many different members of the ensemble of time functions, an estimate could be based on samples taken from a single member of the ensemble.

The question naturally arises as to whether other procedures (besides sampling) would be better for processing T seconds of a set of signals $\underset{\sim}{s}(t)$ to obtain an estimate of $p_n(\underset{\sim}{x})$. Following the success counting approach we now consider this question, restricting the processing methods to linear operations. Specifically, to estimate $p_n(\underset{\sim}{x})$ as a constant** in the region $R_{\underset{\sim}{x},\Delta}$ we define

$$p_n^*(\underset{\sim}{x}) \equiv \frac{1}{\Delta^n} \Pr\left\{\underset{\sim}{s} \in R_{\underset{\sim}{x},\Delta}\right\}, \tag{46}$$

---

*Ibid., p. 238.

**This estimation will produce what is called a histogram approximation to $p_n(\underset{\sim}{x})$.

A "success" counting function can again be defined such that

$$y = f(\underset{\sim}{s}) = \begin{cases} 1, \text{ if } \underset{\sim}{s} \text{ falls in } R_{\underset{\sim}{x}, \Delta} \\ 0, \text{ if } \underset{\sim}{s} \text{ falls outside } R_{\underset{\sim}{x}, \Delta} \end{cases} \tag{47}$$

where the n-tuple $\underset{\sim}{s}$ is actually a function of time.
The function $f(\underset{\sim}{s})$ registers whenever $\underset{\sim}{s}(t)$ falls in the region $R_{\underset{\sim}{x}, \Delta}$. Again let Δ be a (small) positive number. Then, the most general linear operation on the limited data provided by $f[\underset{\sim}{s}(t)]$ in an interval T is

$$\hat{p}_n(\underset{\sim}{x}) = \int_T h(t)\, f[\underset{\sim}{s}(t)]\, dt \tag{48}$$

(Notice again that if h(t) were a comb of impulses, $\hat{p}_n(\underset{\sim}{x})$ would be a sum of samples -- the usual "counting" method.) An unbiased estimate has

$$\begin{aligned} \overline{\hat{p}_n(\underset{\sim}{x})} = p_n^*(\underset{\sim}{x}) &= \int_T h(t)\, \overline{f[\underset{\sim}{s}(t)]}\, dt = \overline{f(\underset{\sim}{s})} \int_T h(t)\, dt \\ &= \int_{R_{\underset{\sim}{x}, \Delta}} 1 \cdot p(\underset{\sim}{x})\, d\underset{\sim}{x} \int_T h(t)\, dt = Pr(\underset{\sim}{s} \in R_{\underset{\sim}{x}, \Delta}) \int_T h(t)\, dt \end{aligned} \tag{49}$$

Using eq. (46), this provides the constraint:

$$\int_T h(t)\, dt = 1/\Delta^n \tag{50}$$

If we minimize $\sigma^2 [\hat{p}_n(\underset{\sim}{x})]$ by choice of h(t), we obtain the following integral equation for the optimum h(t):

$$\int_T h(\tau)\, R_f(t-\tau)\, d\tau = c, \quad t \in T \tag{51}$$

where

$$R_f(t-\tau) = \overline{f[\underset{\sim}{s}(t)]\ f[\underset{\sim}{s}(\tau)]}. \tag{52}$$

Again, the correlation of $\underset{\sim}{s}(t)$ extends only over roughly an interval $\frac{2}{W}$, where W is the bandwidth of $\underset{\sim}{s}(t)$. If $T >> \frac{2}{W}$, an approximate solution for h(t) is

$$h(t) = \frac{1}{T\Delta^n}, \quad t \in T. \tag{53}$$

Thus when TW >> 1 (many Nyquist intervals in the observation interval), optimum estimation is obtained by the equation

$$\hat{p}_n(\underset{\sim}{x}) = \frac{1}{\Delta^n T} \int_T f[\underset{\sim}{s}(t)]\, dt. \tag{54}$$

This estimate is the percentage of time that $\underset{\sim}{s}(t) \in R_{\underset{\sim}{x}, \Delta}$ while $t \in T$, weighted according to the value of Δ chosen.

In summary, if TW >> 1, the best way to use a given amount of continuous data (i.e., a segment of $\underset{\sim}{s}(t)$) for estimating $p_n^*(\underset{\sim}{x})$ is to weight all of it equally.*

To implement this method would require the construction of a device consisting of an integrator followed by a multiplier. The integrator would be required to have a wide dynamic range in both output voltage and integration time. The output voltage dynamic range is a function of the dynamic range of the probabilities to be measured. A 40 db range would be a minimum requirement. The integration time is a function of (a) the bandwidth of the signals being measured, (b) desired accuracy requirements, and (c) the probability being measured.

*The same result was obtained for other statistics (averages of functions of $\underset{\sim}{x}$) of stationary signals in Section 2.2.1.3 above.

As the number of cells (used to cover the signal space) increases, the probability that a set of signals will fall in a given cell decreases; e.g., for a 4-th order uniform probability density function with 3 bit quantization in each dimension, the probability that the input signals will fall into a given cell is 1/4096. If such cells as well as those with probability near unity are of interest, and if, in addition, a wide range of signal bandwidths are to be considered, then the dynamic range of the integration time would be greater than 10^4. The construction of an integrator within tolerances imposed by these requirements would be a formidable task.

Fortunately, the difficulties associated with the optimum success counting method of estimation can be overcome through the use of a non-optimum, but still efficient technique involving sampling the signals, $\underset{\sim}{s}(t)$. Specifically, the optimum method of estimating the probability that $\underset{\sim}{s}(t)$ falls in any specified region*, R, may be approximated by a generalization of eq. (43):

$$\Pr\{\underset{\sim}{s} \in R\} = \frac{1}{N} \sum_{k=1}^{N} f[\underset{\sim}{s}(t_k)]. \tag{55}$$

where $f(\underset{\sim}{s}) = 1$ if $\underset{\sim}{s} \in R$, and $f(\underset{\sim}{s}) = 0$ if $\underset{\sim}{s} \notin R$.

That this estimate is nearly equivalent to the optimum one is indicated by equation (40). The utility of this method follows from the use of digital circuitry. The most significant advantage derived from the use of digital circuitry is in the flexibility of the digital integrator (counter): the dynamic range and resolution can be doubled by the addition of a single stage; the integration time can be varied either by the addition of counter stages or by changes in clock frequency; and there are no drift problems. Some fringe benefits derived from the use of digital circuitry are: (a) the success regions can be conveniently stepped automatically from one cell to another; (b) the results can be easily presented in a form convenient for entry to a computer; (c) the output data can be easily normalized to provide equal resolution of both low and high probabilities.

2.2.2.1 Parallel Processing With a Digital Computer

Having decided that periodic samples of a set of signals $\underset{\sim}{s}(t)$ will form the basis of an estimation, there are two alternative ways of processing these samples. The first, called parallel processing (Figure 5), consists of

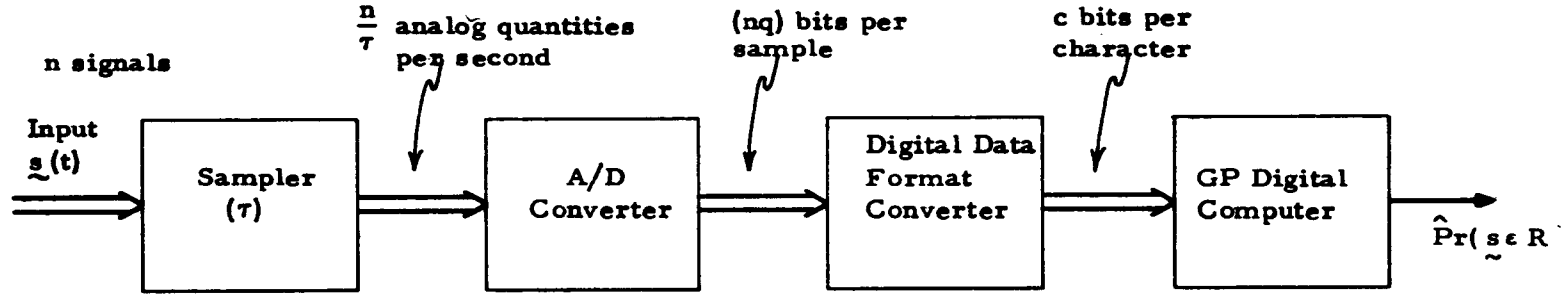


Figure 5. Parallel Success Counting Probability Analyzer

sampling $\underset{\sim}{s}(t)$, analog to digital conversion of the sample values, changing the digital format, and inserting the resulting data into a general purpose digital computer. The computer would then be programmed to count all the successes occurring in specified regions to produce an estimate of $p_n(\underset{\sim}{x})$, $P_n(\underset{\sim}{x})$ or, in general, $\Pr\{\underset{\sim}{s} \in R\}$

To estimate the practical limitations imposed on this technique by the characteristics of a general purpose computer, suppose the available main storage capacity is equal to 4096 computer words.* With this capacity, the data will have to be read into the computer in segments. Assuming that half of the storage capacity is used for recording a histogram estimate of $p_n(\underset{\sim}{x})$, and the other half is used to store a segment of data, the signal space X may be partitioned into at most $\frac{(2048)w}{\nu}$ cells, where each computer word is composed of w bits, and $\hat{p}_n(\underset{\sim}{x})$ is recorded with ν-bit accuracy per cell. With the CDC 160 computer, $w = 12$ bits per word, and the total number of cells allowable by a 6-bit quantization of $\hat{p}_n(\underset{\sim}{x})$ in each cell, is approximately 2^{12} cells. To estimate an n-th order probability density function with a histogram based on a cell structure with q-bit accuracy in each variate, a total of 2^{nq} cells are required. Thus, we obtain the restriction:

$$nq \leq 12 \tag{56}$$

A value of $q = 3$ would seem to be a minimum resolution capability; this quantization would allow for the computation of up to fourth order probability density functions.

The time required to collect and process data with this technique is, of course, largely determined by the computer utilized. However, the cost of operating the computer may be approximately the same for different computers. As an illustration, consider the Recomp II and the problem of estimating a fourth order probability density function with 3-bit quantization per variate, and 5-bit accuracy of representing $\hat{p}_n(\underset{\sim}{x})$ in each cell. By assigning two of the $2^{4 \cdot 3} = 4096$ cells to a single computer word, approximately half of the main storage (2048 words) is made available for raw digital data input. Since each sample consists of 12 bits, approximately 6000 samples

*As is the case with the Packard Bell 440, Recomp II, CDC 160, and other computers.

of data can be handled in a single, continuous segment. The input data rate is limited for this computer to 60 words per second, or 180 samples per second; thus each segment of data can be read into the computer in approximately thirty seconds. If 2^{19} samples are processed*, then approximately 80 segments of data would have to be processed. The total read-in time would therefore be less than one hour. The processing time to calculate probabilities would probably be at most an order of magnitude longer.

Implementation of the parallel processing technique requires only the construction of sampling, analog-to-digital conversion, and digital format conversion equipment. However, a possibly major drawback to this method of probability estimation is its inherent reliance on a general purpose computer. If a computer is not available, then a probability estimation cannot be accomplished. Another unattractive feature of this method is the requirement to obtain and store a large quantity of data in the process of obtaining a probability estimate. Unless the signal samples can be fed directly into the computer, some form of intermediate storage is required. Assuming again the necessity for 2^{19} samples**, and using 3-bit quantization of each of four signal amplitudes, more than 6 million bits would have to be stored for each probability calculation. If paper tape is used as the intermediate storage medium with 6-bit characters, then more than eight thousand feet of tape (or sixteen 500 foot rolls) would be required for each probability calculation. In addition to the possibility of introducing errors, the increased processing time for punching and reading the data would be significant -- as much as 5 hours for punching.

Because of the desire (for this study) to make the probability analyzer independent of other computational aids, and the intermediate storage problem, we have chosen the other method of implementing the success counting probability estimating technique, namely serial processing with a completely self-contained device.

---

*A number estimated (in Section 2.3.2) to be sufficient for a cell structure with 4096 cells.

**Ibid

2.2.2.2 Serial Processing with a Self-Contained Device

Construction of a self-contained device for simultaneously calculating the number of samples which fall into each of a large number of cells in signal space is precluded by its cost, if the number of cells, c, is very large (as it will be for $n > 1$). We must therefore be content with a device which utilizes a period of time, T, for estimating the probability that $\underset{\sim}{s}$ falls in a single cell, and repeats the process for every other cell examined. The block diagram of such a serial estimation device is shown in Figure 6. The first operation performed on $\underset{\sim}{s}(t)$ (by the success region detector) is the registration of intervals during which $\underset{\sim}{s}(t)$ falls in an n-dimensional region, R, as dictated by equation (55). The result of this operation is a success waveform, which takes on the value 1 at times when $\underset{\sim}{s}(t) \in R$, and zero at other times. This waveform is sampled periodically (roughly at the Nyquist rate) until a specified number of samples, or trials, have been taken. The number of successes obtained and samples taken are recorded in a Success Counter and Trial Counter, respectively. By digital operations on the contents of these two counters, a number representing the ratio of successes, n_s, to total number of trials, n_t, is obtained. This quantity is the estimate of the probability that $\underset{\sim}{s} \in R$.

The flexibility of this type of device is unsurpassed by any of the other methods of estimating higher order statistics of signals. By adjustment of the region R, any one of $\hat{p}_n(\underset{\sim}{x})$, $\hat{P}_n(\underset{\sim}{x})$, or in general, $\hat{P}r\{\underset{\sim}{s} \in R\}$ may be calculated. Also, there is no limitation to stationary signals (although samples may no longer be obtained at the Nyquist rate from nonstationary signals*). Perhaps most important, either partial or preliminary results can be obtained without having to perform a complete calculation of $\hat{p}_n(\underset{\sim}{x})$: the value of $\hat{p}_n(\underset{\sim}{x})$ can be ascertained in a few selected cells, or the cell size can be set at a large value to obtain a coarse, but possibly informative, preliminary histogram.

With this method of probability estimation, a significant intermediate data storage problem will never arise, even if further processing of the results of an estimation are desired. At worst, the results of a probability density estimation will produce $(c\nu)$ bits of data, where c is the total number of cells and $\hat{p}_n(\underset{\sim}{x})$ is represented by ν-bit quantized numbers. Assuming the previously mentioned numbers ($c = 12$ and $\nu = 6$), storage of a complete probability density function would require less than one-tenth of a roll of tape.

---

*The effects of nonstationarity of signals on accuracy of estimation are discussed in Section 2.3.4.

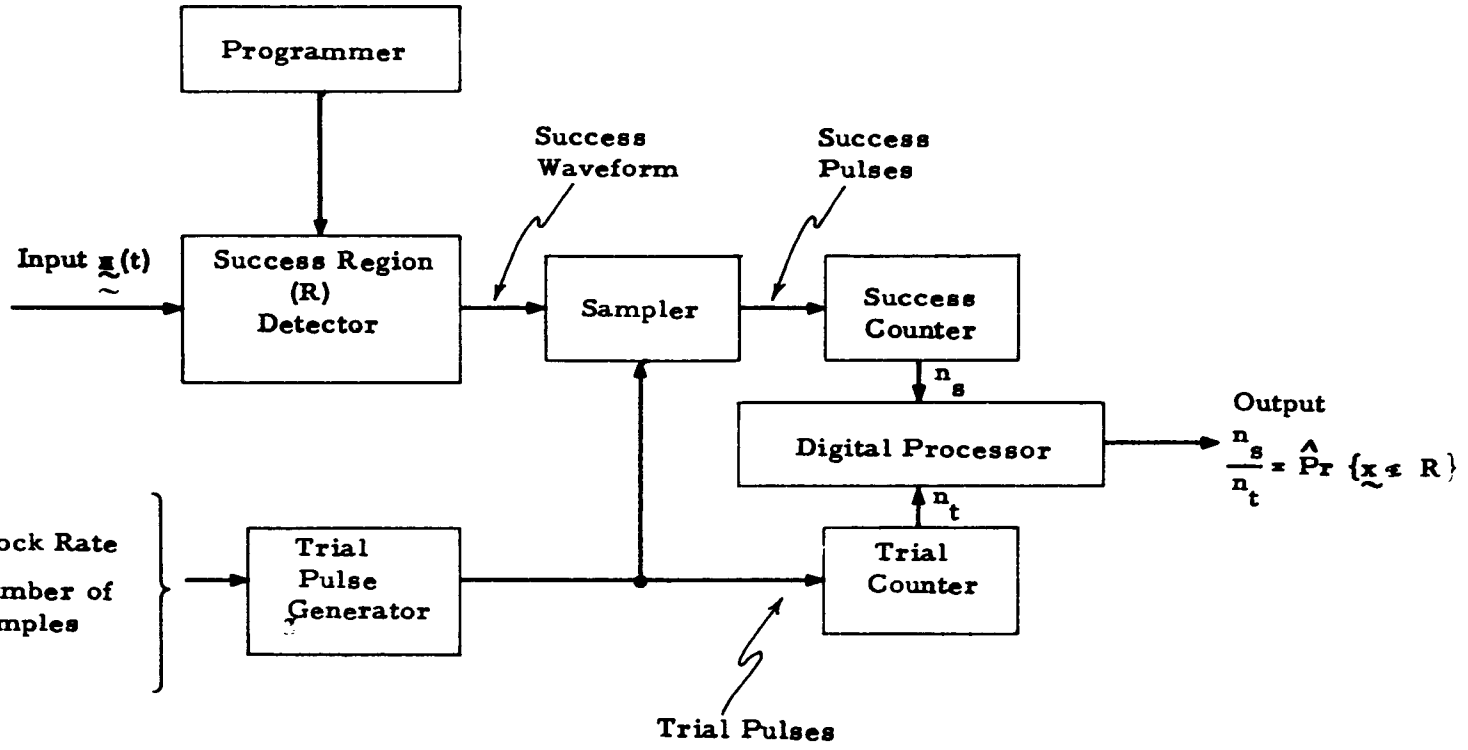


Figure 6. Block Diagram of Serial Success Counting Probability Analyzer

The success region detector can be implemented in many different ways. One fairly general approach to this problem is outlined in the next subsection. Following this is a statement of the manner in which the rate at which samples should be taken from a set of signals to minimize processing time. Following this, in Section 2.3, is a discussion of sources and potential magnitudes of errors associated with the serial processing success counting method of estimation.

2.2.2.3 Establishment of a Success Region

As indicated in Figures 5 and 6, the initial step toward estimating the probability, $\Pr\{\underset{\sim}{s} \in R\}$, of a set of signals falling in a given region, R, is the establishment of a success region detector. In practice, the region of interest may take on a variety of forms. In error probability calculations, for instance, it is often the case that R is defined as all values of $\underset{\sim}{x}$ for which $x_1 \geq x_j$, $j = 2, \ldots, n$. Another region of interest is that for which

$$\sum_{j=1}^{n} x_j^2 \leq K.$$

In general, it is desired that a success region detector be flexible enough to accommodate all conceivable forms of the region, R, but of course this is precluded by the cost of such a device.

One compromise which is relatively easily implemented consists of defining regions with hyperplanes, i.e., linear inequalities. With this method a region R is defined as all points $\underset{\sim}{x}$ for which

$$\sum_{k=1}^{n} x_k a_{jk} > C_j, \quad j = 1, 2, \ldots, m, \tag{57}$$

where the $\{a_{jk}\}$ constitute a set of mn constants which correspond to the region, R. To indicate whether a sample value $\underset{\sim}{s}$ falls in a given region, R, it is sufficient to build m resistive adders and comparators, whose outputs are routed to an AND gate. A device for implementing these operations for n=2 and m=3 is indicated in Figure 7.

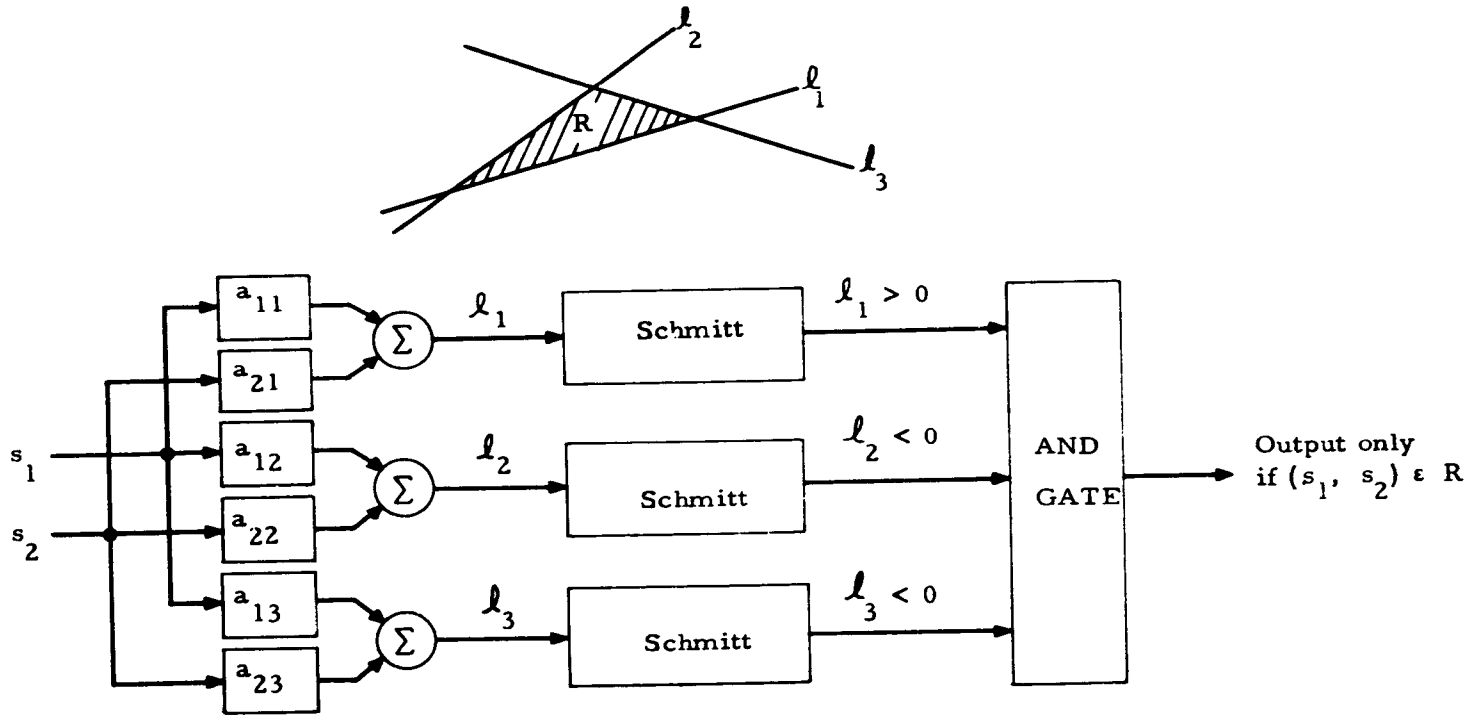


Figure 7. Circuit for Registering $\underset{\sim}{s} \varepsilon R$.

The accuracy with which a region R can be approximated using hyperplanes is quite dependent on the shape of the region. Obviously any region which is bounded by hyperplanes can be implemented perfectly. For instance, the regions associated with probability distributions or histogram estimates of probability density functions (with hypercubic cells) can be realized precisely, if the number of hyperplanes, m, is greater than or equal to the number of variates, n, for distribution functions, or twice that number, 2n for density functions. For regions bounded by curved surfaces, however, errors in specifying the regions will result.

As an indication of the number of hyperplanes required to approximate success regions, the ratios of the volume of inscribed and circumscribed hyperspheres to the volume of a regular n-tope are shown in Table 1 for $n = 2$ and $n = 3$. For $n = 2$, the approximating region is a regular polygon, and for $n = 3$ the approximating region is a regular polyhedron (of which there are only 5).

Table 1. Ratios of Volumes of Polytopes and Hyperspheres for n=2, n=3, and Several Values of m

	Polygon Approximation		Polyhedron Approximation	
m	Circumscribed	Inscribed	Circumscribed	Inscribed
3	2.42	.62		
4	1.57	.79	8.15	.30
5	1.32	.86		
6	1.21	.91	2.72	.52
7	1.14	.93		
8	1.11	.95	3.14	.60
9	1.09	.96		
10	1.07	.97		
11	1.06	.97		
12	1.05	.98	1.50	.75
20	1.00	1.00	1.65	.82

For two dimensional regions, a dozen straight lines would probably suffice for most regions of interest, but for three or higher dimensional regions, perhaps a few dozen hyperplanes might be needed for accurate representation of an arbitrary region. In practice, however, by using only m = 2n hyperplanes, a histogram can be constructed over which an integration could be carried out for any region, with accuracy limited only by the cell size chosen for the histogram. Thus, construction of a success region detector with more than 2n hyperplanes would probably not be justified unless a specific region is to be investigated for a variety of inputs.

2.2.2.4 Sampling Rate Adjustment

In practice, selection of an appropriate sampling rate poses a problem. If a set of signals, $\underset{\sim}{s}(t)$, is sampled at a high rate, dependent data are obtained which carry very little statistical information, perhaps even in a large amount of data. On the other hand, sampling at too low a rate, although yielding independent data, requires a long data collection time. The guide to selection of an intermediate sampling rate is given by the sampling theorem*; if a process is bandlimited to W cycles per second, samples taken $\frac{1}{2W}$ seconds apart just suffice to reconstruct the time function exactly. That is, this rate of sampling, 2W samples per second, does not miss any of the "important changes" in the time function, yet does not yield a large amount of superfluous data.

In order to apply this theorem to the problem at hand, the bandwidth of the process must be known. A quick rough overestimate of the bandwidth W, perhaps by means of a spectrum analyzer, would suffice to allow for proper adjustment of sampling rate.

*[6]

2.3 ACCURACY ATTAINABLE WITH THE SUCCESS COUNTING METHOD

The accuracy with which a probability distribution or density function can be estimated using the success counting method is limited by the quantization of signal space, the processing time, and equipment inaccuracies.

2.3.1 Quantization Error

Consider first the effect of quantization. If unlimited processing time is available, and no equipment inaccuracies exist, then the true probability that a set of signals lies in a particular region R in signal space can be made arbitrarily close to the ratio of (a) the number of trials which produced values of the signal set within the region, n_s, to (b) the total number of trials, n_t. Explicitly, if n_t can be made arbitrarily large, then the law of large numbers states that it is possible to make

$$\left| \int_R p_n(\underset{\sim}{s})\, \underset{\sim}{ds} - \frac{n_s}{n_t} \right| < \epsilon \tag{58}$$

for any given $\epsilon > 0$, and any region R in signal space. The histogram height, h, for a given cell in signal space is defined as $h = \frac{n_s}{n_t \Delta v}$, where Δv is the volume of the cell. The inequality (58) indicates that the volume under the histogram for a cell is the same as the volume under the true probability density function, i.e., the true probability of the signals falling in the cell.

For any region in signal space whose boundary lies only on boundaries between cells, the probability of a set of signals falling in this region can be estimated arbitrarily closely with the success counting method. However, for a region whose boundary deviates from cell boundaries, the estimated probability of a set of signals falling in the region (obtained by integrating under the histogram) can differ from the true probability. To obtain some feeling for the adequacy of a given quantization and cell structure, we may introduce a quantity which provides some indication of the degree to which the estimated probabilities (obtained with histograms) differ from the true probabilities. One such quantity is

$$D = \frac{1}{V} \int_R \left| P_n(\underset{\sim}{x}) - \hat{P}_n(\underset{\sim}{x}) \right| \underset{\sim}{dx} \tag{59}$$

where

$$V = \text{the total volume of signal space}$$

$$= R_1 R_2 \ldots R_n$$

$$P_n(\underset{\sim}{x}) = \Pr\{s_1 \leq x_1,\ s_2 \leq x_2, \ldots, s_n \leq x_n\}$$

$$\hat{P}_n(\underset{\sim}{x}) = \int h(\underset{\sim}{s})\, \underset{\sim}{ds}$$

$$(s_i \leq x_i,\ i = 1, 2, \ldots, n)$$

and R denotes the region in which signals can arise. The quantity D measures the average absolute deviation of the estimated probability, $\hat{P}_n(\underset{\sim}{x})$, from the true probability distribution, $P_n(\underset{\sim}{x})$. For simplicity of illustration, consider the single variate case, and assume that the range R_1 is partitioned into c cells having the same width, Δ. Then, the distribution function, $P_1(x)$, is a function of the single variable, x, and a histogram can be constructed which provides a piecewise linear approximation to $P_1(x)$, as indicated in Figure 8, with $c = 8$. As noted above, the estimated probability is equal to $P_1(x)$ for $x = i\Delta$, $i = 1, 2, \ldots, c$, and generally differs from $P_1(x)$ at other values of x.

The value obtained for the quantity D is dependent on the probability distribution function. From the standpoint of maximizing D, the worst probability distribution consists of a staircase function with jumps at the cell boundaries, as indicated by the dashed curve in Figure 8. For this distribution function, the average absolute difference between the true probability distribution and the estimated value of this quantity, is equal to $\frac{1}{2c}$. Thus to obtain an average deviation of less than 0.1 (in probability), it is necessary only to choose $c \geq 5$. In general, for an n-variate distribution, the average absolute difference between the true probability distribution and the estimated quantity is less than or equal to $\frac{1}{c}\left[1 - \left(\frac{1}{2}\right)\right]^n$, where c is the total number of (equi-volume) cells in the signal space.

2.3.2 Error Caused By Finite Processing Time

As pointed out earlier, if unlimited processing time is available, then the ratio of the number of "successes" to the number of samples, or trials, can be made arbitrarily close to the probability of a set of signal amplitudes falling in a region R. The difference between the estimated and actual values will tend to be greater for shorter processing times.

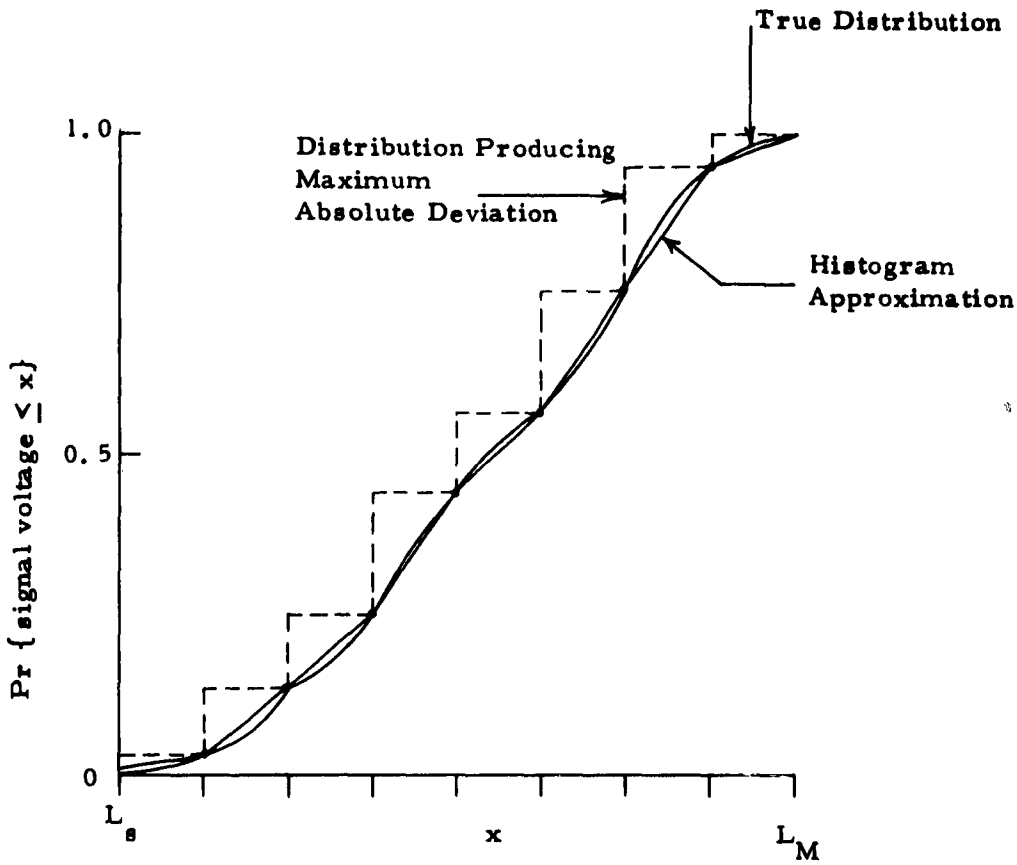


Figure 6. Probability Distribution Function Error Introduced By Quantization

For the success counting method of estimation, the processing time, T, can be related to the number of statistically independent* samples available in that time. Specifically, a process $\{x(t)\}$ which is limited in bandwidth to a range W provides roughly $N = 2TW$ independent samples in a time, T. Therefore, to determine the error imposed by a given limited processing time, it is sufficient to relate the error of an estimation to the number of independent samples used.

The customary method of measuring the error of an estimate, $\hat{P}$, of a probability, P, is in terms of the average square of the difference between these quantities:

$$e^2 = \overline{(\hat{P} - P)^2} = \overline{\left(\frac{N_s}{N} - P\right)^2} \tag{60}$$

where N_s is the number of trials resulting in successes. The quantity N_s has a binomial probability den. ity function,

$$p(N_s) = \sum_{k=o}^{N} \binom{N}{k} P^k (1-P)^{N-k} \delta(N_s - k) \tag{61}$$

and the quantity e is readily calculated:

$$e = \sqrt{\frac{P(1-P)}{N}} \tag{62}$$

A more demanding measure of error which would reflect the relative accuracy attained for any probability P, is provided by

$$e_r = \frac{e}{P} = \sqrt{\frac{1-P}{PN}} \cong \sqrt{\frac{1}{S_e}} \tag{63}$$

where S_e = the expected number of successes in N trials. For a given number of trials, (63) indicates that the relative error is much larger for small probabilities.

It would, of course, be desirable to be able to know in practice precisely the number of samples required to estimate an unknown probability with a specified relative accuracy. However, from (63) we know only the relationship between the

*[2], p. 114.

error and the (unknown) true probability. One way to obtain an estimate of the error (in estimating P) is to substitute the estimate, $\hat{P} = \frac{N_s}{N}$, into the formula for e_r, obtaining

$$e_r \approx \sqrt{\frac{N - N_s}{N N_s}} \tag{64}$$

For small values of P (the more difficult or time consuming range), we would hope to have $N_s << N$ and (64) would then become

$$e_r \approx \sqrt{\frac{1}{N_s}} \tag{65}$$

Thus, to control the relative estimation error, it might be reasonable and economical to operate the success counting device for the period of time it takes to produce a specified number of successes.

Although this approach to error control is appealing at first sight, we have decided not to use it. The primary reasons for not using it are (a) the expression (65) is valid only if the accuracy of estimation is rather high anyway; (b) the actual processing time would be a randomly fluctuating quantity; and (3) for signals with bandwidth in the range 5 kcps - 10 kcps, the error can be controlled in a way which doesn't involve these difficulties.

A convenient,and perhaps more realistic, method of controlling the estimation error due to limited processing time, consists of simply incorporating a wide dynamic range in the success and trial counters. If the input signal bandwidth is so small, or the number of cells so large that an intolerably long processing time is involved in computing probabilities with the maximum number of trials available, then a succession of calculations using fewer trials could usually establish the number above which the estimate can be expected to remain essentially unchanged. While this procedure would economize on processing time, there may exist probability calculations which cannot be accomplished with a given maximum number of trials (except by combining the results of several computations using independent data).

To determine the maximum number of trials it would be reasonable to attempt to record in the trial counter, we must consider the lowest probability we would ever want to compute. A reasonable assumption for the latter quantity is the probability of a set of signals falling in a single histogram cell defined for an n-th order process, where n is the highest order to be considered. It was suggested earlier that it would not be reasonable to attempt to calculate more than fourth order statistics using a cell structure which quantizes each variate into eight intervals. Thus, no more than $2^{12} = 4096$ cells would be involved. If the minimum value of P to be encountered in any practical application is assumed to correspond roughly to the uniform distribution of fourth order signal amplitudes over all of these 2^{12} cells ($P = 2^{-12}$), and if a relative error of at most ten percent is allowed for probabilities down to 2^{-12}, then (63) indicates that the number of trials should be on the order of $2^{12} \cdot 100 \cong 2^{19}$. The trial and success counters in the success counting probability analyzer described in Section 3, have been designed to accommodate this number of samples (or fewer).

2.3.3 Errors Caused By Equipment Inaccuracies

The two major types of errors introduced by imperfect implementation of the success counting method are (1) warping of the success waveform, and (2) sampling discrepancies resulting from the use of non-zero width sampling pulses. Consider first the sources and nature of the first type of error. One source of error in producing a success waveform is the improper location of boundaries of the given region. In general, the error in estimation of a probability which results from improper location of success region boundaries is dependent on the nature of the probability density function of the signals involved. For a uniform probability density function defined over an n-dimensional cell structure consisting of equivolume cells, specification of cell boundaries with α percent accuracy results in a possible probability density estimation error of $n\alpha$ percent. Fortunately, even for the most fine-grained quantization desired (8 levels per dimension), it is quite easy to establish a 1 percent accuracy of boundary placement, which results in a 4 percent error in fourth order probability density function estimation.

Other sources of success waveform distortion are switching delays in the Schmitt triggers, varying delays in different input channels, and logic delays. All of these delays serve to create shifts, compression, or expansion of the success waveform. These effects are portrayed in Figure 9 for a sinewave, using a cell width equal to one-fourth of the peak-to-peak amplitude of that waveform. The ideal success waveform consists of pulses with width w, spaced one period apart. The actual success waveform consists of pulses with one of two widths, w_1 and w_2, spaced approximately one period apart. Every other

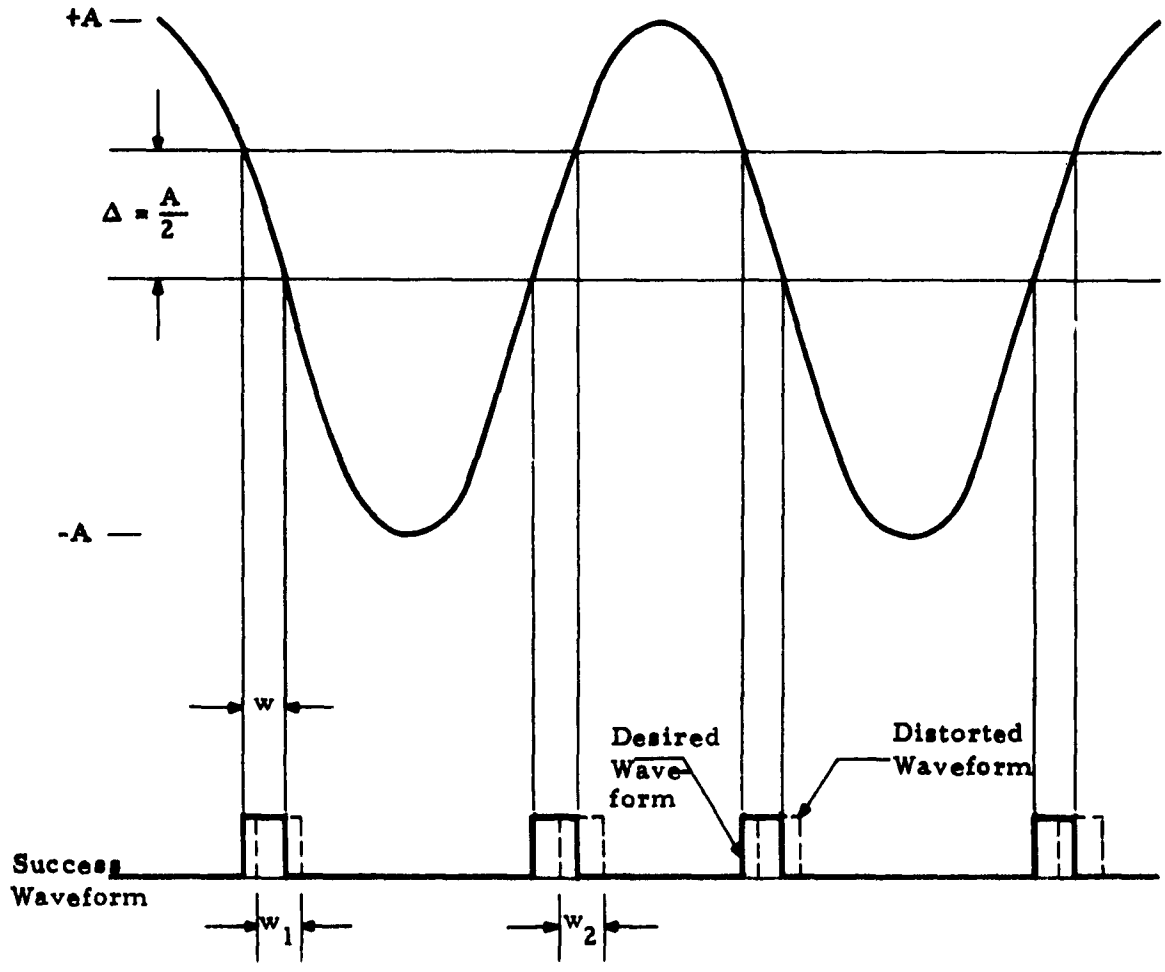


Figure 9. Success Waveform for a Sinewave Input

pulse has width w_1, corresponding to the upward movement of the sinewave through the success region, and these pulses are interlaced in time with pulses with width w_2, corresponding to the downward movement of the sinewave through the success region. If many samples are taken at a rate which is incommensurate with the frequency of the sinewave, then translation of the pulses in the success waveform will not affect a first-order estimation at all. For higher order statistics, differences in translations in success regions for different variates may tend to "smear" a probability density estimation. However, as long as these differences are small in comparison with the minimum success waveform pulse width, the effect on an estimate is negligible. Calibration results which illustrate this effect are presented in Section 3.

The magnitude of the difference between the desired result and the actual result obtained for the probability that the sinewave falls in the given region is $\left| \frac{1}{2} [w_1 + w_2] - w \right|$. In practice, the difference in switching times of different Schmitt triggers will be the major cause of deviations in w_1 and w_2 from w. However, for signals and regions for which roughly as many upward traversals are produced as downward traversals, it is easily seen that error cancellation takes place: $w_1 + w_2$ will very nearly equal 2w. This cancellation effect takes place whether or not the desired success waveform pulse widths are fixed. Therefore, the error introduced by distortion in the success waveform can be considered insignificant. This conclusion is verified by calibration results presented in Section 3.

The other type of error is introduced through the use of non-zero width sampling pulses. For a sampling pulse width δ, and a given width of a success waveform pulse, w, the maximum effective change in the success waveform produces a relative error of $\frac{\delta - 2\mu}{w}$, where μ is the minimum overlap between a counting pulse and a success waveform pulse which will be counted as a success. This can be seen by noting that the counting pulse can be located at any point in an interval of duration $w + \delta - (2\mu)$ to cause a success to be registered. To determine the net effect of this counting error on a probability estimate would require that the quantity $\frac{\delta - 2\mu}{w}$ be averaged over all possible values of success waveform pulse width, w, weighting each value by the probability of that value occurring. Except for deterministic signals, this problem evidently cannot be solved.

Fortunately this calculation can be circumvented by examining two special cases: (1) the deterministic signal and success region which produces the smallest value for w (and therefore the largest estimation error), and

(2) a random process with known characteristics. In each case we want to know the smallest value of w which can be expected to occur.

The deterministic signal which produces the shortest success waveform is a sinewave with the maximum frequency, W, for which it is desired to obtain probability estimates. In this study, W = 10 Kcps. For a sinewave, the minimum success waveform pulse width, w_o, is given by approximately

$$w_o = \frac{1}{\pi c W} = \frac{10^{-4}}{\pi c} \text{ seconds} \tag{66}$$

where c = the number of equi-amplitude cells covering the sinewave amplitude range. For c = 8, this minimum pulse width is $w_o = 4\ \mu$sec. Thus, if the sampling pulse width δ, is equal to 2 μsec, and the minimum overlap between a counting and success waveform pulse which will produce a success is 0.9 μsec, the relative error in estimation will be at most 5 percent. It should be pointed out that this is an upper bound for the worst combination of deterministic signal and success region. For other signals and success regions, the relative error should be less than that implied by (66).

To check this result for a specific noise waveform, the probability that the magnitude of the slope of a bandlimited Gaussian signal will be greater than a given value, α, has been found

$$\Pr\{\,|\text{slope}| > \alpha\,\} \cong 2\,\Phi\left(-\frac{3\alpha}{\pi R W}\right) \tag{67}$$

where $\Phi(x) \equiv \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\,du$, and where R is the range of signal amplitudes being considered, and W is the noise bandwidth. By substituting the value $\alpha = \frac{R}{cw}$ in (67), the probability of obtaining a success waveform pulse width less than w can be written* $\Pr\{\text{success pulse width} < w\}$

$$= 2\,\Phi\left(-\frac{3}{\pi c w W}\right) \tag{68}$$

*Assuming that the noise waveform does not leave the success region at the same extremity through which it entered the region.

The probability that a success waveform pulse will be generated with a width less than the minimum pulse width of a deterministic signal ($w = w_o$), is therefore

$$\begin{aligned} &\Pr\{\text{success pulse width} < w_o\} \\ &\quad = 2\,\Phi(-3) \\ &\quad = 2.6 \times 10^{-3} \end{aligned} \tag{69}$$

Thus, it appears unlikely that a greater counting error will result with a noisy waveform than is obtained with the high frequency sinewave. Results of calibration tests using a sinewave are reported in Section 3.

2.3.4 Effects of Non-Stationarity of Signals

A few comments on the effects of non-stationarity of a process on a probability density function estimate are appropriate at this point. First, for a stationary process, as pointed out in the earlier sections, the accuracy of p.d.f. estimates is limited primarily by the processing time, improving as the processing time increases. However, for a non-stationary process, additional errors can result, depending on the rate of fluctuation of the statistics. There are three types of situations which can arise: where the time constant of the statistical (non-stationary) fluctuations is small, intermediate, and large, respectively, compared with the time required to estimate the entire p.d.f.

For the first case, where the time constant is small, the process will pass through all its "modes" many times during the processing interval. In this case, the output of the estimating device is an average p.d.f. of the actual input process p.d.f. If the estimation procedure were duplicated on a different serial section of input data, the same average p.d.f. would result, and an observer would not even be aware of the rapid statistical changes in the input process. This is not necessarily a deleterious effect; however, it is well to be aware of its possible presence and effect on short term decisions.

At the other extreme of a large time constant, processing of the input samples is completed before the input process changes its statistical behavior significantly. (For example, the power in the input process may slowly increase and double its value in an hour.) For this case, the estimation device yields a local estimate of the true input p.d.f. If the estimation procedure were repeated on a serial section of input data, gradual changes in the estimated p.d.f. would appear, thereby indicating conclusively a non-stationary trend in the input data.

For intermediate time constant values, significant input statistical changes occur in a time comparable with the processing time. For this situation, a sequential search of p.d.f.space can yield inconsistent results (e.g., the area under the p.d.f. not equal to unity), caused by a "jumping" about of the input process in signal space. The best way to alleviate (but not eliminate) this situation is to record a section of input data which is long enough to accurately evaluate the p.d.f. in one cell of p.d.f. space and rerun the same record for each and every cell of p.d.f. space sequentially. If the time constant is comparable with the processing time for one cell p.d.f. estimation, there will still be significant changes in the total estimated p.d.f., record to record. However, local estimates of the p.d.f. will result which will be consistent (sum up to unity) and will indicate accurately the degree of non-stationarity and its rate of change.

3. A FOURTH ORDER SUCCESS COUNTING PROBABILITY ANALYZER

3.1 GENERAL DESCRIPTION

A breadboard of a success counting probability analyzer has been constructed in accordance with the method* described in Section 2.2.2.2 of this report. A photograph of the unit is shown in Figure 10. The breadboard is basically a flexible, digital 12-th order binary probability analyzer. It is arranged functionally into four octal channels, making it a fourth order analyzer with 3 bit quantization in each dimension.

The serial success counting method whose mathematical description is given by (55), is used for measurement of $\Pr\{\underset{\sim}{s} \in R_{\underset{\sim}{x},\Delta}\}$, with twenty bit trial and success counters. A block diagram of the device is shown in Figure 11. Each of the four channels is provided with the necessary analog and digital circuits required for establishing the success region, $R_{\underset{\sim}{x},\Delta}$, and generating the success waveform $f(\underset{\sim}{s})$ as described by equation (47). An internal clock and a pulse shaper for an external clock are provided for sampling of the success waveform at frequencies up to 20 kc with sample sizes of 2^8 through 2^{19} selectable in 12 steps.

The ratio of the numbers in the success and trial counters is normalized and displayed in a row of lights as a 6 bit binary mantissa and a 4 bit binary (1's complement form) characteristic. The output is also automatically punched as two 6 bit characters on paper tape in a format compatible with CDC 160 computer (the extra two bits are used as a control code).

The size of the success region $R_{\underset{\sim}{x},\Delta}$ for each component of $\underset{\sim}{s}$ may be set (independently for each channel) to 1.25, 2.5, or 5 volts, and the center of the success region may be stepped through 8, 4, or 2 equally spaced intervals in each channel (step size independently selected for each channel). The cell location (center of the region $R_{\underset{\sim}{x},\Delta}$) may be stepped either manually or automatically to a new location at the end of each measurement. The entire 4096 cell space or selected regions may be covered automatically by appropriate settings of front panel controls.

---

*See Figure 6 for a Generic Block Diagram of Method

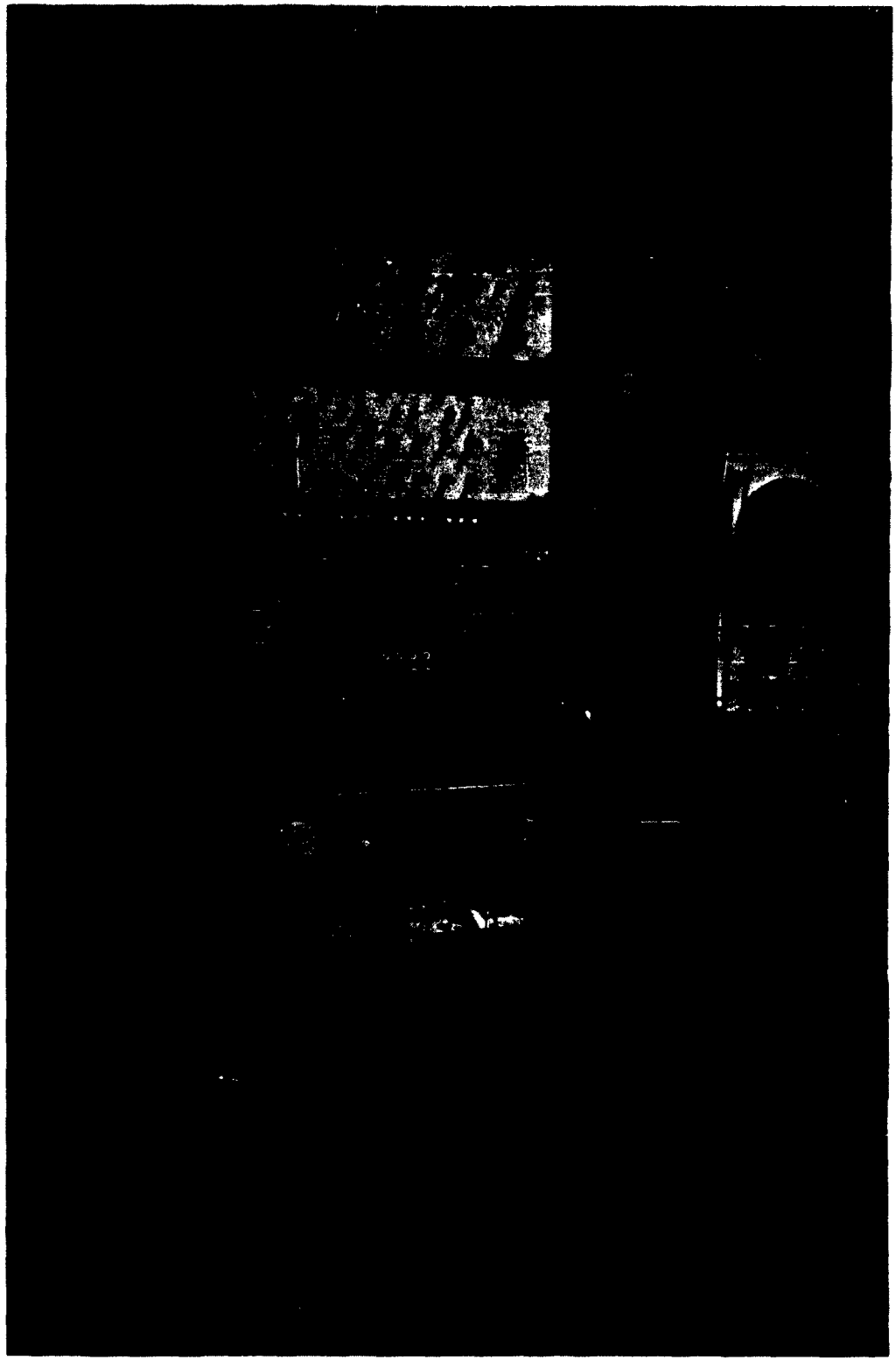


Figure 10. Probability Analyzer

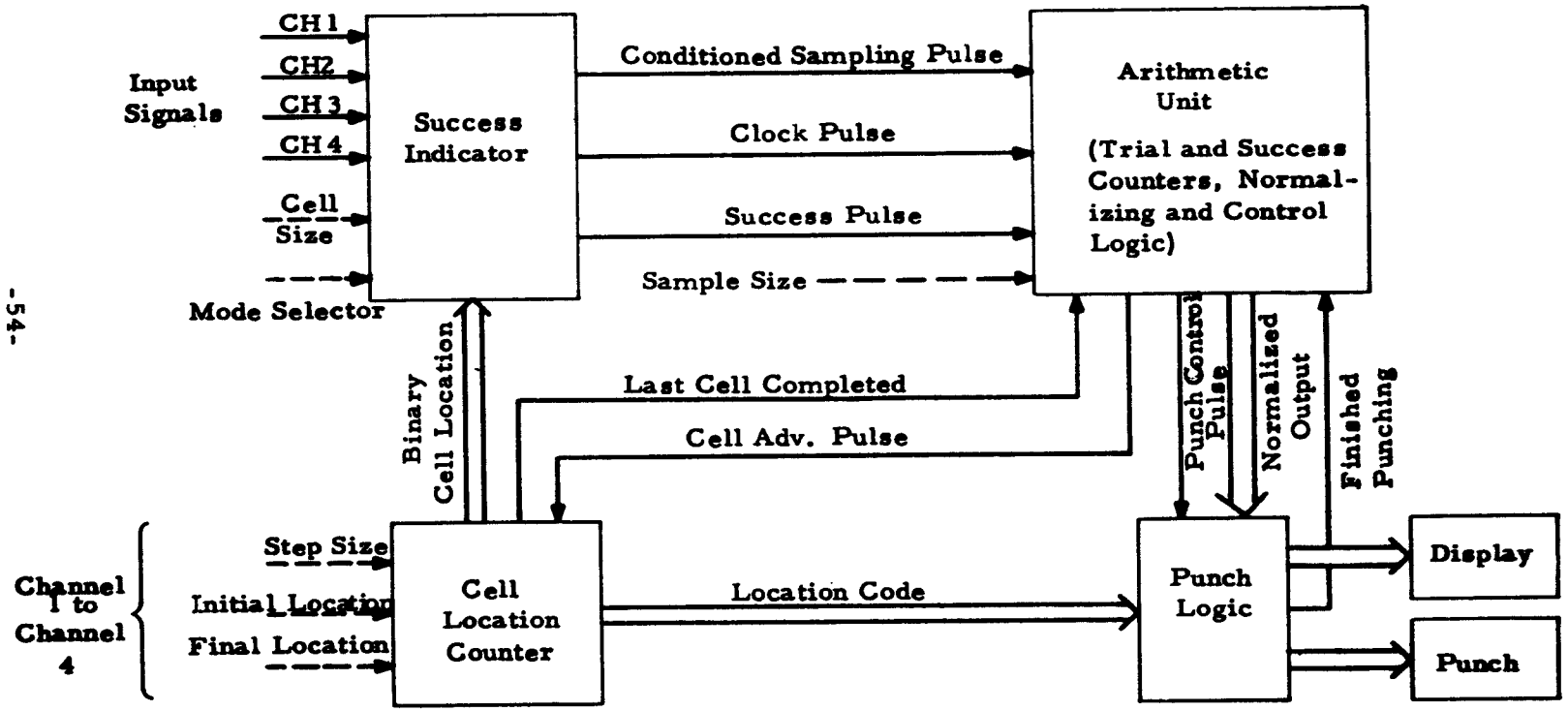


Figure 11. Probability Analyzer Block Diagram

As indicated in Figure 12, two logical AND gates, A and B, are provided with means for switching any combination of the four success waveforms independently to either AND gate. Gate A is sampled by the clock pulses and Gate B may be sampled by either success pulses from the output of Gate A or delayed success pulses from Gate A. The trial counter may be switched to count either clock pulses or Gate A success pulses. The success counter may be connected to the output of either Gate A or Gate B. Thus the analyzer may be set up to measure 1) the probability of a success at Gate A, 2) the probability of a success at Gate B, given a success at Gate A, 3) the probability of a success at Gate A and at a later time a success at Gate B or 4) the probability of a success at Gate B given that a success occurred at some fixed earlier time at Gate A, where success at Gate A or Gate B may independently indicate the joint occurrence of up to four input signals in corresponding independently set regions.

3.2 DETAILED DESCRIPTION OF EQUIPMENT

3.2.1 Basic Units of the Analyzer

The block diagram of Figure 11 shows the major units of the breadboard. The function of the Success Region detector is performed by the Cell Location Counter and the Success Indicator. The Arithmetic Unit includes the Trial and Success Counters plus the control logic required to: a) set the sample size, b) normalize the result, c) initiate the punch cycle and d) automatically advance the cell location counter at the end of each measurement cycle.

3.2.2 Success Indicator

A Simplified Block Diagram of the Success Indicator is shown in Figure 12. A success waveform is generated for each channel by the use of two hyperplane circuits (comparators), a digitally controlled reference voltage, and two offset voltages $+\frac{\Delta}{2}$ and $-\frac{\Delta}{2}$. One comparator has a "1" output when the input signal is less than the reference voltage plus $\Delta/2$ and the other has a "1" output when the input signal is greater than the reference voltage minus $\Delta/2$. If both comparator outputs are "1" then the success waveform f(s) is "1" and s is within the selected cell (or success region) in that channel. The binary input which controls the reference voltage comes from the Cell Location Counter (see Section 3.2.6). Both the cell size, Δ, and the cell center in each channel can be set with an accuracy of

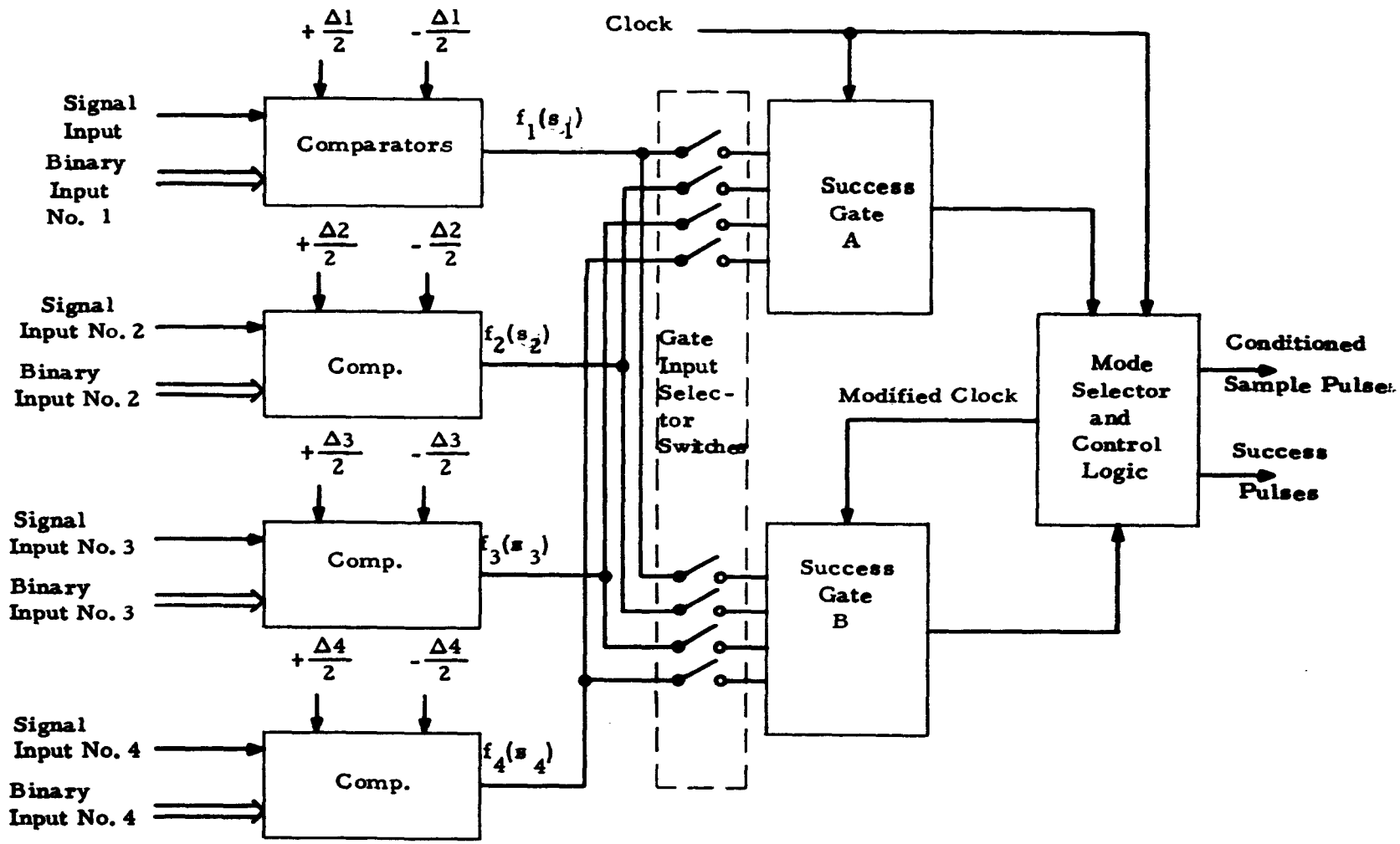


Figure 12. Simplified Diagram of Success Indicator

better than $\pm$ 0.5 percent with a long term stability (several days) of better than 0.5 percent. The effective cell size accuracy is degraded near the upper frequency limit of the analyzer, to a worst case value of 10 percent at 10 kc.

A fourth position of the cell size control sets $\Delta = 0$ and allows for selection of either comparator so that the success waveform indicates success for input signals greater than (or less than) the reference input. This position may be used for measurement of distribution functions.

Two independent 1-st to 4-th order success waveforms may be obtained by switching any combination of the four first order success waveforms to either or both of two logical AND gates: Success Gate A and Success Gate B. The Gate Input Selector Switches may be set independently for each gate. The clock pulses are applied as a 5-th input to Gate A as well as to the Mode Selector and Control Logic. The modified clock pulses, applied to the 5-th input to Gate B, may be either the output pulses directly from Gate A or delayed Gate A output pulses. The Conditioned Sample Pulses are obtained either from the clock directly or from Gate A. The Success Pulses are obtained from either Gate A or Gate B. The mode selector is therefore able to set up the analyzer to measure any one of four quantities:

1) P(A): the joint probability that the signals at the inputs of all channels whose outputs are connected to Gate A will fall within the region selected for their respective channels.

2) P(B/A): the conditional probability of a success at Gate B given a success at Gate A, where success at Gate A or Gate B is defined as in Mode 1.

3) P(B, A delayed): the joint probability of a success at Gate A and at some specified time later a success at Gate B.

4) P(B/A delayed): the conditional probability of a success at Gate B given a success at some specified earlier time at Gate A.

A fifth position of the mode selector is made available for future expandability of the breadboard, namely: P(A or B).

3.2.3 Arithmetic Unit

The Success Indicator outputs (i.e., the conditioned sample pulses and success pulses) are fed to the Arithmetic Unit where they are counted in two 20 bit counters, the Trial Counter and Success Counter respectively. Control circuits are provided to allow sampling to continue until the Trial counter has reached a selected (by the Sample Size Selector Switch) integral power of 2 between 2^8 and 2^{19}, at which time the counting is inhibited. The completion of sampling is detected by the presence of a "1" in the selected (control) bit of the trial counter (and "0's" in all other bits). The Sample Size Selector Switch also selects the Success Counter bit corresponding to the selected Trial Counter Control bit plus the five next less significant bits as the probability output.

On completion of the sampling process, the contents of the success counter are normalized by shifting until the most significant output bit is "1", or for a maximum of 15 shift pulses. The shift pulses are counted in the least significant four bits of the Trial Counter (which are all zero on completion of the sampling interval), which therefore contain the characteristic of the probability. The six selected bits of the success counter contain the Mantissa. On completion of the normalization process, a pulse is transmitted to the punch logic to initiate the punch cycle and to the cell location counter to advance it to the next cell location.

3.2.4 Cell Location Counter

The Cell Location Counter consists of a 12 bit counter arranged as four octal counters, one for each Success Indicator Channel. Each octal unit functions as a separate counter with the overflow from the channel 1 counter indexing the channel 2 counter, overflow from channel 2 indexing channel 3 and so on through channel 4. Overflow from channel 4 stops the automatic recycling of the Arithmetic Unit. Each octal unit has separate front panel control of both initial and final locations, and would be reset to the initial location at the start of a computation, count up to the final location, produce an overflow pulse, reset itself to the initial location, and repeat the process until overflow from the channel 4 counter stops automatic recycling of the arithmetic unit. A two bit location code is generated to indicate whether the last advance pulse produces overflow from the first, second, or third octal unit or no overflow.

The Step Size may be set (independently for each channel) to step through 1/8, 1/4 or 1/2 of the total range on each index (or overflow) pulse. The output of each octal unit is connected to the most significant 3 bits of the corresponding (4-bit) digital-to-analog converter in the success indicator. The least significant bit is controlled by a toggle switch (next to the step size switch), and is used to shift the cell center to the upper edge of one of the eight equal voltage intervals, when the step size is greater than 1/8.

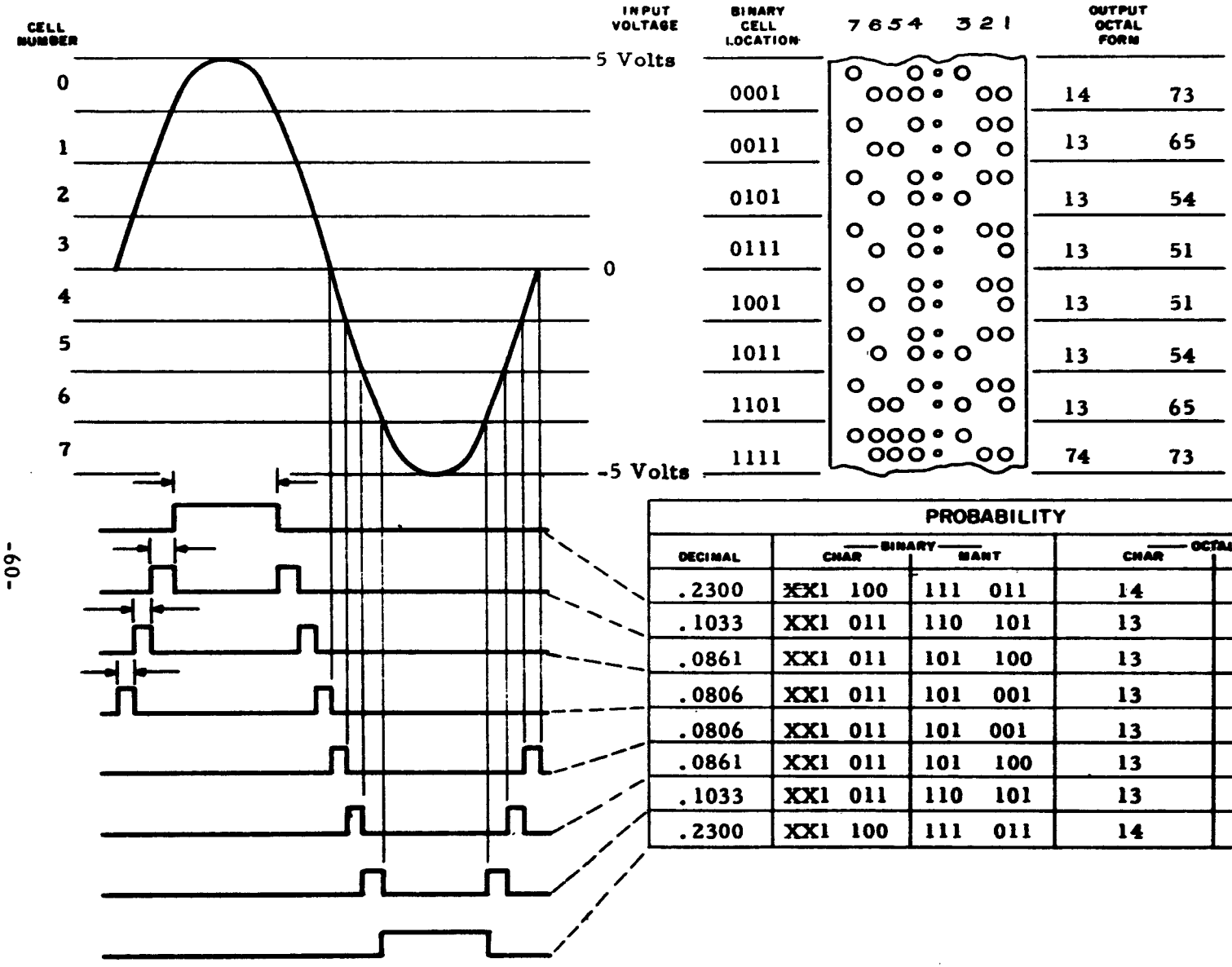
The reason for the flexibility designed into the Cell Location Counter is to allow for automatic stepping through selected limited four dimensional signal space to explore, in more detail, regions of particular interest.

3.2.5 Punch Logic and Data Format

The punch logic serves to commutate the data into the paper tape perforator and generate the necessary timing pulses to punch two 7 bit characters for each measurement and a pulse to recycle the Arithmetic Unit.

The output data is arranged as a 12 bit computer word compatible with the CDC 160 computer. The least significant six bits are the Mantissa and are obtained from the Sample Size Switch output. The next four bits are the one's complement of the characteristic. The complement form is used since the characteristic is always negative. By making the characteristic the more significant part of the word (than the Mantissa) and using its complement form, a monotonic relationship between the probability and its binary representation is maintained, thus simplifying certain computer processing of a probability density function. The most significant two bits are the location code generated in the Cell Location Counter and will be useful in checking the data while entering it in the computer.

The computer word is punched as two characters, the most significant half being punched first and indicated by a "1" in the seventh hole position. A typical tape for a Sine Wave, first order density function is shown in the upper right of Figure 13. The numbers punched are indicated in a convenient octal form as shown to the right of the tape. Four octal characters are used



PROBABILITY				
DECIMAL	BINARY CHAR	BINARY MANT	OCTAL CHAR	OCTAL MANT
.2300	XX1 100	111 011	14	73
.1033	XX1 011	110 101	13	65
.0861	XX1 011	101 100	13	54
.0806	XX1 011	101 001	13	51
.0806	XX1 011	101 001	13	51
.0861	XX1 011	101 100	13	54
.1033	XX1 011	110 101	13	65
.2300	XX1 100	111 011	14	73

Figure 13. Graphical Calculation of Sine Wave Density Function

for two punched characters or one computer word. This notation is convenient since the data can be easily typed out in this form using standard programs. Only the least significant bit of the most significant octal digit is part of the characteristic and so this digit should be considered as 1 for odd numbers (1, 3, 5, or 7) and 0 for even numbers (0, 2, 4, 6). Since the most significant 2 bits of this octal digit represent the location code, this code can be obtained by subtracting 1 from odd numbers and 0 from even numbers and dividing by 2.

3.2.6 Use of Paper Tape Output

For convenience in processing the output of the probability analyzer in a digital computer, a paper tape punch has been provided with the breadboard. The format of punched data has been chosen for convenience in using the CDC 160 computer, as discussed in Section 3.2.5.

Even though the Probability Analyzer is a versatile device which can provide probability estimates directly at its output, problems may arise for which further processing is required. Two types of analyses which are especially useful for evaluating the characteristics of n-order statistics (after the n-dimensional probability function has been calculated) are the examination of two-dimensional cross sections, and identification of the modes (high density regions) of a probability density function. A third function which might be performed with a computer is computation of $\Pr\{\underset{\sim}{s} \in R\}$, where R is a complicated region not easily instrumented in the Success Region Detector, and where the probability density function, $\hat{p}_n(\underset{\sim}{x})$ has been calculated in the Probability Analyzer.

To illustrate the manner in which operations such as these might be performed using the paper tape output of the Probability Analyzer, flow charts of programs for (a) inserting data into the computer, and (b) computing and displaying two dimensional cross sections of joint probability density functions, $\hat{p}_n(\underset{\sim}{x})$ (all but two of the $\{x_i\}$ fixed), and conditional probability densities, $\hat{p}_n(x_j \mid \text{all } x_i,\ i \neq j)$, have been included in Appendix I.

3.3 CALIBRATION DATA

3.3.1 D.C. Calibration Data

A simplified block diagram of the comparator circuits, which determine the cell size and cell location, is shown in Figure 14. The data for the final DC calibration of these circuits (taken about 1 week prior to delivery) is shown in Tables 2, 3, and 4. The cell location calibration (Table 2) was made by measuring the input signal (from a precision voltage reference source) required to set the output of the summing amplifier to zero volts, for each binary input. Data for the two extremes and the center point are presented. The worst case error for all points of all four channels was less than 0.1 percent of full scale. The data in Table 2 were rechecked after delivery to RADC and all points were within 10 mv without further re-adjustment.

The cell size data of Table 3 was obtained by adding a small (50 mv p-p) low frequency sinewave to the DC input, setting the Binary input to 1000 (see Table 2) and measuring the positive and negative input voltages required to produce a square success waveform, f(s). The cell size indicated in Table 3 are all within 1 percent of their nominal value. The values for the 1/8 cell size were rechecked after delivery to RADC and were all still within the 1 percent tolerance and had not changed by more than 5 mv from the initial calibration points.

The Schmitt Trigger hysteresis (Table 4) was also rechecked at RADC and all values were again within 5 mv of the initial calibration.

3.3.2 Low Frequency Sine Wave Calibration

A graphical method of calculating the histogram of the first order probability density function of a sinewave using 3 bit (8 level) quantization is shown in Figure 13. The cell numbers, binary representation of the cell centers, and voltages of the cell boundaries are shown with the sinewave in the upper left corner. The corresponding success waveforms f(s) are shown (for each cell) below the sinewave. The probability for each of the eight cells is shown in the table to the right of the success waveform in decimal, binary, and octal form. The binary and octal forms are in the normalized form available as the analyzer output, with the complement of the

Binary Input From Location Counter	Input Voltage required to produce zero output from Summing Amplifier				
	Channel 1	Channel 2	Channel 3	Channel 4	Nominal
0000	4.998	4.997	5.000	4.997	5.000
1000	0.000	0.004	0.002	0.000	0.000
1110	-3.748	-3.748	-3.748	-3.748	-3.750

Table 2. Calibration of Cell Location

Cell Size	Boundary	Channel 1	Channel 2	Channel 3	Channel 4	Nominal
1/8	Upper	+ .618	+ .628	+ .633	+ .630	+ .625
	Lower	- .630	- .616	- .622	- .623	- .625
1/4	Upper	+1.247	+1.248	+1.250	+1.239	+1.250
	Lower	-1.240	-1.249	-1.248	-1.252	-1.250
1/2	Upper	-2.457	-2.470	-2.490	-2.475	-2.500
	Lower	-2.457	-2.470	-2.490	-2.475	-2.500

Table 3. D. C. Calibration of Cell Size

St. No.		Trigger Level, mv				
		Channel 1	Channel 2	Channel 3	Channel 4	Nominal
1	Positive	12	13	15	18	15
	Negative	18	13	15	12	15
2	Positive	15	20	15	15	15
	Negative	20	15	10	12	15

Table 4. Low Frequency Calibration of Schmitt Trigger Hysteresis

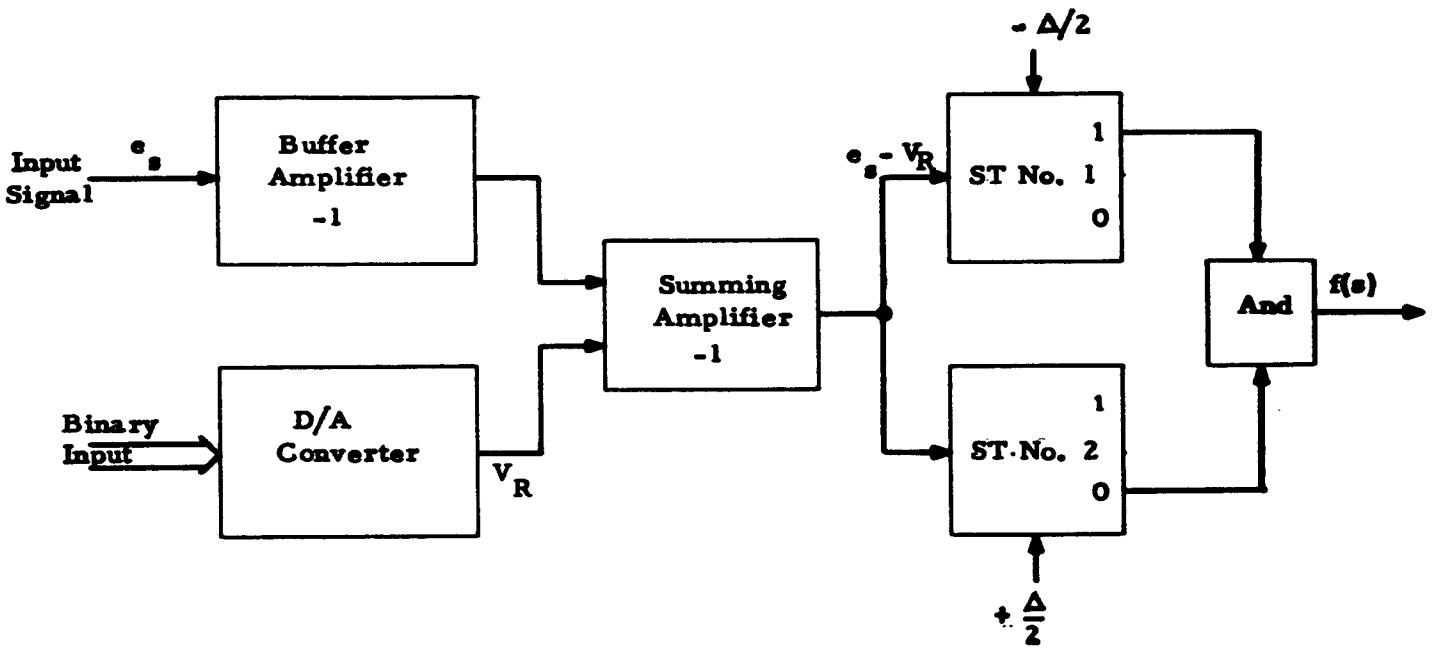


Figure 14. Simplified Diagram of Comparator Circuit

characteristic indicated. The x's in the binary form are location code bits and may take on any value, depending on the channel being used, and the cell being measured as well as the setting of certain switches. A paper tape representation of a typical sinewave density function is shown above the table.

For calibration of the analog circuits the density function of a low frequency sinewave was determined by measuring the duty cycle of the success waveform, f(s). Table 5 lists these results for all four channels with a 100 cps sinewave. These numbers agree with the calculated values within the accuracy (a few percent) with which the scope could be interpreted. A copy of a paper tape output from the Probability Analyzer automatic measurement of the data in Table 5 is shown in Figure 15. The Cell numbers and octal representation of the punched data are shown next to the tape. Note that the most significant octal characters are different for each channel. This difference is the result of the location code bits, which are different for each of the measurements. The results on this tape are all within 1 bit (about 2 percent) of the calculated value (see sample tape in Figure 13.

3.3.3 High Frequency Calibration Using a 10 kc Sinewave

Table 6 gives the results of the analog measurement and Table 7 the corresponding digital measurement of the density function of a 10 kc sinewave. These results were obtained at RADC after delivery of the equipment. The worst case error was about 10 percent, however the digital and analog measurements agree within a few percent. Slightly better results were obtained previously at Waltham. Some of the errors in the measurements given in Tables 6 and 7 may be due to the 50 mv noise level (about 4 percent of the cell size) of the sinewave generator. As a result of the noise level, the sinewave had to be set slightly less than 10 v pp in order to contain the whole density function within the $\pm$ 5v range of the instrument. Because of time limitations and the relative significance of the errors the exact cause was not tracked down.

Cell Number	Channel 1	Channel 2	Channel 3	Channel 4	Calculated (Fig. 15)
0	.235	.235	.235	.235	.230
1	.104	.104	.104	.104	.1033
2	.085	.085	.085	.085	.0861
3	.080	.080	.080	.080	.0806
4	.080	.080	.080	.080	.0806
5	.086	.086	.086	.086	.0861
6	.104	.102	.103	.103	.1033
7	.230	.220	.230	.225	.230

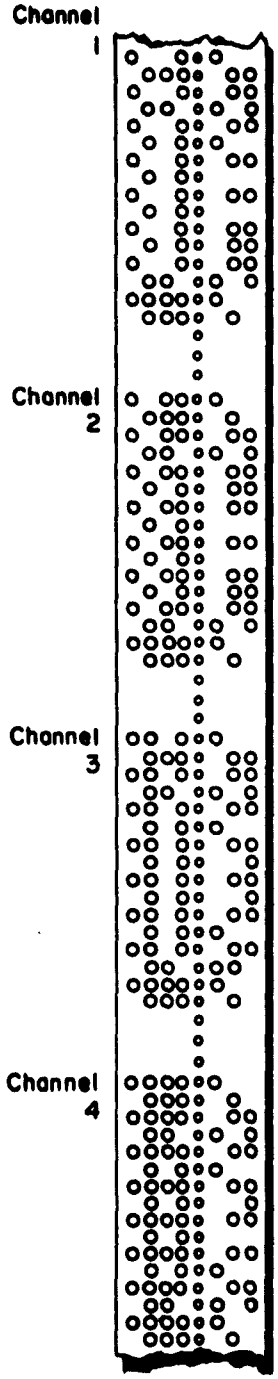
Table 5. Low Frequency Calibration of the Four Channels by Measuring the Duty Cycle of the Success Waveform f(s) for a 100 cps Sine Wave

Cell Number	Channel 1	Channel 2	Channel 3	Channel 4	Calculated (Fig. 15)
0	.220	.210	.215	.220	.230
1	.103	.103	.111	.110	.1033
2	.086	.086	.089	.089	.0861
3	.082	.080	.083	.082	.0806
4	.080	.080	.085	.080	.0806
5	.087	.087	.090	.086	.0861
6	.106	.105	.110	.104	.1033
7	.230	.235	.228	.235	.230

Table 6. High Frequency Calibration of the Four Channels by Measuring the Duty Cycle of the Success Waveform for a 10 Kc Sine Wave

Cell Number	Channel 1		Channel 2		Channel 3		Channel 4		For 10V p-p Sinewave Calculated	
0	14	67	14	65	14	67	14	70	14	73
1	13	64	13	66	13	66	13	65	13	65
2	13	53	13	52	13	54	13	54	13	54
3	13	51	13	50	13	53	13	51	13	51
4	13	50	13	47	13	52	13	50	13	51
5	13	53	13	52	13	55	13	53	13	54
6	13	65	13	65	13	70	13	65	13	65
7	14	72	14	73	14	71	14	73	14	73

Table 7. Digital Calibration of the Four Channels Using a 10 Kc Sine Wave



	CELL NUMBER	OCTAL OUTPUT	
Channel 1	0	14	73
	1	13	65
	2	13	54
	3	13	50
	4	13	50
	5	13	53
	6	13	65
	7	74	72
Channel 2	0	34	72
	1	33	65
	2	33	53
	3	33	50
	4	33	50
	5	33	53
	6	33	65
	7	74	72
Channel 3	0	54	73
	1	53	65
	2	53	54
	3	53	51
	4	53	51
	5	53	54
	6	53	66
	7	54	72
Channel 4	0	74	72
	1	73	65
	2	73	54
	3	73	51
	4	73	50
	5	73	54
	6	73	65
	7	74	72

Figure 15. Low Frequency Sinewave Calibration of the Four Channels

3.3.4 Noise Measurement

The density function of a noise generator output filtered by a single pole 5 kc lowpass filter was measured for each channel with the results shown in Table 8. The density function for the "Typical Channel" was obtained by taking the mean or in a few cases the mode of the four channels. The decimal equivalent of the histogram for the typical channel is given as well as a calculated histogram. The basis for the calculations was a Gaussian distribution with zero mean and a value of σ chosen for best fitting of the typical channel curve ($6.64\sigma = 10$ v). The fit is good near the center cells and gets progressively worse at the outside cells. The $\pm$ 5 percent asymmetry of the curve suggests the possibility of non-symmetrical limiting which could occur in the analyzer due to differences in the recovery time from positive and negative overloads. This source of distortion could be investigated (and removed) by limiting the input signal (using fast recovery diode limiters) to the $\pm$ 5 volt range of analyzer. The typical and calculated histogram from Table 8 are plotted in Figure 16 for comparison.

As a check on the higher order capabilities of the analyzer the joint probability of a success at Gate A and at about 1 ms later a success at Gate B was measured where channel 2 was connected to Gate A and channel 3 connected to Gate B and the previous noise signal connected to both channels 2 and 3. With binary inputs to both channels set to cell 3, the measured probability was, in octal form: 1355 which is exactly the product of the probabilities given for the corresponding cell and channels in Table 8.

3.4 EXPANSION CAPABILITIES OF THE BREADBOARD

3.4.1 Modifications of the Analog Circuits

While the breadboard was designed to be a fairly versatile general purpose probability analyzer, some limitations had to be imposed by cost considerations. The limitations, however, are primarily in the analog circuitry. Numerous simple modifications to these circuits could be made to extend their usefulness in the solution of specific problems. Incorporation of these modifications did not seem practical since their particular configuration would be a function of the particular problem, and they might better be made as temporary wiring changes than as unnecessarily complicated panel controls.

Cell Number	Channel 1		Channel 2		Channel 3		Channel 4		Typical Channel		Typical Channel	Calc. for 6.64σ=10V
0	07	44	07	40	07	51	07	42	07	44	.0044	.0060
1	12	52	12	46	12	50	12	51	12	50	.039	.042
2	14	51	14	47	14	50	14	47	14	50	.156	.155
3	15	50	15	46	15	46	15	46	15	46	.297	.297
4	15	47	15	45	15	45	15	45	15	45	.289	.297
5	14	47	14	46	14	47	14	46	14	46.5	.150	.155
6	12	53	12	53	12	53	12	53	12	53	.0419	.042
7	07	63	07	64	07	66	07	64	07	64	.0054	.0060

Table 8. Calibration of the Four Channels Using 5 Kc Band Limited Noise

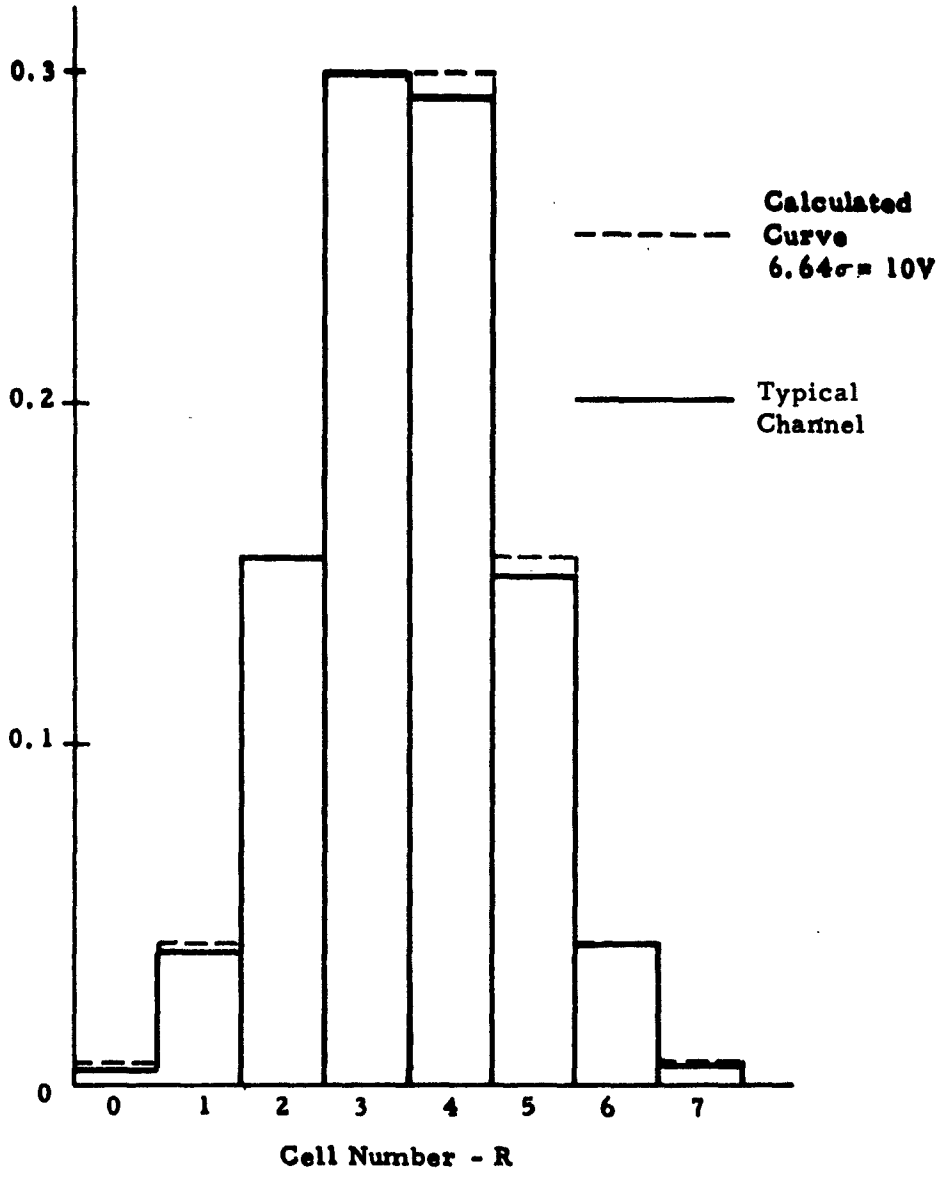


Figure 16. Histogram for Gaussian Noise Waveform

One such modification would be provision for the generation of irregularly shaped success regions. Since there are presently eight hyperplane circuits in the equipment with convenient means for connecting these circuit outputs to the success gates, only minor rewiring would be necessary to provide for convenient insertion of additional summing resistors.

A second modification would be to provide additional limiting circuits so that the signals could be amplified further and examined in more detail in narrower amplitude ranges of interest.

3.4.2 Use of a Digital Comparator

As mentioned in the general description, the breadboard is a 12-th order binary or 4-th order octal probability analyzer. A considerable increase in its flexibility could be obtained by providing for the use of digital comparators as well as the present analog comparators. This modification could be accomplished very easily with the addition of a small amount of additional wiring in such a way that either analog or digital comparators could be plugged into the same connector, and the switching would be automatic.

With this modification, any binary input could be applied to the analyzer and compared with the cell location counter setting. The success waveform would thus be "1" when both binary numbers are the same and zero otherwise.

For example a 12 bit D-A converter could be connected between the signal and the analyzer, and a high resolution first order density function measured automatically with each conversion representing one sample. With two 6 bit D-A converters a higher resolution second order density function could be measured.

If a set of signals were A-D converted (using any combination of A-D converters with up to 12 bits output) and the resulting set of numbers recorded on a magnetic tape loop, then the tape output could be compared

with the Cell Location Counter and a corresponding density function calculated. This option would have the advantage of eliminating the variation of signal characteristics with time which might be inherent in an analog recording.

For pattern recognition problems it is often desirable to analyze some analog waveforms along with the outputs of parameter extractors which may have digital outputs. For such problems it would be convenient to use one or more of the analog inputs to the analyzer and at the same time use digital comparators in the other inputs. With modifications cited above, such problems can be solved readily.

4. CONCLUSIONS AND RECOMMENDATIONS

For estimating n-th order probabilities of the form, $\Pr\{\underset{\sim}{x} \in R\}$, where $n > 1$, the success counting method of estimation is the most versatile and accurate of the various methods examined. If extremely low probabilities are to be calculated (as when investigating "tails" of distribution functions), the sampling and counting technique utilising digital circuitry is preferable to the analog technique of integrating under a success waveform.

The choice between parallel processing of samples utilising a general purpose computer and serial processing with a special purpose device hinges on (1) the availability of a computer, (2) the capability to feed data directly into the computer without intermediate storage, and (3) whether many small-scale probability analyses will be performed, or relatively few large-scale analyses. Consideration of (2) and (3) has led to implementation of the serial processing method on this project.

Salient features of the serial processing breadboard are:

(1) A rough estimate of an entire probability density function can be obtained quickly through the use of a coarse histogram cell structure.

(2) Partial results (e.g., a portion of a probability density function) can be obtained quickly without having to go through a complete calculation.

(3) Answers to practical problems of the form $\hat{\Pr}\{\underset{\sim}{x} \in R\}$, where R may be a complicated region, can be obtained directly without going through the intermediate step of calculating $\hat{p}_n(\underset{\sim}{x})$.

(4) Accuracy requirements can be translated to processing time with a fairly high degree of confidence.

(5) Expansion or modification of the device to accommodate special requirements is readily implementable.

A device of the sort described in Section 3 will provide answers to a large variety of statistical questions, if suitable modifications are made in the success region detector. Since these modifications will in most cases be extremely inexpensive, but cannot be foreseen for all problems, it is suggested that this breadboard be regarded and utilized as a general purpose digital probability analyzer, taking full advantage of the provision for changing the success region detector.

LIST OF REFERENCES

[1] Davenport, W. B., Jr., Johnson, R. A., and Middleton, D., "Statistical Error in Measurements on Random Time Functions", J. Applied Physics, Vol. 23, pp. 377-388; April 1952.

[2] Feller, W., An Introduction to Probability Theory and its Applications, John Wiley and Sons, Vol. I, 2nd Ed., 1957.

[3] Jordan, K. L., "A Digital Probability Density Analyzer", S. M. Thesis, MIT; August 1956.

[4] Nuttall, A. H., "Optimum Linear Filters for Finite Time Integrations", Quarterly Progress Report, RLE, MIT, pp. 65-75, October 15, 1957.

[5] Schreiber, W. F., "The Measurement of Third Order Probability Distributions of Television Signals, IRE Trans., PGIT, Vol. IT-2, No. 3, pp. 94-105; September 1956.

[6] Shannon, C. E., "Communication In The Presence of Noise", Proc. IRE, Vol. 37, No. 1, pp. 10-21; January 1949.

[7] Van Meter, D., and Middleton, D., "Modern Statistical Approaches to Reception in Communication Theory", Trans. IRE, PGIT-4, pp. 119-145; September 1954.

[8] White, H. E., "An Analog Probability Analyzer", Technical Report No. 326, RLE, MIT; April 1957.

[9] Wiener, N., Extrapolation, Interpolation, and Smoothing of Stationary Time Series, John Wiley and Sons; 1949.

[10] Wolf, A. A., and Dietz, J. H., "A Method of Measurement and Display of Probability Functions of Ergodic Random Processes by Orthogonal Series Synthesis", Proc. IRE, Vol. 50, No. 12, Correspondence, pp. 2503-2504; December 1962.

[11] Easter, B., "An Electronic Amplitude Probability Distribution Analyzer", S. M. Thesis, MIT; 1953.

APPENDIX I. COMPUTER PROGRAM FLOW CHARTS

I. 1 The following is a description of a method (Figure I-1) for storing the data output from the Probability Analyzer into a computer in the same arrangement that it held in the Probability Analyzer.

Each word in the computer is initialized to a code to indicate, at the program's completion, the locations which do not contain pertinent information. Control data consists of two four-digit octal location readings (the initial and final locations for the data) and a four-digit switch setting. Each digit in the latter is used to step the corresponding digit position in the read location during program execution. An overflow condition is generated when any digit position in the current read location exceeds its corresponding digit in the final location. The type overflow to be generated can be computed and compared to the overflow indicator on the punched tape input to verify the data transmission.

Explanation of tape input (i. e., control data)

Start Location $D_4 D_3 D_2 D_1$

Final Location $S_4 S_3 S_2 S_1$

Switch Settings $SW_4 SW_3 SW_2 SW_1$ [Possible settings, 1, 2, 4]

Overflow Codes 01 into D_2
10 into D_3
11 into D_4

The D's are the octal digits of the initial read location; S's are the final settings; SW's are the increments for stepping through each of the digit positions.

Ex. Start 2023
Stop 4065
Switches 2141

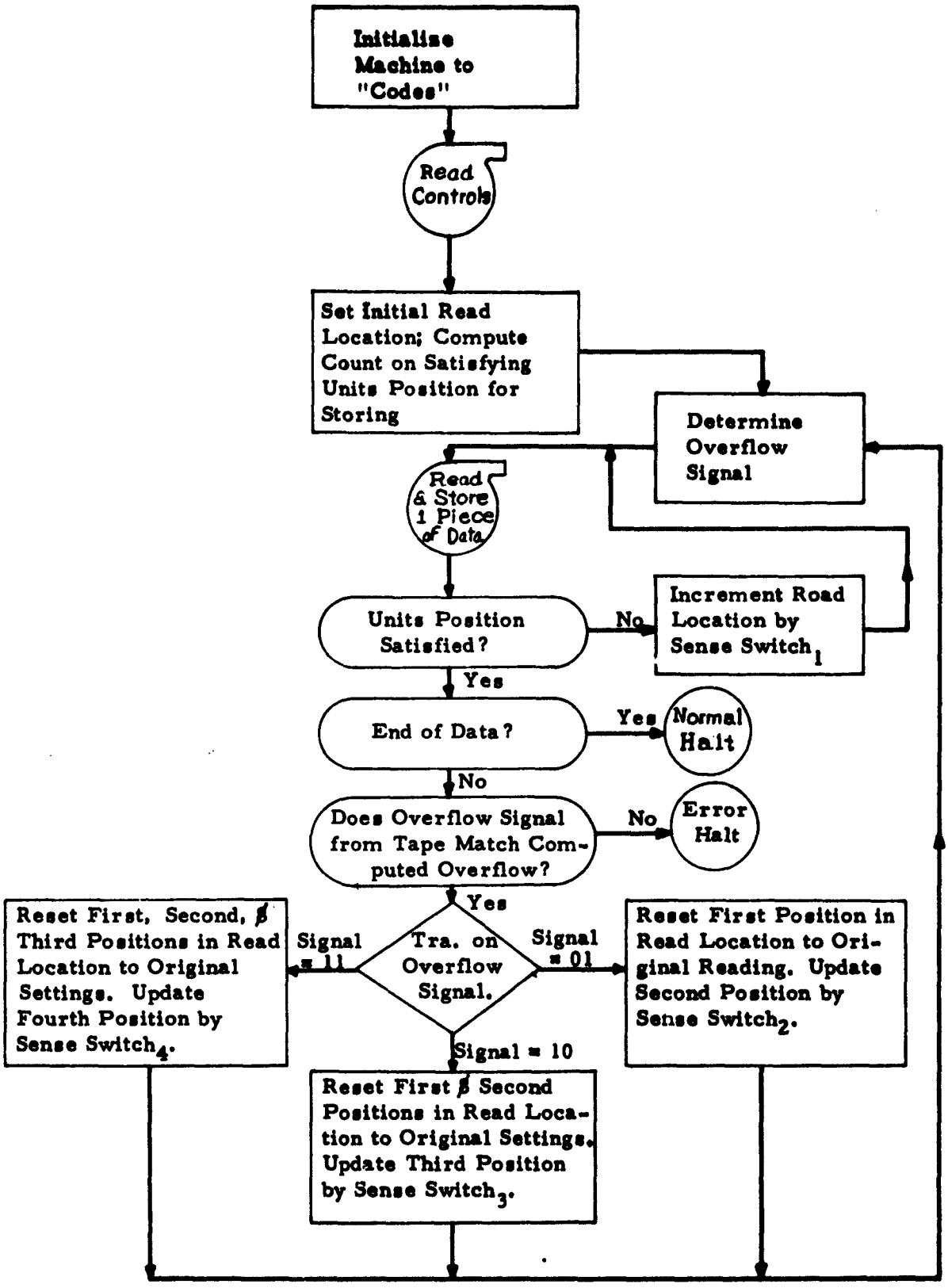
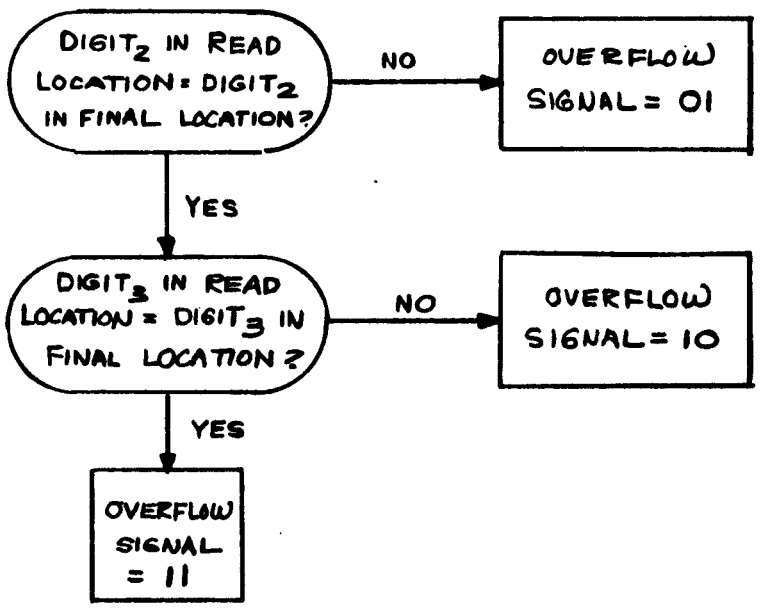


Figure I-1. Data Input Program

Read Data into	**2023**	
	2024	
	2025	**01 overflow**
	2063	
	2064	
	2065	**11 overflow**
	4023	
	4024	
	4025	**01 overflow**
	4063	
	4064	
	4065	
	End of Data	

"Codes" = any signal to indicate locations which do not contain pertinent data after the execution of the program.

A method for determining what type overflow signal will be found:



$$\text{Count for Digit}_1 = \frac{(\text{Digit}_1 \text{ in Original Loc.} - \text{Digit}_1 \text{ in Final})}{\text{Switch Setting}_1} + 1$$

I.2 Figure I-2. is the flow chart of a technique which can be used in calculating two-dimensional cross sections of n-th order probability density functions, $P_n(x_1, \ldots, x_n)$ for $1 \leq n \leq 4$. The equations below apply to said flow chart.

$$f(x_{JI}) = P_N(a_1, \ldots, a_{j-1}, x_{JI}, a_{J+1}, \ldots, a_N)$$

where the $\underset{\sim}{a}^{(J)} = (a_1, \ldots, a_{J-1}, a_{J+1}, \ldots, a_N)$

are specified quantities.

$$g(x_{JI}) = P_N(x_{JI} \mid \underset{\sim}{a}^{(J)}) = \frac{f(x_{JI})}{\sum_{s=1}^{c} f(x_{JS})}$$

where $I = 1, 2, \ldots, c$ and $1 \leq c \leq 8$.

I.3 In order to plot an array of two dimensional data on a typewriter, the data should be ordered on the independent variable. The range of the dependent variable is scaled across a given number of type positions. If it is desired to scale the independent variable, the range is based on the number of lines per page. Each parameter is scaled over the required range, the number of spaces or carriage returns is computed and a decimal point typed in the appropriate position.

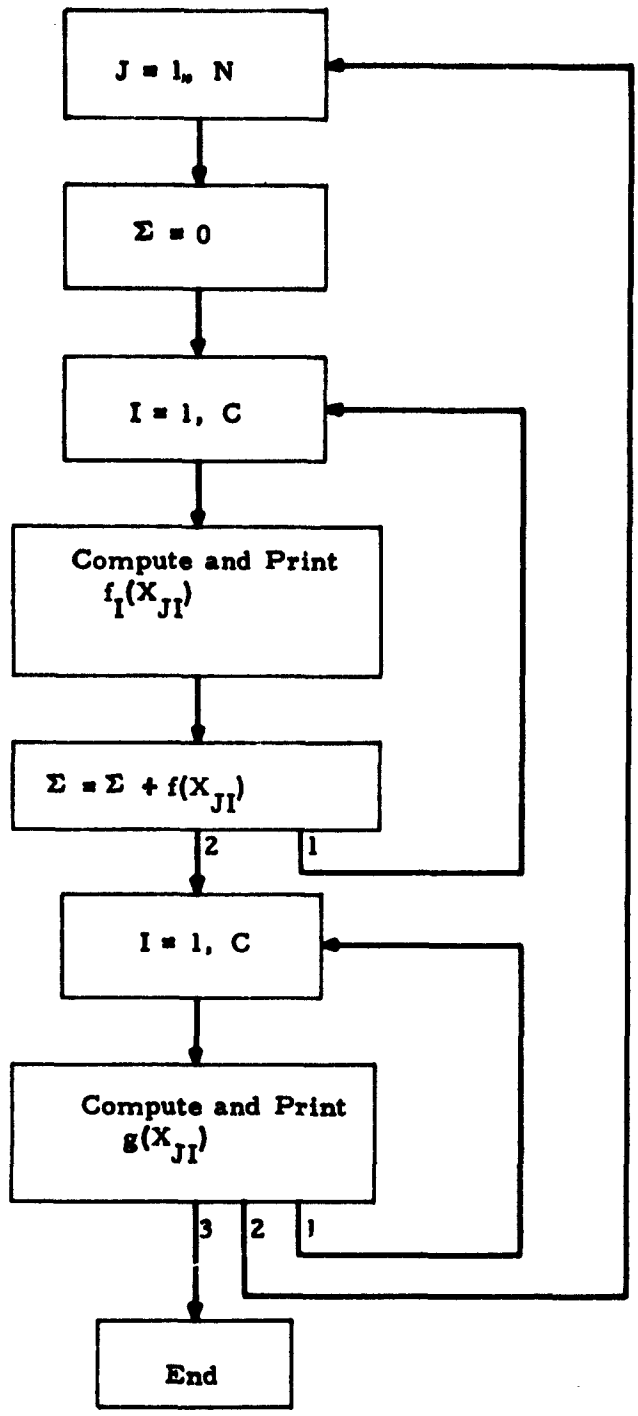


Figure I-2. Calculation of 2-Dimensional Cross Sections of n-th Order Probability Density Functions

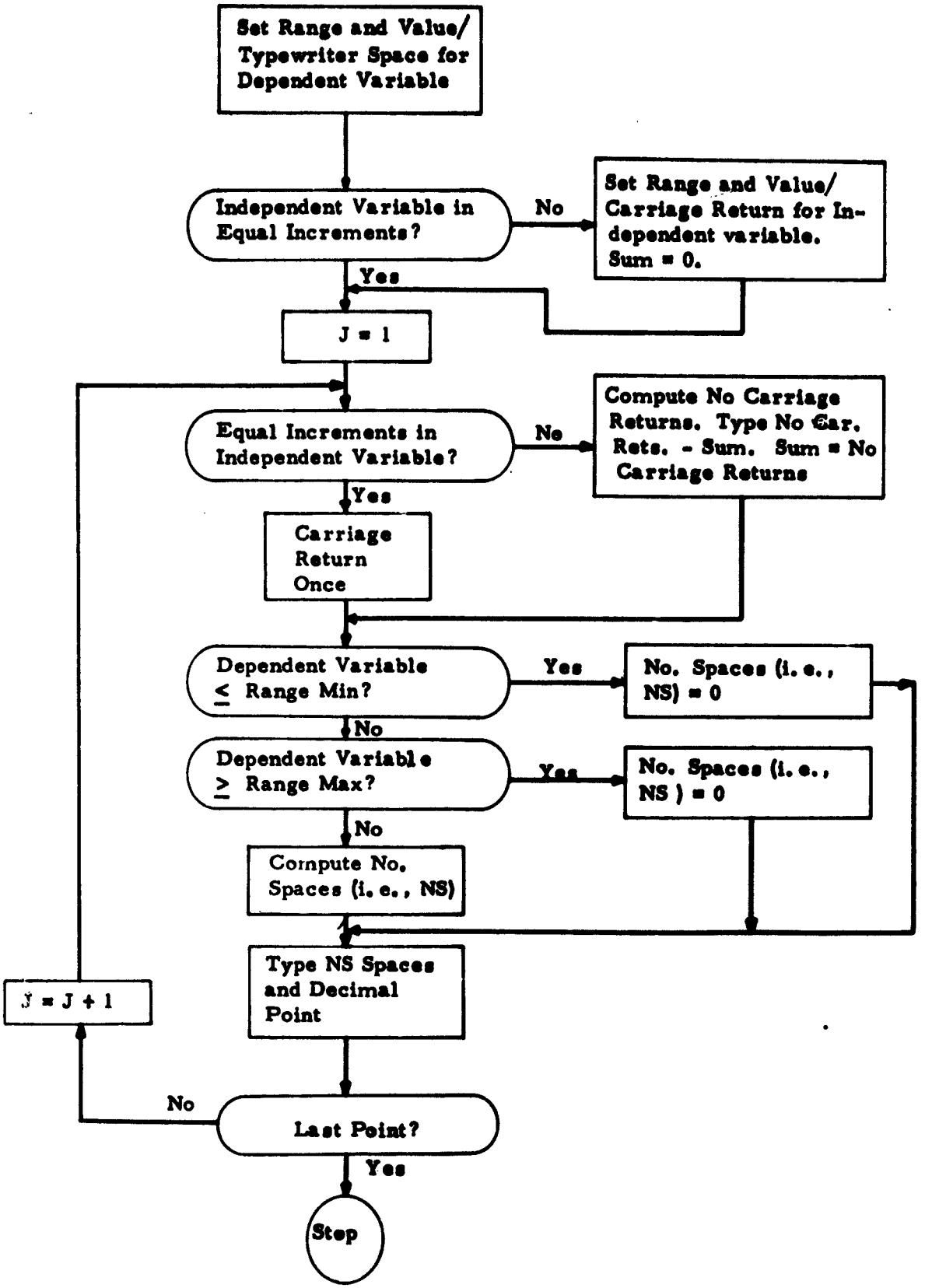


Figure I-3. To Plot An Array of Two Dimensional Data
Note: Independent Variable must be Ordered