April 24, 1974

To the Graduate Council:

I am submitting herewith a thesis written by Ralph E. Klautsch entitled "A Statistical Procedure for Improving the Precision of Data Acquired on a Digital Data Acquisition System." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.

E. C. Huebschmann, Major Professor

We have read this thesis
and recommend its acceptance:

Fletcher W. Donaldson

Roy D. Joseph

Accepted for the Council:

Hilton A. Smith
Vice Chancellor
Graduate Studies and Research

A STATISTICAL PROCEDURE FOR IMPROVING THE PRECISION OF DATA ACQUIRED ON A DIGITAL DATA ACQUISITION SYSTEM

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee

Ralph E. Klautsch

June 1974

ACKNOWLEDGMENTS

The author wishes to express appreciation to Dr. E. C. Huebschmann for his advice in this study, to the management of ARO, Inc., for permission to attend The University of Tennessee Space Institute, and to his wife, Faye, for her continued patience and encouragement.

The work reported herein was supported by Headquarters Arnold Engineering Development Center, Arnold Air Force Station, Tennessee within the terms of Contract F40600-C-74-0001 with ARO, Inc. Reproduction to satisfy the requirements of the United States Government is authorized.

1147094

ABSTRACT

The precision desired in a measuring process sometimes exceeds that which available hardware is capable of producing with a single reading of the parameter of interest. A method is presented for improving the precision of data acquired on a data acquisition system by applying statistical averaging to replicate measurements. Procedures for determining confidence intervals associated with typical data-taking situations are detailed. Computer programs, developed for modeling various data acquisition processes, are used to test the reliability of the statistical theory. Random number generation routines are used to simulate data noise variations. Experimental data acquired on an operational data acquisition system are presented and compared with predicted results. Data precision improvements of an order of magnitude over single reading uncertainty are demonstrated. Primary emphasis is placed on static and quasi-static data; however, some attention was given to low frequency non-static data. It is concluded that the replicate measurement technique can be used to significant advantage in typical data acquisition situations involving slowly varying data, and further efforts toward determining the applicability of similar statistical processes to higher frequency data are suggested as follow-on work.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
II.	ANALYSIS	5
	Measurement Error	5
	Sampling Statistics	7
	Confidence Intervals	10
	Computer Modeling	26
III.	EXPERIMENTAL EXERCISE	32
	Experimental Procedure	37
	Data Analysis	38
	Elimination of Outliers	44
	Other Program Features	47
IV.	NON-STATIC DATA APPLICATIONS	50
V.	CONCLUSIONS	54
BIBLIOGRAPHY		55
APPENDIXES		58
A.	VARIANCE OF A UNIFORM DISTRIBUTION OF ERROR	59
B.	COMPUTER PROGRAM FOR CALCULATION OF STANDARD DEVIATION	62
C.	COMPUTER MODEL OF DATA ACQUISITION PROCESS	66
D.	COMPUTER MODEL FOR DETERMINING STANDARD DEVIATION OF A UNIFORMLY DISTRIBUTED POPULATION	73
E.	COMPUTER PROGRAM FOR DATA ANALYSIS	77

CHAPTER PAGE

F. COMPUTER MODELS FOR NON-STATIC DATA 87

VITA 91

LIST OF TABLES

TABLE		PAGE
I.	Two Tailed Student's "t" Tables	12
II.	Mean Relative Range	15
III.	Sample Averages within Confidence Interval	29
IV.	Thompson's Tau	46
V.	Effects of Removing Outliers	48

LIST OF FIGURES

FIGURE		PAGE
1.	Distribution of Errors of Unbiased Measurement Processes	8
2.	Confidence Intervals for Sample Averages	17
3.	Uniform Distribution of X	20
4.	Curves for Estimating Standard Deviation	23
5.	Ninety-Five Percent Confidence Interval Comparison	25
6.	Distribution of Sample Averages (Computer Model)	27
7.	Standard Deviation of a Uniform Distribution (Computer Model)	31
8.	VKF Data Processing System	33
9.	Ninety-Five Percent Confidence Interval, C_u-C_n Thermocouple Measurement	36
10.	Experimental Results, Subcommutator Data	40
11.	Experimental Results, Main Commutator Data	42
12.	Computer Simulation, Acquisition of Time Varying Linear Data	52
13.	Computer Simulation, Acquisition of Sinusoid Data	53
14.	Uniform Distribution of Error	60
15.	Calculation of Standard Deviation	63
16.	Modeling of Data Acquisition Process	67
17.	Model for Determining Standard Deviation	74

NOMENCLATURE

a	Constant
b	Constant
B	Magnitude of sinusoid
CI	Confidence interval
C_u-C_n	Copper constantan
d_2	Mean relative range
°F	Degrees Fahrenheit
Hz	Hertz
m	Slope of straight line
N	Sample size
p(x)	Probability density
P	Level of significance (Thompson's τ)
R	Range
$\bar{R}$	Average range
S	Estimate of standard deviation
S^2	Estimate of variance
t	Distribution function (Student's t); time
X	Measured value
X_i	Specific measured value
$\bar{X}$	Average value of X
Δx	Difference between X values
VKF	Von Karman Facility
y	Dependent variable
Z	Standard normal deviate

γ	Confidence coefficient
δ_i	Difference between parameter values
ε	Error of measurement
μ	Mean of repeated measurements
σ	Standard deviation
σ^2	Variance
τ	True value of a parameter
ω	Angular frequency

CHAPTER I

INTRODUCTION

A problem familiar to those responsible for making measurements is the requirement for producing results that are seemingly beyond the capabilities of the available measuring hardware. One solution to this problem is to improve the measuring process by replacing the existing equipment with instruments of better quality. While this may often be a most desirable move, it frequently is not a practical one. Consider, for example, the instrumentation situation of an environmental test facility. Test programs are regularly encountered which include requirements for measuring large numbers of temperatures, voltages, forces, or other data parameters. A multi-channel digital data acquisition system is usually employed to satisfy this measurement need. Systems of this type are large, complex, and expensive, and obviously cannot be easily replaced; yet there may be times when the accuracy desired in the measurement of a test parameter exceeds that which can be justly associated with a single system reading of the parameter. The objective of the work described in this study is to establish a method, using the statistics of replicate measurements, to improve the ultimate quality of data acquired on a digital data acquisition system.

As one might intuitively suspect, the average of

several readings of a parameter proves to be a better representation of the true value than a single reading. This fact serves as the basis for development of a procedure that uses the average and the standard deviation of repeated measurements of a data parameter to establish an interval within which the true value of the parameter may be expected to lie. The probability associated with this expectation can also be determined. A plot of the confidence interval expression, with the number of readings averaged as the variable, shows that a significant improvement in precision can be realized from the replicate measurement process. The fact that this improvement is in precision, and not necessarily in absolute accuracy, is worthy of note, and will be discussed later in more detail. It should also be pointed out that improvement of the data by this method is not without cost. It is accomplished at the expense of (1) an increase in time required to acquire the replicate readings, (2) a significant limitation on the rate of change of the input data, and (3) an added amount of computation in the data reduction process. There are, however, many data taking situations where these restrictions can be tolerated, and the statistical process can be used to great advantage.

As a method of testing the validity of the confidence interval statistics, a computer model of a data acquisition process was developed and run on a Xerox 9300 computer. A random number generator was used to simulate variations (noise) in the data. Consideration was given to the

distribution of the data dispersion; that is, the confidence interval curves associated with a normal (Gaussian) distribution were compared with those of a uniform distribution. In order to make these comparisons, a relationship between maximum data dispersion (range) and standard deviation for a uniform distribution was needed. An expression was developed analytically, and modeled on the computer, which shows that standard deviation approaches a linear function of range as the number of data readings increases towards infinity.

An experimental exercise was conducted using a digital data acquisition system in the von Karman Facility (VKF) at the Arnold Engineering Development Center. Confidence intervals were calculated for different input configurations, and a series of test runs were made with known data applied to the system. Average values were calculated and the results are compared with the statistical predictions. The computer program, written to perform the calculations, includes a routine for eliminating erratic data readings caused by electrical noise spikes, temporary system malfunctions or other transient irregularities, that occasionally appear. Thompson's Tau technique was used for eliminating these outlying readings.

The primary emphasis of this work was placed on handling techniques for very low frequency (quasi-static) data; however, some consideration was given to the possibility of applying similar procedures to non-static data. Additional work should be done in this area to determine

the usefulness of the replicate measuring technique with time varying input data. The potential advantages must be weighed against input frequency limitations that would be associated with acceptable confidence intervals.

CHAPTER II

ANALYSIS

I. MEASUREMENT ERROR

The error ε of any measurement X may be defined as the difference between the value X and the true value τ of the magnitude of the quantity concerned. Thus $\varepsilon = X - \tau$. The numerical value of ε will usually be unknown (and unknowable), since the true value τ is rarely a knowable quantity. However, certain inferences can be made regarding the bounds or limits of the error ε.

Statisticians have classically defined two types of error, (1) systematic error, and (2) errors of precision. If μ denotes the mean of repeated measurements of magnitude of a quantity, the systematic error (also frequently referred to as bias) is defined as the difference between μ and the true value, i.e., $\mu - \tau$. The systematic error of a measuring process may be thought of as its tendency to measure something other than was intended. Systematic errors will frequently have both constant and time variable components. Zero offset or scaling errors, for example, will bias all measurements made on the affected system. Variations caused by changes in temperature or atmospheric conditions introduce a bias which will be time dependent. Fortunately, the cause of systematic errors can frequently be

identified and they can either be removed or minimized by calibration.

Many authors define accuracy as the absence of bias. Complete accuracy then would be the total absence of bias, i.e., $\mu - \tau = 0$. No measuring process, however refined, is completely accurate. Nor can it be expected to produce a succession of identical numbers (true value or otherwise) upon repeated measurement of the same quantity. This was stated more eloquently, if somewhat less succinctly, by Eisenhart (1):[1]

> In what sense can we say that the measurements yielded by a particular measurement process serve to determine a unique magnitude, when experience shows that repeated measurement of a single quantity by this process yields a sequence of nonidentical numbers?

The magnitude of this disagreement between repeated measurements of a single quantity is an indication of the precision of the measuring process. Lack of precision, that is, imprecision, is evidenced in the amount of scatter exhibited in the results. Precision therefore is not related to true value, but only to the degree of conformity between individual measurements. It has been said (2) that accuracy has to do with closeness to the truth, and precision only with closeness together. Hillebrand (3) used similar language when he stated that "precision is a measure of reproducibility and accuracy is a measure of correctness."

---

[1] Numbers in parentheses refer to similarly numbered references in the bibliography.

Both of these parameters are of significant importance to the user of a measuring device. His ultimate concern is for the accuracy or "correctness" of the data. However, the system accuracy is constrained by the lack of repeatability between readings. System precision therefore is of equal concern. As a graphic illustration consider Figure 1. The curves show the distributions of measurements yielded by different hypothetical measuring processes. Curves A, B, and C are centered around the true value τ, indicating that these processes have no bias--the systematic error is zero. It is clear that the process of highest precision, C, is also the most accurate, since a larger percentage of the data measurements made by this process are closer to the true value than those made by processes A and B. Process A is obviously inferior to B and C; and the probability is high that a single measurement made by process A will be less accurate than a single measurement made by B or C. Results obtainable from process A can, however, be significantly improved by taking the average of a number of replicate measurements. In fact, it can be shown (1) that averages of 12 readings taken by process A will have the distribution of curve B, and averages of 144 readings by process A will increase the precision to that of curve C.

II. SAMPLING STATISTICS

In the terminology of applied statistics each reading produced by a measuring process is a member of a statistical

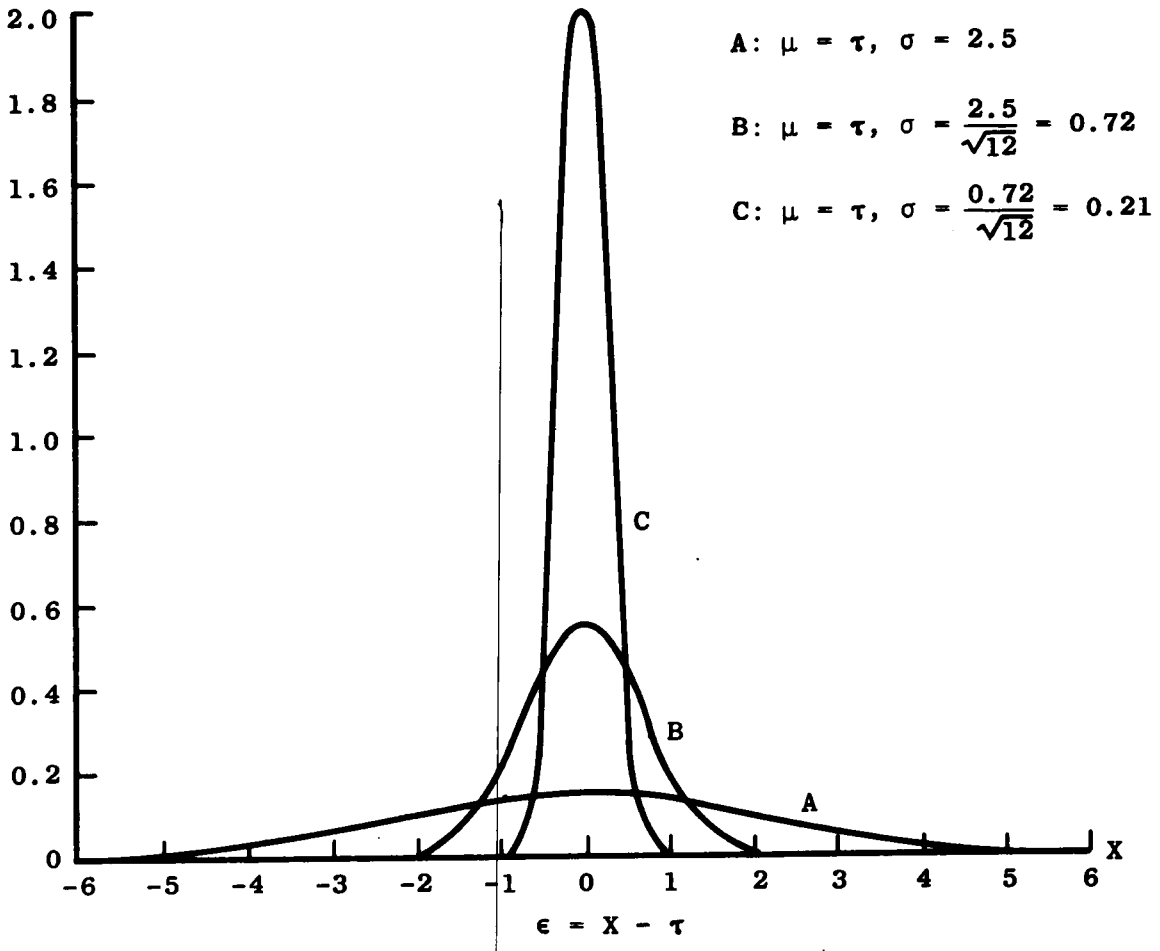


Figure 1. Distribution of errors of unbiased measurement processes.

Source: C. Eisenhart, "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems," Journal of Research of NBS, 67C:161-187, April-June, 1963.

population (or universe). In theory the process could be continued indefinitely forming a hypothetical infinite population. Since it is impossible to determine the value of every member of an infinite population, the readings obtained by any practical measuring process, regardless of their number, must be considered a statistical sample. The science of statistics establishes relationships between samples and populations which allow one to judge the reliability of the sample as a representation of the total population. The necessary prerequisite for application of these relationships is that the sample selection process satisfies the condition of randomness; that is, the samples must show statistical control. Denning (4) explains it this way:

> No . . . statistical theory (can be) applied unless the observations show statistical control. In this state the samples when cumulated over a suitable interval of time give a distribution of a particular shape, and this shape is produced hour after hour, day after day . . . so long as the process remains in statistical control, i.e., exhibits the properties of randomness. In a state of control n observations may be regarded as a sample from the universe of whatever shape it is. A big enough sample, or enough small samples, enables the statistician to make meaningful and useful predictions about future samples.

Quantities of specific interest in sampling statistics are the mean (μ) and variance (σ^2) of a distribution. The latter quantity (or its square root, the standard deviation) is the most commonly used measure of precision. Since the actual population parameters, μ and σ, are usually unknown, they must be estimated from sample observations. A commonly used method of making these estimates is well

described in the literature, for example, by Bendat and Piersal in Reference (5).

If X_1, X_2, X_3, . . . X_N represent a random sample of size N from any population, an estimate of μ, the population mean, is given by the arithmetic average of the sample, thus:

$$\text{Estimate of } \mu = \bar{X} = \frac{\sum_i X_i}{N} \tag{1}$$

The sample estimate of the standard deviation, represented by the symbol S, is:

$$\text{Estimate of } \sigma = S = \left(\frac{\sum_i (X_i - \bar{X})^2}{N - 1} \right)^{1/2} \tag{2}$$

III. CONFIDENCE INTERVALS

By using estimates of sample mean and standard deviation, a confidence interval can be established for the mean μ of a normally distributed (Gaussian) population. The concept of the confidence interval is useful since it provides an estimation of an interval which will include the true value of the parameter being measured, with a known degree of uncertainty. For the case of a mean value estimate, a mathematical probability statement for the confidence interval based on $\bar{X}$ and S can be shown (5) to be:

$$\text{Prob}\left[\bar{X} - t \frac{S}{\sqrt{N}} \leq \mu \leq \bar{X} + t \frac{S}{\sqrt{N}} \right] = \gamma \tag{3}$$

where γ is the confidence coefficient and t is a distribution function explained below. Equation 3 indicates that the probability is 100 γ percent that the true value μ falls within the interval $\bar{X} \pm t \frac{S}{\sqrt{N}}$. Or, stated in slightly different terms, if many replicate samples of N measurements were taken, and $\bar{X}$ and S were computed for each sample, one can expect the interval $\bar{X} \pm t \frac{S}{\sqrt{N}}$ to include the true value 100 γ percent of the trials.

The confidence interval depends on a probability distribution function known as "Student's t" (6). This distribution, unlike the standard normal distribution, is a function of the size of the sample associated with $\bar{X}$ and S. Thus, tables of Student's t, which can be found in any statistical text or handbook, contain a parameter called "degrees of freedom." If the confidence interval is based on a sample of size N the number of degrees of freedom is N - 1. The value of t is also dependent on the level of probability to be established on the confidence interval--the higher the confidence coefficient γ, the greater the value t. Table I is an excerpt from a standard Student's t table, listing values of t for γ equal to .80, .95, and .99.

The t-distribution is particularly applicable in situations where the sample size N is small. For values of N greater than 40 the normal distribution gives a good approximation of the t-distribution. In fact, it can be shown that the t-distribution tends to the normal distribution as N approaches infinity. A small confidence interval,

TABLE I

TWO TAILED STUDENT'S "t" TABLES

Degrees of Freedom	Value of t		
	$\gamma = 0.800$	$\gamma = 0.950$	$\gamma = 0.990$
1	3.078	12.706	63.657
2	1.886	4.303	9.925
3	1.638	3.182	5.841
4	1.533	2.776	4.604
5	1.476	2.571	4.032
6	1.440	2.447	3.707
7	1.415	2.365	3.499
8	1.397	2.306	3.355
9	1.383	2.262	3.250
10	1.372	2.228	3.169
11	1.363	2.201	3.106
12	1.356	2.179	3.055
13	1.350	2.160	3.012
14	1.345	2.145	2.977
15	1.341	2.131	2.947
16	1.337	2.120	2.921
17	1.333	2.110	2.898
18	1.330	2.101	2.878
19	1.328	2.093	2.861
20	1.325	2.086	2.845
30	1.310	2.042	2.750
∞	1.282	1.960	2.576

Source: J. Mandel, The Statistical Analysis of Experimental Data. (New York: John Wiley and Sons, Inc., 1964).

of course, is desirable; and, as Equation 3 indicates, the interval decreases as the number of measurements N is increased. Actually, the expression $S/\sqrt{N}$ is referred to in statistics (6) as the "standard error of the mean" and represents the improvement in distribution that can be achieved by averaging replicate measurements.

As stated earlier, the objective of this work is to establish a procedure for improving the usefulness of a data acquisition system by taking advantage of the statistics of replicate measurements. In a practical system, it may well be possible to effectively minimize the systematic error by performing a system calibration against a reference standard. Assuming that this can be done, random fluctuations in the data caused by system noise become the dominant error source. The system precision is therefore of primary concern, and a quantitative measure of expected performance is available in the confidence interval. It remains then to establish a value for the expression $t \frac{S}{\sqrt{N}}$ given in Equation 3. Of these parameters, the only real unknown is the estimated standard deviation S, since N is the number of replicate measurements used to obtain the average $\overline{X}$, and the statistic t, for various confidence levels (γ), is available from a table.

As previously mentioned, the value of S can be computed from Equation 2 if a number of sample measurements are made. However, another very simple method can be used to provide a good approximation of S.

If range is defined as the difference between the

largest and smallest values of a sample of N observations, the approximate value of the standard deviation can be obtained by dividing the range by a quantity referred to in the literature on quality control statistics (7) as the mean relative range. The standard symbolic representation for mean relative range is d_2. A listing of d_2 values applicable to a normal distribution is given in Table II.

Denoting the range by R, the estimate of standard deviation is given by:

$$S = \frac{R}{d_2} \tag{4}$$

This proves to be an efficient estimate since, in small samples from a normal population, the range and standard deviation are likely to fluctuate together.

In addition to providing an acceptable estimate of standard deviation, the range is relatively easy to determine. Consider an operating data acquisition system making repeated measurements of a non-changing parameter. The difference between the highest and lowest readings observed over a selected time period is a measure of the range. An improved measure, and one more consistent with the statistical basis of the ensuing calculation, is that determined by taking the average of these reading excursions over a number of identical time periods. This average range may be represented by $\overline{R}$.

Assume, as an example, that a measuring system has a

TABLE II

MEAN RELATIVE RANGE

Sample Size N	Mean Relative Range d_2	$1/d_2$
2	1.128	.887
3	1.693	.591
4	2.059	.486
5	2.326	.430
6	2.534	.395
7	2.704	.370
8	2.847	.351
9	2.970	.337
10	3.078	.325
11	3.173	.315
12	3.253	.307
13	3.336	.300
14	3.407	.294
15	3.472	.288
16	3.496	.286
24	3.802	.263
32	4.098	.244
40	4.291	.233
56	4.651	.215
72	4.830	.207
96	5.025	.199

Source: A. J. Duncan, Quality Control and Industrial Statistics. (Homewood, Illinois: Richard Irwin, Inc., 1969).

range of 16 counts (the units are of no particular significance in this illustration; however, in the terminology associated with data acquisition systems, they would normally be in counts which are proportional to input voltage). Calculations of conficence intervals for various values of N can now be made using Equations 3 and 4 and Table I, page 12, and Table II.

$$CI = 2t \frac{S}{\sqrt{N}} = 2t \frac{\bar{R}}{d_2} \frac{1}{\sqrt{N}} \tag{5}$$

A typical calculation shows that for a sample size N = 8 and a confidence coefficient $\gamma = .95$,

$$CI = (2)\ (2.365)\ \frac{16}{2.847} \frac{1}{\sqrt{8}} = 9.39 \text{ counts}$$

The factor of 2 is required to account for the plus and minus excursions indicated in Equation 3. Figure 2 shows normalized 80, 95, and 99 percent confidence intervals associated with the averages of replicate sample sizes ranging from N = 4 to N = 64. Taking N = 20 and $\bar{R}$ = 16 as an example, it can be determined from Figure 2 that the probability is 95 percent, that the average value of 20 replicate readings produced by this presumed system will be within ±2 counts of the true value. This assumes that systematic errors can be neglected.

Normal Versus Uniform Distribution

In the derivation of the expression for confidence

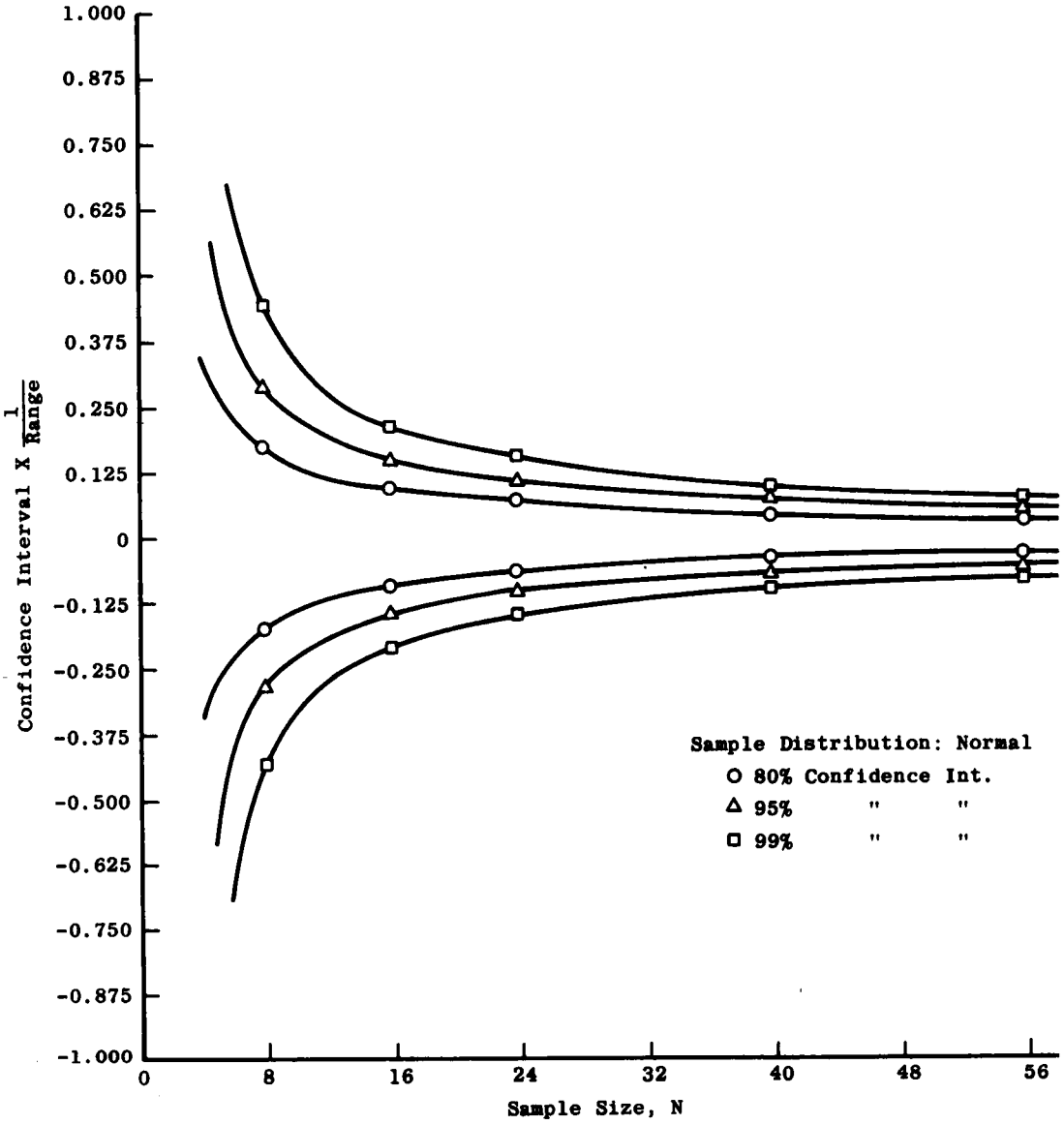


Figure 2. Confidence intervals for sample averages.

interval, Equation 3, one of the justifying criteria is that the population under consideration obeys the standard normal distribution. More specifically, the Student's t statistic, which is the ratio of an observed deviation from the mean, $\bar{X} - \mu$, to its estimated standard error, $S/\sqrt{N}$, is applicable only when the X population is normally distributed. This apparent restriction on the use of the confidence interval equation can be overcome by taking advantage of a property of average values that is described in statistical literature by the central limit theorem. Very simply this theorem states that the distribution of sample averages approaches normality, and the larger the sample, the closer is the approach. This is expressed more precisely by Mandel (6):

> Given a population of values with a finite variance, if we take independent samples from this population, all of size N, then the population formed by the averages of these samples will tend to have a Gaussian (normal) distribution, regardless of what the distribution is of the original population; the larger N, the greater will be this tendency towards normality.

The complete generality of this theorem provides justification for an assumption of normality for the sampling distribution of $\bar{X}$, without regard for the distribution of the base population. Therefore, if the statistic t is replaced by the standard normal deviate Z (6), the confidence interval equation can be applied when the distribution of X is not Gaussian. The normality assumption becomes reasonable for $N > 4$ and quite accurate for $N > 10$.

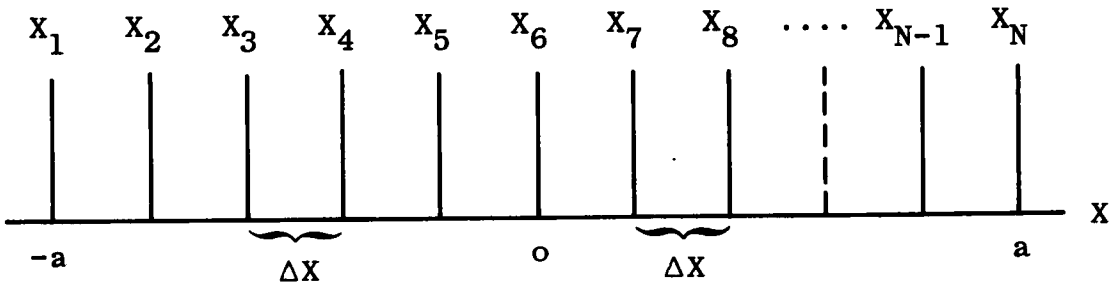
There is, however, one additional consideration. The

use of Equation 4, $S = \frac{R}{d_2}$, as an estimator for standard deviation is also based on the premise that the population is Gaussian. In a data system application this would mean that the distribution of the individual readings used to compute the average must be presumed to be normal. Although this may not appear to be a wholly unreasonable assumption, a more cautious approach requires a less optimistic view. Perhaps the system produces data readings with a pronounced tendency toward a uniform distribution. To account for this possibility, it was decided to pursue the development of a method for using the range to estimate the standard deviation of a uniformly distributed population, and ultimately compare results with those obtained using Equation 4.

As a point of beginning consider Equation 2, the expression for estimating the variance (square of the standard deviation) of a random sample.

$$\text{Variance of } X = S^2 = \sum_{i=1}^{N} \frac{(X_i - \mu)^2}{N - 1} \qquad (6)$$

This will be applied to the values of X given in Figure 3 which are shown to be uniformly distributed between -a and +a. The separation between adjacent X values is equal to ΔX and the number of positive X values is identical to the number of negative X values. The average value of X must therefore equal zero. If ΔX is decreased and the number N of X values thereby increased ($N = 2a/\Delta X + 1$), the above conditions remain unchanged. From an analytical standpoint



Number of X Values (N) = $(2a/\Delta X) + 1$

Figure 3. Uniform distribution of X.

this population may be considered the equivalent of one with a random uniform distribution centered around zero, since in the latter, the probability of occurrence of any one value is equal to the probability of occurrence of any other, and the average of all values approaches zero as N increases.

By substituting for N and setting $\mu = 0$ in Equation 6, an expression for estimating the variance of a sample from a uniformly distributed population is obtained.

$$S^2 = \sum_{i=1}^{2a/\Delta X} \frac{X_i^2}{2a/\Delta X} \tag{7}$$

As ΔX approaches zero in the limit, the summation is replaced by an integral and

$$S^2 = \int_{-a}^{+a} \frac{X^2}{2a}\, dX$$

$$= \frac{1}{a} \int_{o}^{a} X^2 dX$$

$$= \frac{a^2}{3}$$

and consequently,

$$S = \frac{R}{2\sqrt{3}} \tag{8}$$

where $R = 2a$ is the range, which has been previously defined as the difference between the maximum and minimum sample

values.

Equation 8 indicates that the standard deviation of a random sampling from a uniformly distributed population with $\mu = 0$ approaches a limit as the number of samples N increases.[2] That limit is directly proportional to the range.

A means of estimating standard deviation from the range for smaller values of N was obtained by solving Equation 7 for S, using values of ΔX ranging from 2 to 20,000. A Fortran language program for making this computation was prepared (see Appendix B) and run on the VKF Xerox 9300 computer. The results are shown in Figure 4 as a plot of N versus the standard deviation, normalized for a value of $R = 1$.

The curve is seen to approach the constant $\frac{1}{2\sqrt{3}}$ for large N, as anticipated from Equation 8. Also of interest is the indication that very little change in standard deviation would be expected to occur as the number of sample increases beyond 25 or 30. The second curve of Figure 4 was obtained by plotting the reciprocals of the d_2 factors (see Equation 4), applicable for a normal distribution, against the sample size N.

Values taken from these curves can be multiplied by the range R to obtain an estimate of the standard deviation of N samples derived from a homogeneous population. The

---

[2]The derivation of a general expression for the variance of a uniformly distributed population is given in Appendix A.

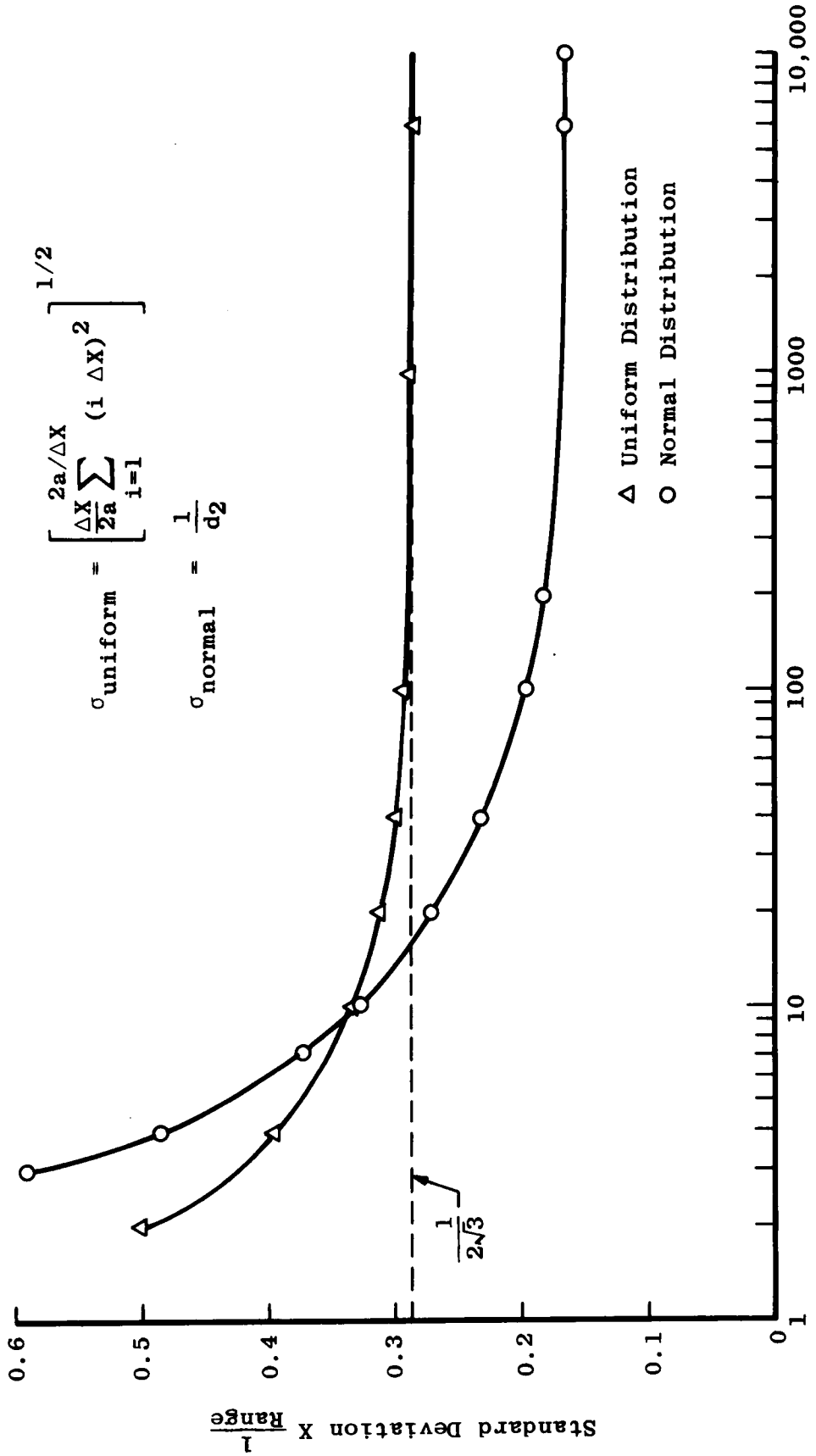


Figure 4. Curves for estimating standard deviation.

distribution of the population under investigation would, of course, determine which curve is applicable.

As a comparative example, the 95 percent confidence intervals for range $\overline{R}$ = 16 were calculated for a uniform distribution. Equation 5 was used, with values taken from the curve for a uniform distribution (Figure 4) being substituted for $1/d_2$ and with the standard normal deviate for 95 percent probability, Z = 1.96, replacing t. The results are shown in Figure 5, together with the previously determined 95 percent confidence interval plot for a normal distribution.

The difference between the two curves is small. Certainly it can be regarded as insignificant when the computational basis and ultimate use of the curves is considered. However, since the confidence intervals predicted for a uniform distribution are more conservative (however slightly) for large N, and since the theory is subsequently found to be pessimistic for small N, these values are used in all subsequent references to confidence interval.

The favorable comparison of the curves in Figure 5 suggests that the distribution of the population of readings derived from a measuring process does not significantly influence the precision of the averages of samples of N readings made by that process. This fact is used to advantage in a computer modeling exercise carried out to test validity of the confidence interval determinations, since the uniform distribution proves to be more convenient to

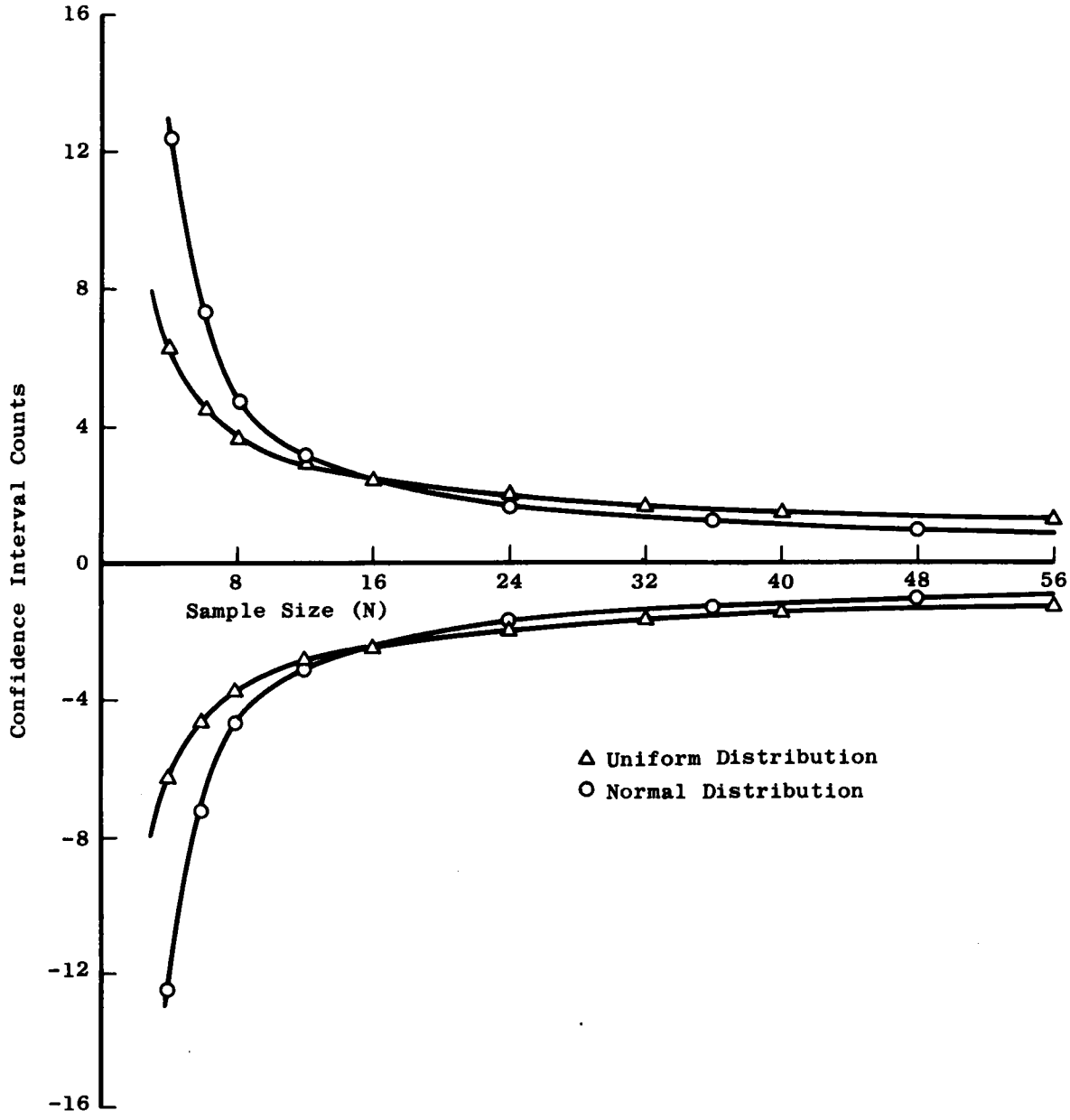


Figure 5. Ninety-five percent confidence interval comparison.

work with than the normal distribution.

IV. COMPUTER MODELING

In an effort to establish some measure of confidence in the statistics presented in the preceding section, a model was developed to verify the results. The model took the form of a computer program, written to simulate a data acquisition process and compute average values and standard deviations for different sample sizes. The simulation was accomplished by adding varying components, representing system noise, to a constant data value. The "noise" source was a random number generator which produced a non-repeating sequence of values between 0 and 1. The maximum noise level (range) could be easily varied by changing a multiplier applied to the random numbers. The simulated data were quantized to more closely represent the integer counts produced by a real digital data system. A detailed description of the computer program is given in Appendix C.

Repeated computer runs were made, calculating averages of the simulated data for selected values of sample size N and range R. The results of approximately 40 such runs with R = 16 are shown in Figure 6. Also included in the figure are the 95 percent confidence intervals which had been previously determined (Figure 5) for a measuring process with this range. The distribution of the averages of the artificially produced data is reasonably well contained within the 95 percent confidence limits. Indeed, if those

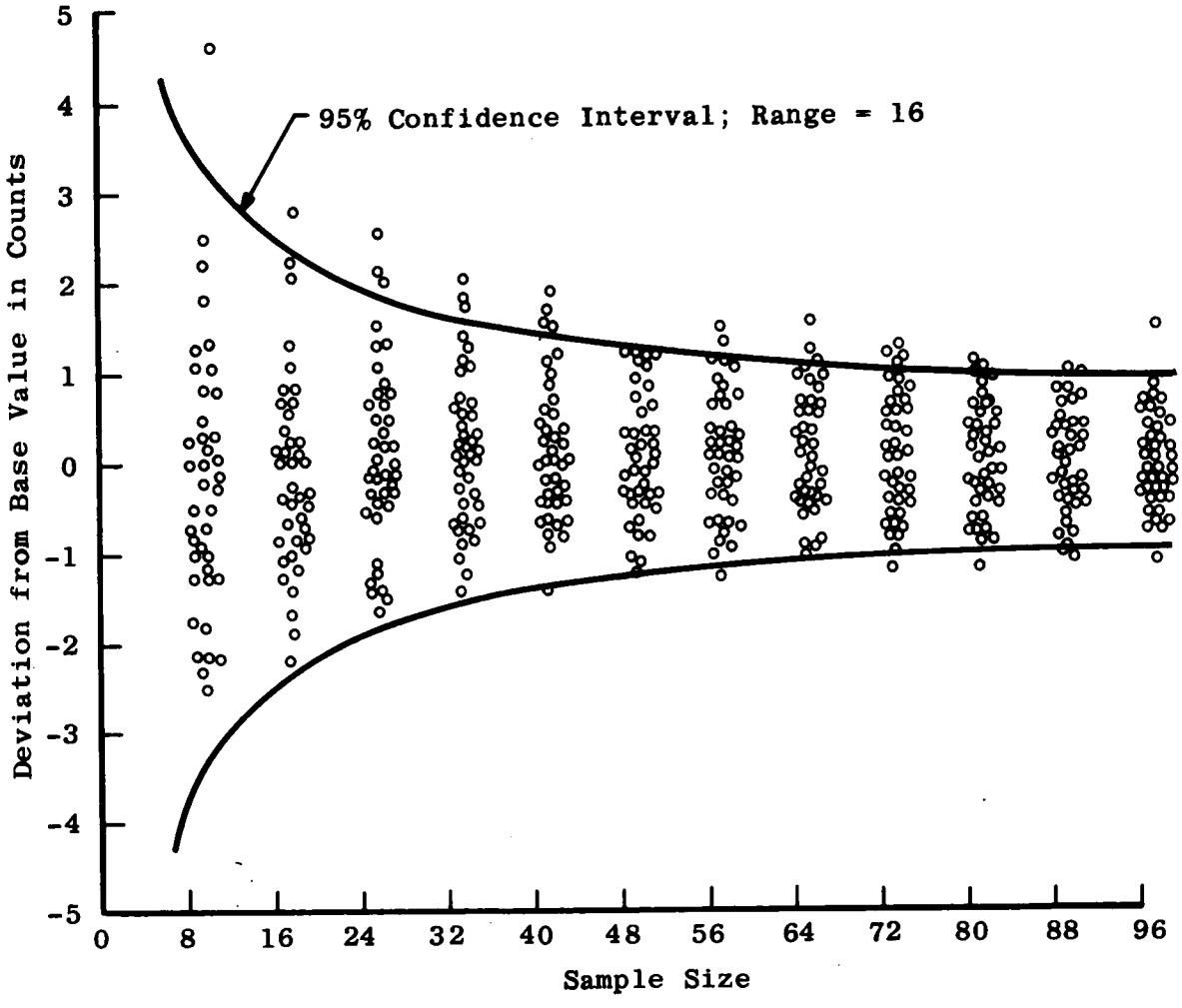


Figure 6. Distribution of sample averages (computer model).

points exceeding these limits are counted, they can be seen to represent approximately 5 percent of the total.

In a supplementary exercise, the computer modeling program was used to calculate 1,000 averages for each sample size, and to make a comparison between each average and the appropriate confidence limits. The number of average values exceeding the limits for the various sample sizes was accumulated. A uniformly distributed random number population was used to simulate the noise source. The resulting data are shown in Table III. The tabulated figures indicate the percentage of the 1,000 average readings that fell within the 95 percent confidence interval curves. Agreement between theory predictions and the modeling results is reasonably good. With a uniformly distributed noise source the percentage of values within range tends to be slightly higher than 95 percent, indicating that the confidence interval theory may be somewhat pessimistic. This is particularly true for the smaller sample sizes. Since the discrepancy is in the conservative direction, it is not really disturbing. One can thus have more positive assurance that the calculated confidence limits will contain the data a larger percentage of the time.

The above results tend to support statistical theory and indicate the methods used in deriving confidence intervals are sound.

As an additional interesting exercise, it was decided to use the computer modeling technique to test Equation 8,

TABLE III

SAMPLE AVERAGES WITHIN CONFIDENCE INTERVAL

Sample Size	Percent within Interval
8	98.7
16	97.2
20	97.8
24	96.9
28	97.1
32	96.8
36	96.2
40	96.5
44	95.1
48	95.1
52	95.0
56	95.5
60	94.6
64	94.9
68	95.2
72	95.1

$S = R/2\sqrt{3}$. This is the expression developed for the limiting value of the standard deviation of a uniformly distributed population as estimated from the range. A program was prepared (see Appendix D) to compute averages and standard deviations of a large number of simulated data values (N = 8,000) for ranges varying from 2 to 20 counts. Again the random number generator was used to produce an artificial noise for modulating a base data value.

The results are shown in Figure 7 as a plot of standard deviation versus range. A linear curve fits nicely through the plotted points; and the slope of the curve is remarkably close to $\frac{1}{2\sqrt{3}}$. The extent of the comparison is further evident in the tabulation of results from the computer modeling and those calcualted from Equation 8 also shown in the figure.

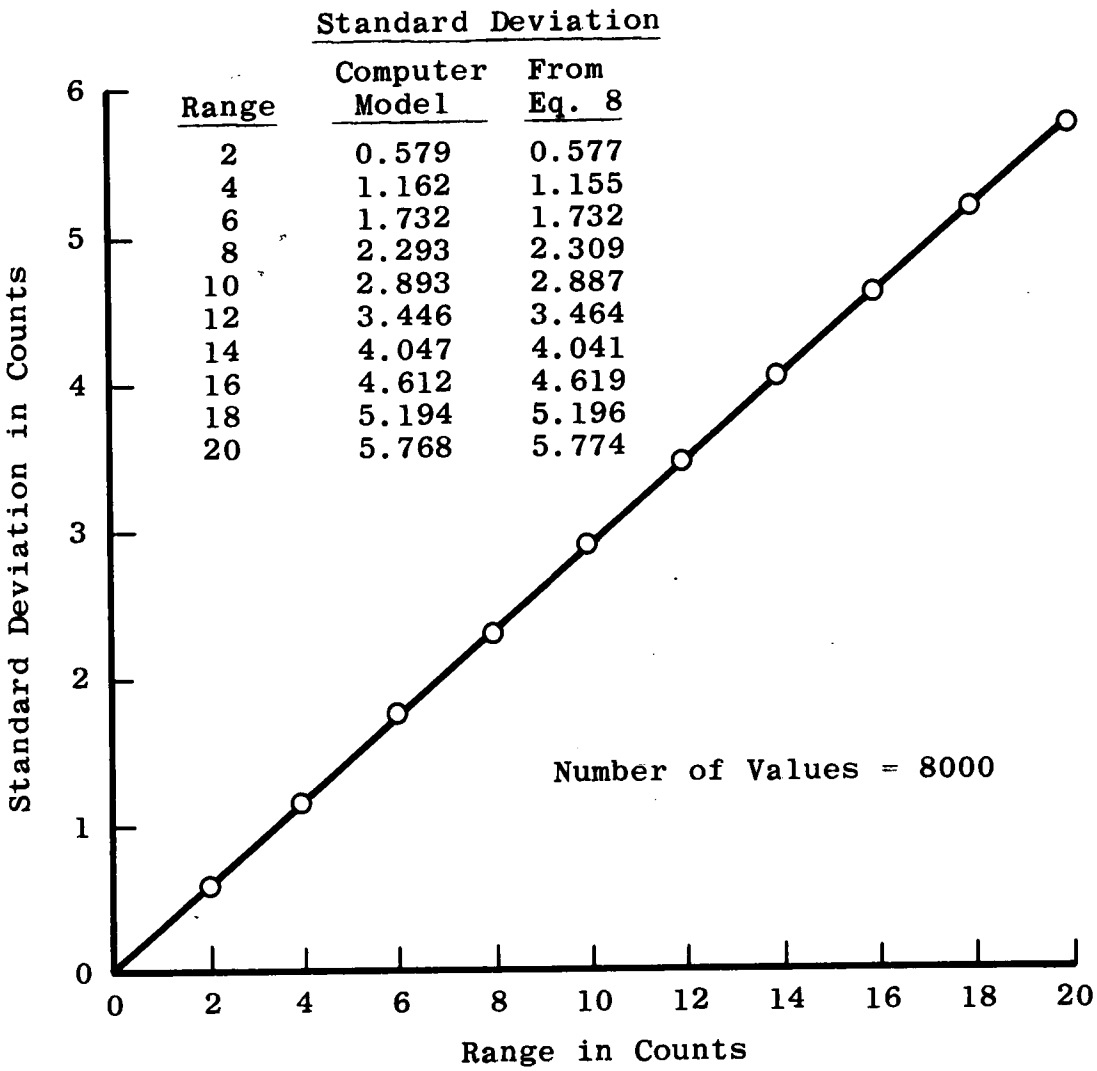


Figure 7. Standard deviation of a uniform distribution (computer model).

CHAPTER III

EXPERIMENTAL EXERCISE

The digital data acquisition system of concern in this work is used to support test programs in the VKF aerospace environmental chambers at the Arnold Engineering Development Center. Tests are typically conducted on the performance of models, components or space vehicles as they are subjected to environments similar to those encountered during space flight. Usually the test programs include requirements for making large numbers of measurements of test article conditions at frequent intervals throughout the course of the test. The VKF data system, shown in block diagram form in Figure 8, was procured to satisfy this requirement. The system incorporates a number of sophisticated features which are described in Reference (12), but for the purposes of this experimental exercise, we are interested primarily in the low level analog input section. Millivolt level signals can be applied to differential preamplifiers, the outputs of which are fed through a high speed, solid state main commutator to an analog to digital converter; there are 50 of these preamplifier channels available. Or the input signals may be routed through lower speed, mechanical relay subcommutators before being amplified and passed through the main commutator. Using the latter method of operation, nearly 1,000 individual data

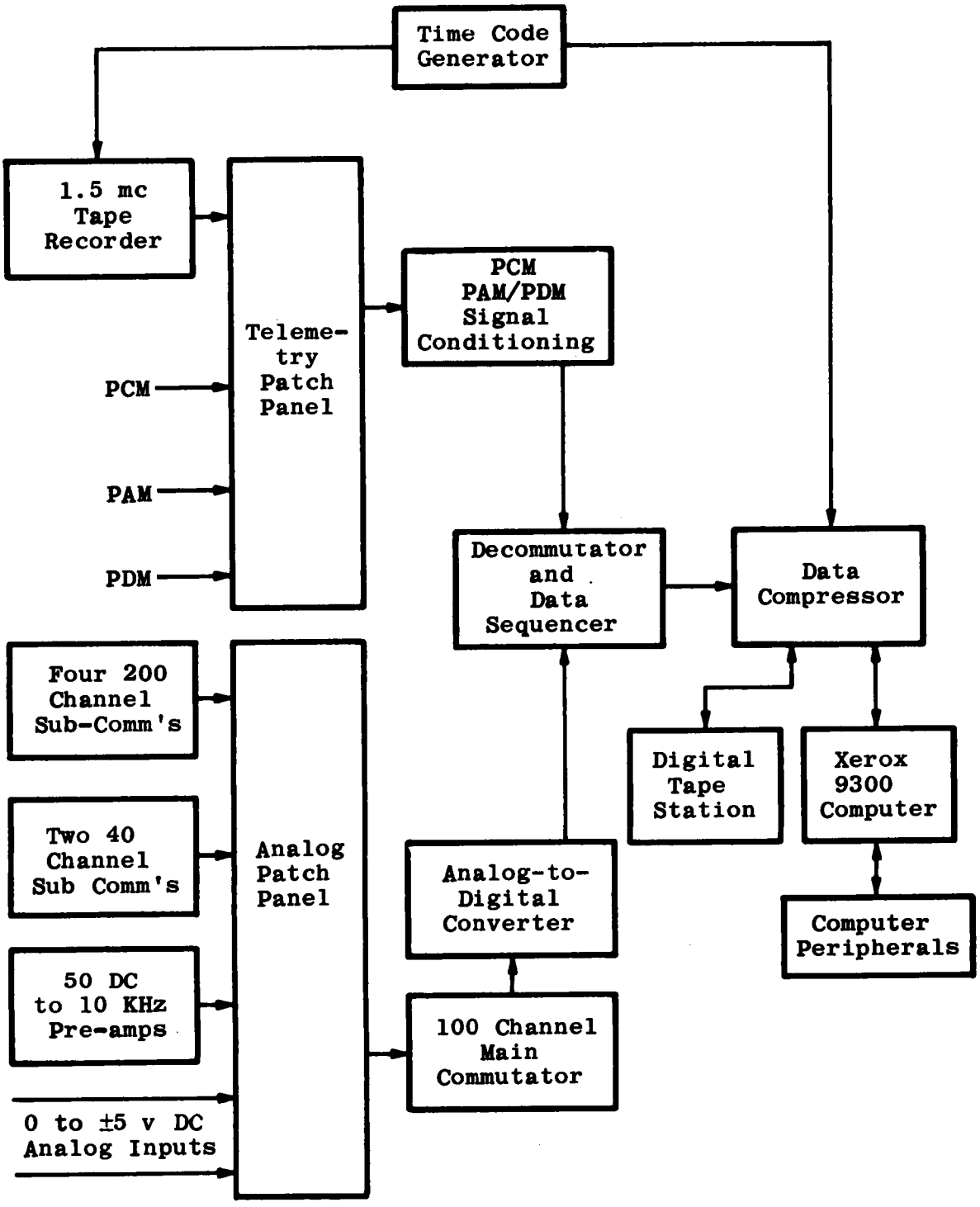


Figure 8. VKF data processing system.

channels can be sequentially processed through the system, recorded on magnetic tape and applied to a digital computer.

The data system resolution is ±2047 counts for full scale input, which may be scaled to accommodate input levels ranging from ±5 millivolts to ±5 volts. The noise level on data processed through the preamplifier/main commutator route (hereafter referred to as main commutator data) is quite low. Random variations of ±1.5 counts in the digital output with a constant input, and with the most sensitive amplifier gain settings, are considered normal. This would be equivalent to a range of 3 in the terminology used in the preceding chapters. The subcommutator data are more noisy, by almost an order of magnitude. Variations of ±8 to 10 counts may be expected, with the additional noise being introduced by the mechanical relay contacts. The impedance of the input transducers and the quality of the interconnecting cabling installation also have an effect on the noise level.

To obtain an indication of the significance of these noise variations, consider a measurement made on the output of a copper-constantan (C_u-C_n) thermocouple. A ±1.5 count fluctuation in the digital output (when the gain is set for ±5 millivolts full scale) is equivalent to an uncertainty of approximately ±0.2 °F in the temperature reading. Similarly, a ±8 count noise level causes nearly a ±1 °F temperature uncertainty.

In most test programs involving large numbers of

measurements, the thermal profile of the test article in certain specified environments is of primary concern. Thermocouples are used in abundance to provide the desired temperature indications. Normally, the ±1 °F precision of data system readings will satisfy the test requirements; and the majority of the temperature measurements can be made through the subcommutator channels. Should a few of the parameters be particularly critical, they can be applied through the main commutator channels, where more precision is obtainable. It must be emphasized again that the above numbers refer to precision and not to absolute (true value) temperatures. The system must be calibrated against a traceable standard to establish a reference to absolute value. In thermocouple measurements, errors associated with the isothermal reference junction, extension wire, the thermocouple attachment process, and the thermocouple itself must be considered in arriving at a value for the absolute accuracy of a temperature reading.

There are measuring situations however, where precision is most important-- where the change in voltage or temperature from a base reading is of primary interest rather than the absolute magnitude. It is in these situations that the statistical improvement obtainable by averaging replicate measurements can be used to advantage.

Figure 9 shows the 95 percent confidence interval curves, established by the methods detailed in Chapter II, that are applicable to a C_u-C_n thermocouple measurement on

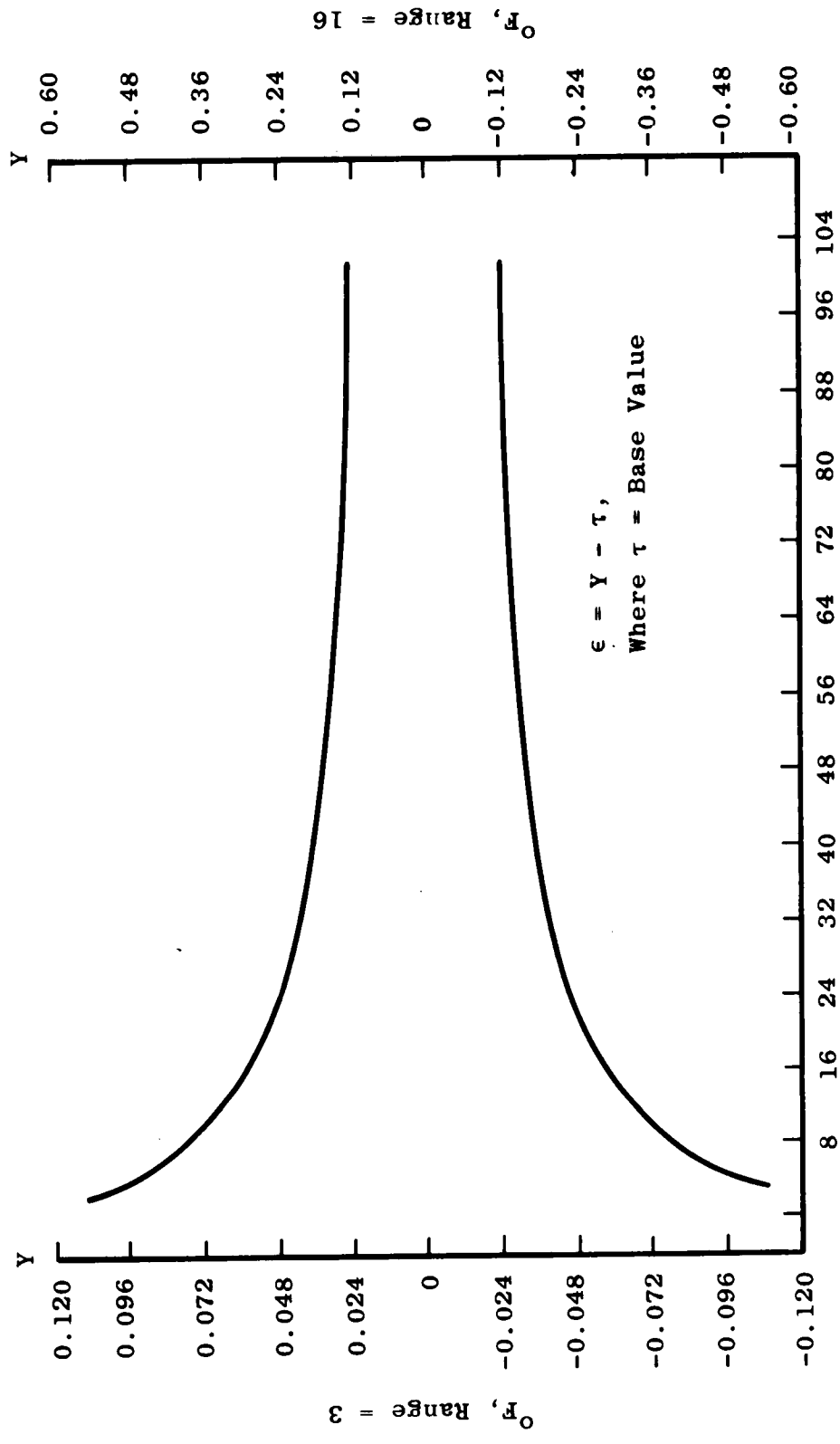


Figure 9. Ninety-five percent confidence interval, C_u-C_n thermocouple measurement.

the VKF data system. The ordinate is labeled with two temperature scales, the left hand scale representing a range of 3 (±1.5 counts) and the right scale a range of 16 (±8 counts). These values will be recognized as the nominal noise figures associated with the data system main commutator and subcommutator channels, respectively. It can be seen from the figure that the precision expected by averaging 80 main commutator readings is ±0.024 °F, very nearly an order of magnitude improvement over the single reading uncertainty of ±0.2 °F. For smaller sample sizes, the expected precision decreases; however, a factor of two improvement over the single reading uncertainty is realized by averaging only eight samples.

I. EXPERIMENTAL PROCEDURE

To test the statistical theory in a practical measuring situation, a number of test runs were made on the VKF system. In one exercise four thermocouples were attached to a metal block mounted inside one of the VKF environmental chambers. The existing cabling and reference junction system was used to connect the thermocouples to the data system approximately 500 feet away. Two of the thermocouple outputs were applied to the system through subcommutator channels and two through main commutator channels. Several data scans were taken, reading and recording the output of each thermocouple several dozens of times. Each few moments the process was repeated, with all of the data

being recorded on a digital magnetic tape. No effort was made to control the temperature of the thermocouple block, since it was a sizeable mass that had been allowed to assume the ambient chamber temperature, and could be considered an isothermal body for the less than two second duration of a data scan. The presumption was that variations in successive readings of an individual channel, during a particular scan, could be attributed exclusively to noise within the system, between the thermocouple output and the recorded reading, and not to changes in the block temperature being measured.

II. DATA ANALYSIS

Analysis of the data thus accumulated was carried out using the VKF Xerox 9300 computer. A program was written to sort each recorded data record and store the readings of the individual input channels in the computer memory, in blocks identified with the appropriate channel numbers. It was then possible to recall the data in the same chronological sequence in which it was recorded, and make the computations necessary for data analysis.

Prior to each series of data runs, zero readings were recorded on the data tape by making a number of scans through the data channels with the input circuits shorted. The program was written to use this data to compute average zero values for each data channel. These values were subsequently recalled from storage and subtracted from all readings of the appropriate channels to minimize biasing errors that

might be caused by zero offsets in the preamplifier or commutator.

The average value and standard deviation of cumulative readings of each data channel were calculated. These averages were ultimately converted to temperatures, using a thermocouple table stored in the computer memory. Computations were made on increments of data, beginning with that originally acquired first and proceeding in chronological sequence. Each succeeding increment was added to that already processed. In this way information on the pattern of behavior of the average value relative to the number of samples in the average could be obtained. By making these computations on repeated runs, which had been conducted under nearly identical conditions, data were available for comparing the performance of an operational system to that which might be predicted from the confidence interval theory.

Subcommutator data from four separate runs are included in Figure 10. The indicated bands represent the maximum spread observed in the averages of readings taken through two different subcommutator channels. Since the true value of the temperature being measured was not known, the final readings of each run were averaged to establish a base temperature. The data in Figure 10 represent variations from this base. This procedure is not completely rigorous since it does not account for any systematic error that may exist between the two data channels. The effect of a systematic error would be to concentrate the preponderance

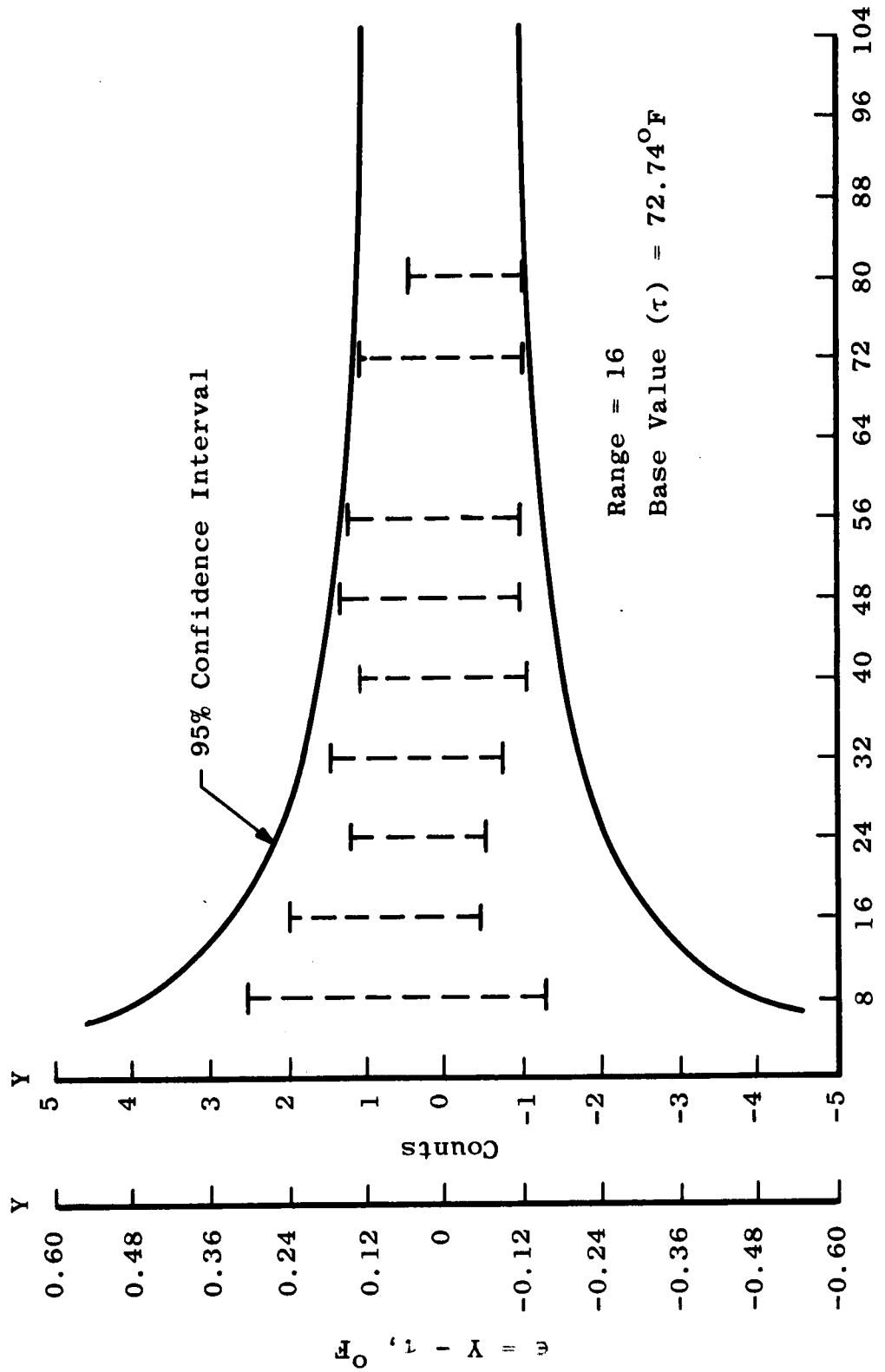


Figure 10. Experimental results, subcommutator data.

of readings of one channel above the base temperature and those of the other channel below the base. This influence can be ignored only if it is of such small magnitude that it does not significantly increase the uncertainty bands.

In this particular case, since the interchannel systematic error was a very small fraction of a degree, no correction was made for its effect. As seen in the figure, all of the reading averages are contained within the calculated 95 percent confidence interval curves. The theory appears to be somewhat pessimistic for small sample sizes, but agrees well with the experimental results for larger values of N.

A similar analysis was made of experimental temperature data read through the system main commutator channels. Figure 11 shows the spread in the averaged values of two data channels on four separate runs. As noted previously, the noise level on the main commutator data is low (range = 3 counts); hence, the 95 percent confidence interval limits are very narrow. Any systematic error that exists therefore becomes a prominent factor. For instance, differences in preamplifier gains, thermocouple electrical outputs, reference junction temperature, etc., that cause a fractional count error in the averaged value, introduce a bias that may be a significant fraction of the predicted confidence interval. To eliminate this bias, the final average of every run on each channel was taken as the base value for the plots in Figure 11. Thus, the uncertainty bands

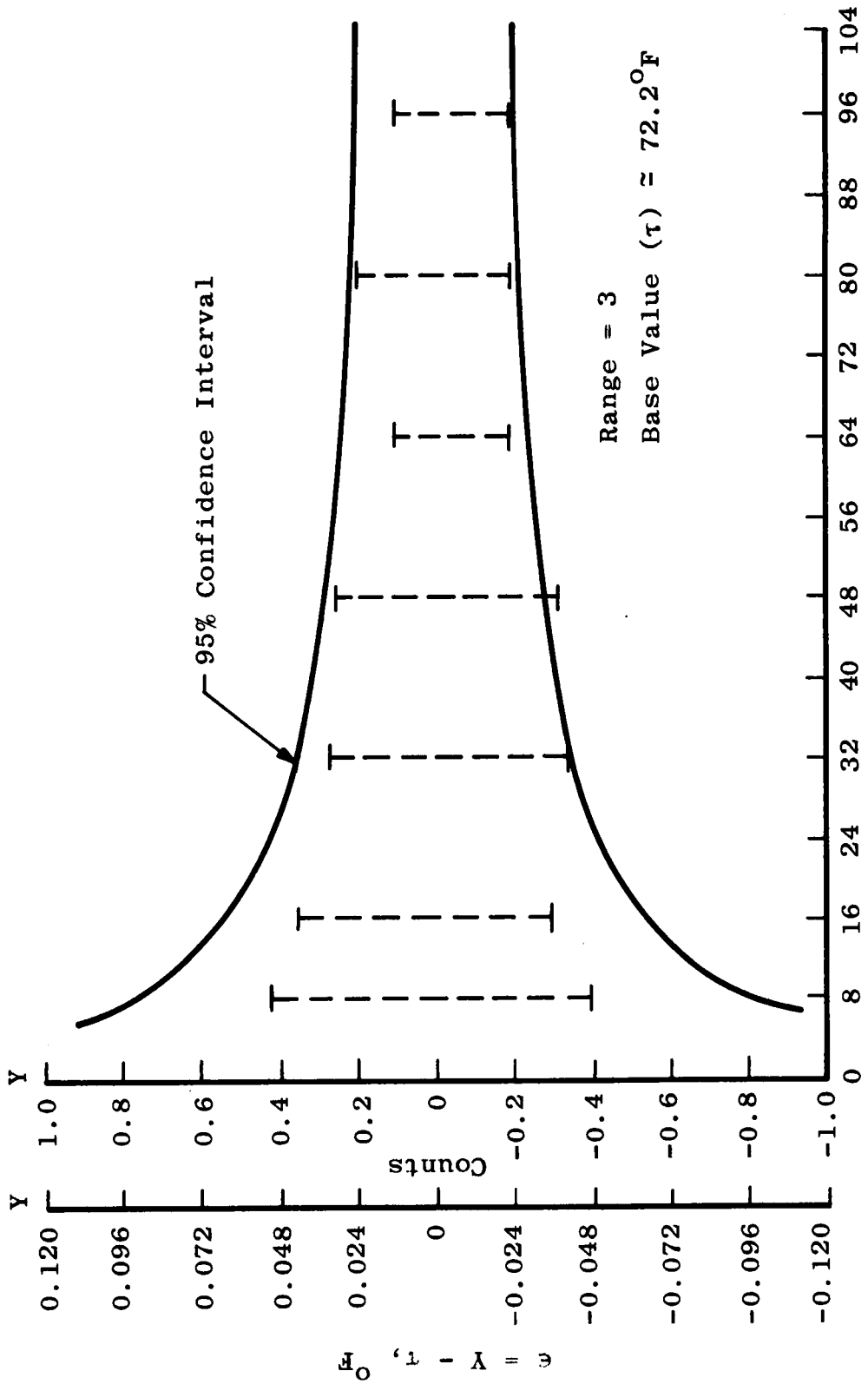


Figure 11. Experimental results, main commutator data.

indicated can reasonably be attributed to the system noise. The improvement in precision observed as the number of values averaged increased, again followed the 95 percent confidence interval predicted from the replicate measurement theory.

It is interesting to note that, by using the averaging technique, it is theoretically possible to achieve a measuring precision approximately five times greater than the system resolution (1 count = 0.12 °F) would supposedly permit. One must realize however, that the substance of this statement may be of more academic than practical value. The systematic errors, neglected in Figure 11, are very real and must be dealt with if absolute values are of interest. Although it is not likely that they can be reduced below the fractional count value indicated above, very good interchannel agreement can be obtained. No effort was made to match thermocouples or preselect amplifiers during these test runs, yet temperature readings observed consistently agreed within 0.1 °F. Certainly, some improvement over this would be possible if a strenuous calibration effort were pursued.

The results obtained in the temperature measuring experiment were found to be typical. Numerous test runs were made, some using thermocouples in different environments, and others with a known millivolt source as an input. Analysis of the resulting data showed no substantive differences from that already presented.

III. ELIMINATION OF OUTLIERS

During the experimentation and data analysis process, the presence of occasional erratic data readings was noted. These were apparently caused by transient noise spikes, momentary subcommutator failure, or other intermittent system malfunctions. In a normal data-taking exercise the data points generated as a result of such erroneous readings would be regarded as bad data and ignored. One such erratic reading (which may be referred to as an outlier) appearing in the average of a group of readings may or may not have an appreciable effect on the final average value. This would depend upon the number of readings being averaged, and upon the magnitude of departure of the outlier from the norm. It would however, always introduce an error that can be detected and eliminated.

The presence of an outlier is identifiable by an increase in the standard deviation of its associated data group. Its effect is to introduce a bias error that detracts from the system precision. Several methods have been developed for separating outliers from legitimate members of a data population. Some of the available outlier tests that may be used include: the 3σ rejection rule, Grubb's method, Anscombe's test, Fisher's t, Thompson's τ, and truncation of largest and smallest values.

An evaluation of the effectiveness of eight of the more prominent outlier tests by using a Monte Carlo simulator

is described in Reference (13). The results indicate that the Thompson's τ technique was far superior to the other methods tested because of its ability to distinguish a larger proportion of outliers closer to the sample value without rejecting good data. Consequently, a routine for subjecting the data system readings to the Thompson's τ test was included in the computer program.

The testing procedure is relatively simple. After the average and standard deviation are computed in the normal manner, an absolute difference is calculated for each reading used in computing the average:

$$\delta_i = |X_i - \bar{X}|$$

where X_i is an individual reading and $\bar{X}$ is the average.

A comparison is then made. If

$$\delta_i > \tau\sigma$$

the data point X_i is considered an outlier, its value is eliminated from the averaging summation, and a new average and standard deviation are computed.

The parameter τ is obtained from a statistical table compiled by Thompson (see Table IV). Its value is a function of the number of readings in the sample (N) and the level of significance (P) desired. The latter term refers to the probability of the test rejecting a good data point. For instance, a level of significance of 0.05 suggests that the odds of rejecting a good point are 100 to 5. Using a

TABLE IV

THOMPSON'S TAU

Sample Size N	Level of Significance P = .1	P = .05	P = .02	P = .01
3	1.3968	1.4099	1.41352	1.414039
4	1.559	1.6080	1.6974	1.7147
5	1.611	1.757	1.869	1.9175
6	1.631	1.814	1.973	2.0509
7	1.640	1.848	2.040	2.142
8	1.644	1.870	2.087	2.207
9	1.647	1.885	2.121	2.256
10	1.648	1.895	2.146	2.294
11	1.648	1.904	2.166	2.324
12	1.649	1.910	2.183	2.348
13	1.649	1.915	2.196	2.368
14	1.649	1.919	2.207	2.385
15	1.649	1.923	2.216	2.399
16	1.649	1.926	2.224	2.411
17	1.649	1.928	2.231	2.422
18	1.649	1.931	2.237	2.432
19	1.649	1.932	2.242	2.440
20	1.649	1.934	2.247	2.447
21	1.649	1.936	2.251	2.454
22	1.649	1.937	2.255	2.460
23	1.649	1.938	2.259	2.465
24	1.649	1.940	2.262	2.470
25	1.649	1.941	2.264	2.475
26	1.648	1.942	2.267	2.479
27	1.648	1.942	2.269	2.483
28	1.648	1.943	2.272	2.487
29	1.648	1.944	2.274	2.490
30	1.648	1.944	2.275	2.493
31	1.648	1.945	2.277	2.495
32	1.648	1.945	2.279	2.498
∞	1.64485	1.95996	2.32634	2.57582

Source: ICRPG Handbook for Estimating the Uncertainty in Measurements Made with Liquid Propellant Rocket Engine Systems. CPIA 180. Silver Springs, Maryland: ICRPG Performance Standardization Working Group, 1969.

smaller level of significance reduces these odds, while using a higher level of significance increases the probability of rejecting bad data points. Since this was our primary aim in applying the test, a higher value (P = 0.10) was used in the program.

As an example of the results which can be achieved, consider the data in Table V. The raw data, taken through one of the system subcommutator channels, contained a number of erroneous readings. This is evidenced by the high standard deviation figures and the excessive count spread between runs. Removal of the outliers reduces these numbers to an acceptable level, and gives a more representative indication of the parameter being measured.

IV. OTHER PROGRAM FEATURES

The computer program used to perform the statistical computation on the experimental data also contains a number of diagnostic routines intended to check the data system operation. For example, running counts of data channels and time information words are accumulated for comparison against the known total record length. This provides a means of detecting errors in channel identifying information, should they exist on the tape. Also, counts are maintained of various flags used by the system to denote specific operational conditions. The occurrence of an improper flag serves as a warning of possible error in the recorded information, and the associated data can then be closely

TABLE V

EFFECTS OF REMOVING OUTLIERS

Run No.	Raw Data (Counts)+		Outliers Removed*	
	Ave.	Std. Dev.	Ave.	Std. Dev.
1	746.69	8.83	748.27	5.11
2	747.96	9.09	749.33	3.56
3	748.32	3.26	748.69	2.64
4	746.79	8.35	748.40	3.87
5	752.18	28.43	748.86	2.67

+Max. Δ_{ave} = 5.49 counts

*Max. Δ_{ave} = 1.06 counts

examined.

Various provisions relating to retrieval of improperly formatted data, preparation of histograms, and data printouts are included to add to the program versatility. Actually, the program was modified many times to include new features or delete portions that were no longer useful. A Fortran listing of the program in its present form is given in Appendix E.

CHAPTER IV

NON-STATIC DATA APPLICATIONS

Thus far, discussion of the replicate measuring technique has been limited to static or quasi-static input data. It has been assumed that variations in the measured parameters are insignificant during the relatively short time period required for a data scan. This is not a particularly discomforting restriction, since the preponderence of environmental testing applications for the data system involve slowly varying parameters. It is not difficult, however, to envision a situation in which the measurement of a large number of non-static data inputs would be desirable. Should this be the case, the replicate measuring technique might also be applicable. As purely an introductory effort in this direction, computer simulation models (see Appendix F) were prepared for a linear time varying signal and for a sinusoid, each modulated with a uniform random noise component.

In the first model, the slope of the straight line was such that the simulated data value increased the equivalent of 20 counts per second. A burst of 80 samples, 4 milliseconds apart, was taken each second. The maximum noise level imposed was ±5 counts. The computer program provided first for the generation of the artificial data points and computation of average values. The midpoint in

time between the first and last data point of a burst was then determined, and the "actual" value of the linear input was computed for that time from the straight line equation. Finally, a comparison was made of this latter value and the average value derived from the 80 point sample. Figure 12 contains a graph of the straight line with the averaged data points plotted at the appropriate time values. An error table has also been included to permit a more accurate evaluation of the results. It can be seen that with the particular slope, noise level, and sampling scheme selected, the deviation from true value is contained within the nominal range of ±1/2 count.

A similar procedure was followed using a sinusoid as an input. An arbitrary frequency of 0.05 Hz was selected. Sampling information and noise level are given in Figure 13 along with the results of the simulation. The error range for this particular experiment was approximately ±1 count.

Obviously, other selections of sampling rates, noise levels, linear slopes and sinusoid frequencies could produce vastly different results; however, this limited effort gives an indication of the potential use of a statistical process for improving time-sampled, non-static data. A further study directed toward the development and testing of confidence intervals as a function of these different variables would be an interesting follow-on to this work.

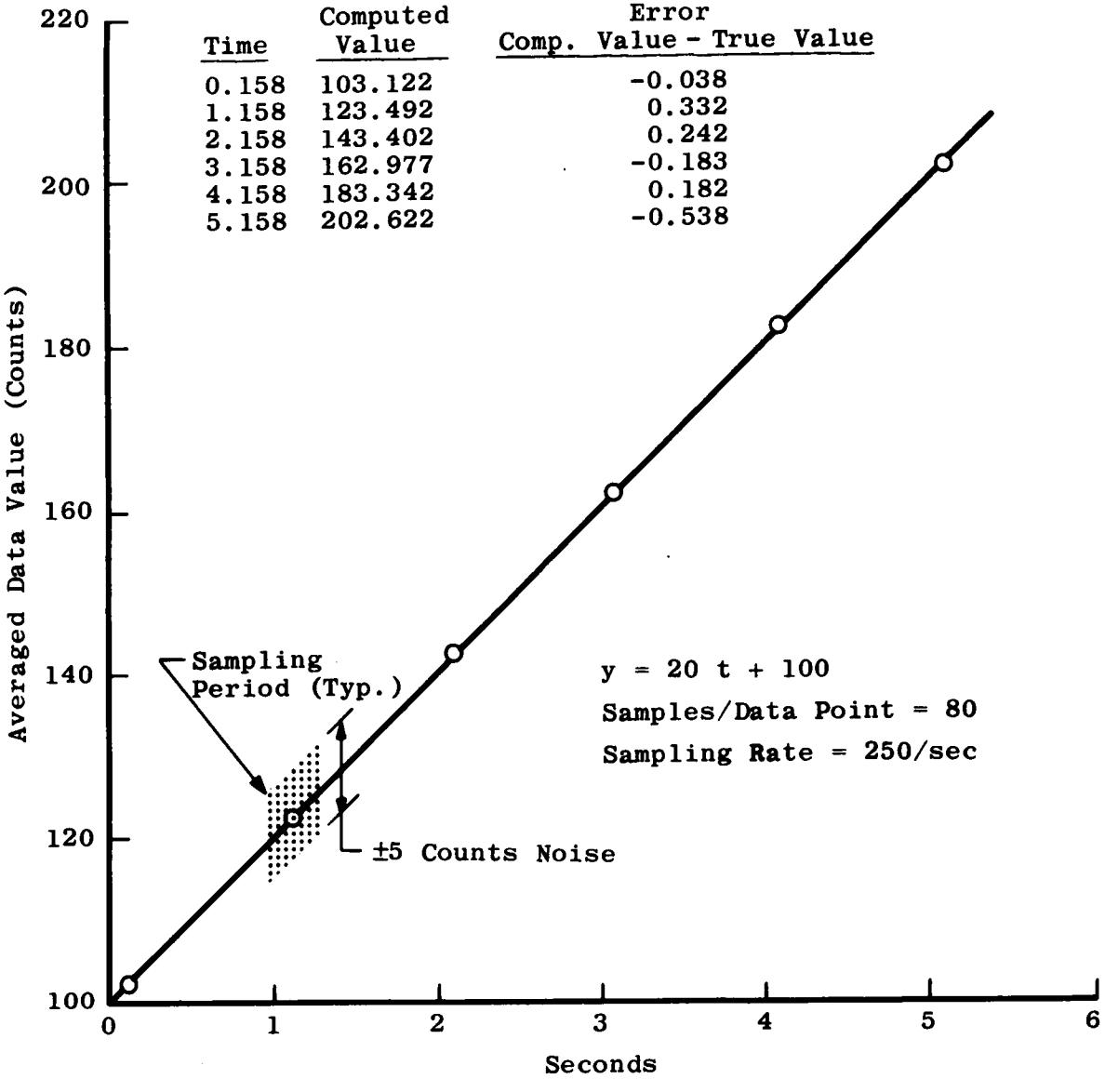


Figure 12. Computer simulation, acquisition of time varying linear data.

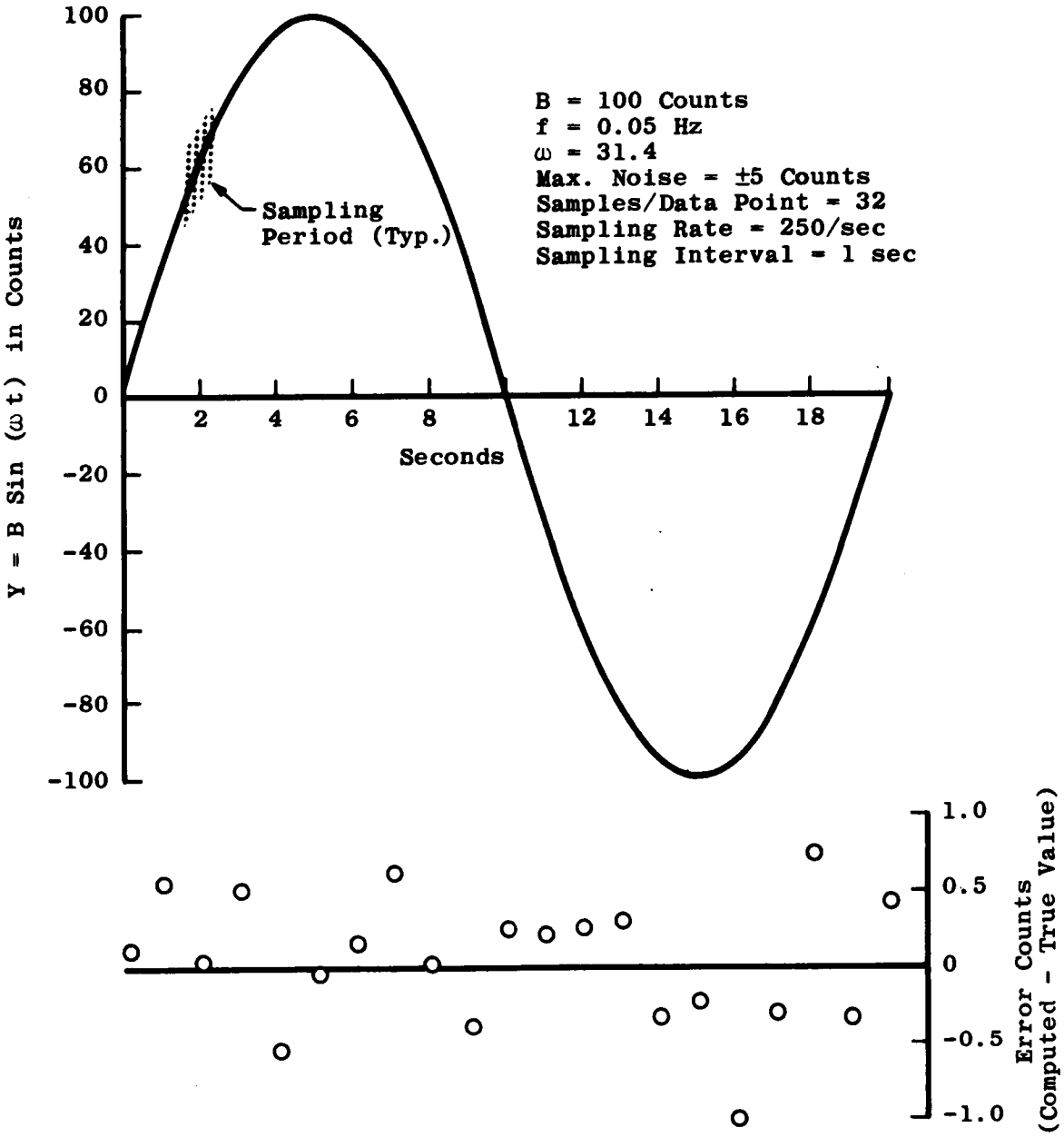


Figure 13. Computer simulation, acquisition of sinusoid data.

CHAPTER V

CONCLUSIONS

The precision of a measuring process can be improved, over that obtainable in a single measurement, by taking the average of a number of replicate measurements. Using well-established statistical theory, confidence intervals can be calculated which predict a factor of two improvement when as few as eight readings are averaged, and nearly order of magnitude improvement as the sample size is increased toward 80. A computer modeling exercise reveals that these statistical predictions are dependable, particularly for the larger sample sizes. The calculated confidence intervals tend to be pessimistic for sample sizes less than 20.

The replicate measurement technique can be used to significant advantage in typical data taking situations involving large scale, digital data acquisition systems. It is possible to achieve a measuring precision up to five times greater than the system resolution, if static biasing errors can be effectively eliminated.

It has been established that the technique is useful for measurement of slowly varying, quasi-static input data. Further effort would be well spent in determining its value in measuring processes involving higher frequency data.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Eisenhart, C. "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems," Journal of Research of NBS, 67C:161-187, April-June, 1963.

2. Eisenhart, C. "Expression of the Uncertainties of Final Results," Science, 160:1201-1204, June, 1968.

3. Hillebrand, W. F., G. E. F. Lundell, H. A. Bright, and J. I. Hoffman. Applied Inorganic Analysis. New York: John Wiley and Sons, Inc., 1953.

4. Denning, W. E. Some Theory of Sampling. New York: John Wiley and Sons, Inc., 1950.

5. Bendat, J. S., and A. G. Piersol. Measurement and Analysis of Random Data. New York: John Wiley and Sons, Inc., 1966.

6. Mandel, J. The Statistical Analysis of Experimental Data. New York: John Wiley and Sons, Inc., 1964.

7. ASTM Manual on Quality Control of Materials. 15C. Philadelphia, Pennsylvania: American Society for Testing Materials, 1951.

8. Duncan, A. J. Quality Control and Industrial Statistics. Homewood, Illinois: Richard D. Irwin, Inc., 1959.

9. Spence, J. T., B. J. Underwood, C. P. Duncan, and J. W. Cotton. Elementary Statistics. New York: Appleton-Century-Crafts, Inc., 1954.

10. Annual Summary Statements of the Accuracy of AMR Tracking Instrumentation. Patrick Air Force Base, Florida: Radio Corporation of America, 1963.

11. Farr, W. R. "Uniform Random Number Generators." NASA Technical Memorandum TMX-64637, Marshall Space Flight Center, Alabama, November, 1971.

12. Klautsch, R. E. "A Universal Data Processing System for the Mark I Aerospace System Environmental Chamber," IEEE Transactions on Aerospace, AS-2: 1152-1160, October, 1964.

13. ICRPG Handbook for Estimating the Uncertainty in Measurements Made with Liquid Propellant Rocket Engine Systems. CPIA 180. Silver Springs, Maryland: ICRPG Performance Standardization Working Group, 1969.

APPENDIXES

APPENDIX A

VARIANCE OF A UNIFORM DISTRIBUTION OF ERROR

An expression for the precision index for a uniform distribution of error may be derived (13) from the general formula definition of the variance:

$$\sigma^2 = \int_a^b (x - \mu)^2 p(x) dx$$

where μ is the mean value and $p(x)$ is the probability density function.

For a uniform distribution, Figure 14, between the limits a and b

$$\mu = \frac{a + b}{2}, \text{ and } p(x) = \frac{1}{b - a}$$

therefore

$$\sigma^2 = \int_a^b \left(x - \frac{a + b}{2}\right)^2 \frac{1}{b - a} dx$$

$$= \int_a^b \left(x^2 - (a + b)x + \frac{a^2 + 2ab + b^2}{4}\right) \frac{1}{(b - a)} dx$$

$$= \frac{1}{b - a}\left(\frac{b^3 - a^3}{3}\right) - \frac{(a + b)}{(b - a)}\left(\frac{b^2 - a^2}{2}\right) + \frac{a^2 + 2ab + b^2}{4(b - a)}[b - a]$$

which reduces to

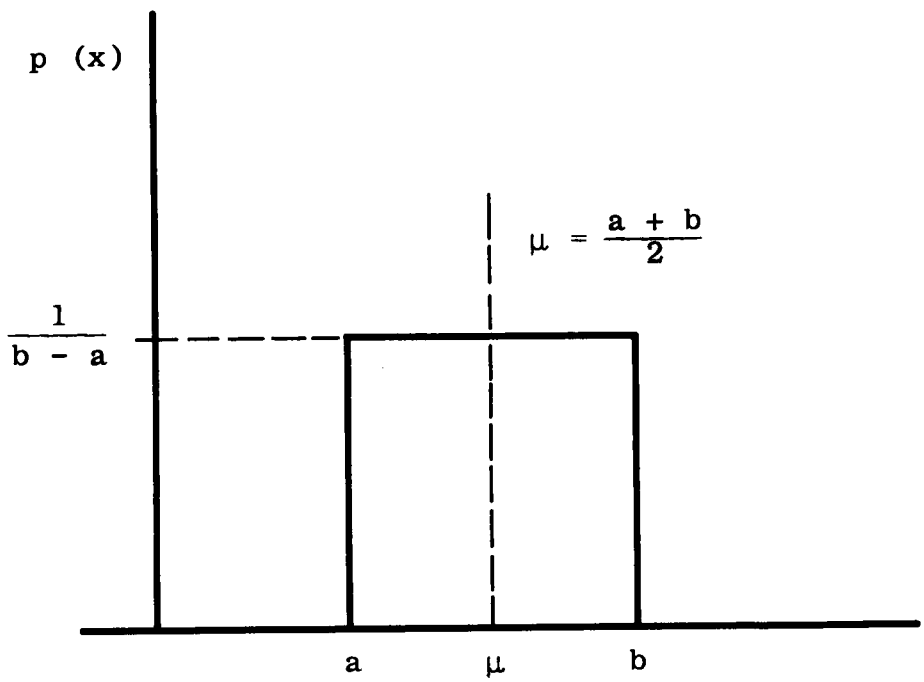


Figure 14. Uniform distribution of error.

$$\sigma^2 = \frac{(b - a)^2}{12}$$

or

$$\sigma = \frac{b - a}{\sqrt{12}}$$

When a = -1 and b = +1

$$\sigma = \frac{1 + 1}{\sqrt{12}} = \frac{1}{\sqrt{3}}$$

APPENDIX B

COMPUTER PROGRAM FOR CALCULATION OF STANDARD DEVIATION

An expression for estimating the variance of a sample from a uniformly distributed population was developed in Chapter II (see Equation 7):

$$S^2 = \sum_{i=1}^{2a/\Delta x} \frac{X_i^2}{2a/\Delta x}$$

In this equation X_i represents values along the X axis separated by Δx; and $2a/\Delta x$ is the number of increments between +a and -a (see Figure 3, page 20). Since X_i is the distance of an X value from the zero axis, it can be seen that $X_i = i\ \Delta x$, therefore

$$S^2 = \frac{\Delta x}{a} \sum_{i=1}^{a/\Delta x} (i\ \Delta x)^2$$

A flow diagram for a computer program written to solve this equation for values of Δx ranging from 2 to 20,000 is given in Figure 15.

The initial steps of the program establish the magnitude and number of Δx increments. The computation is made for a = 1. This is representative of a range of 2. Values for other ranges R may be obtained by multiplying the results by (1/2)R.

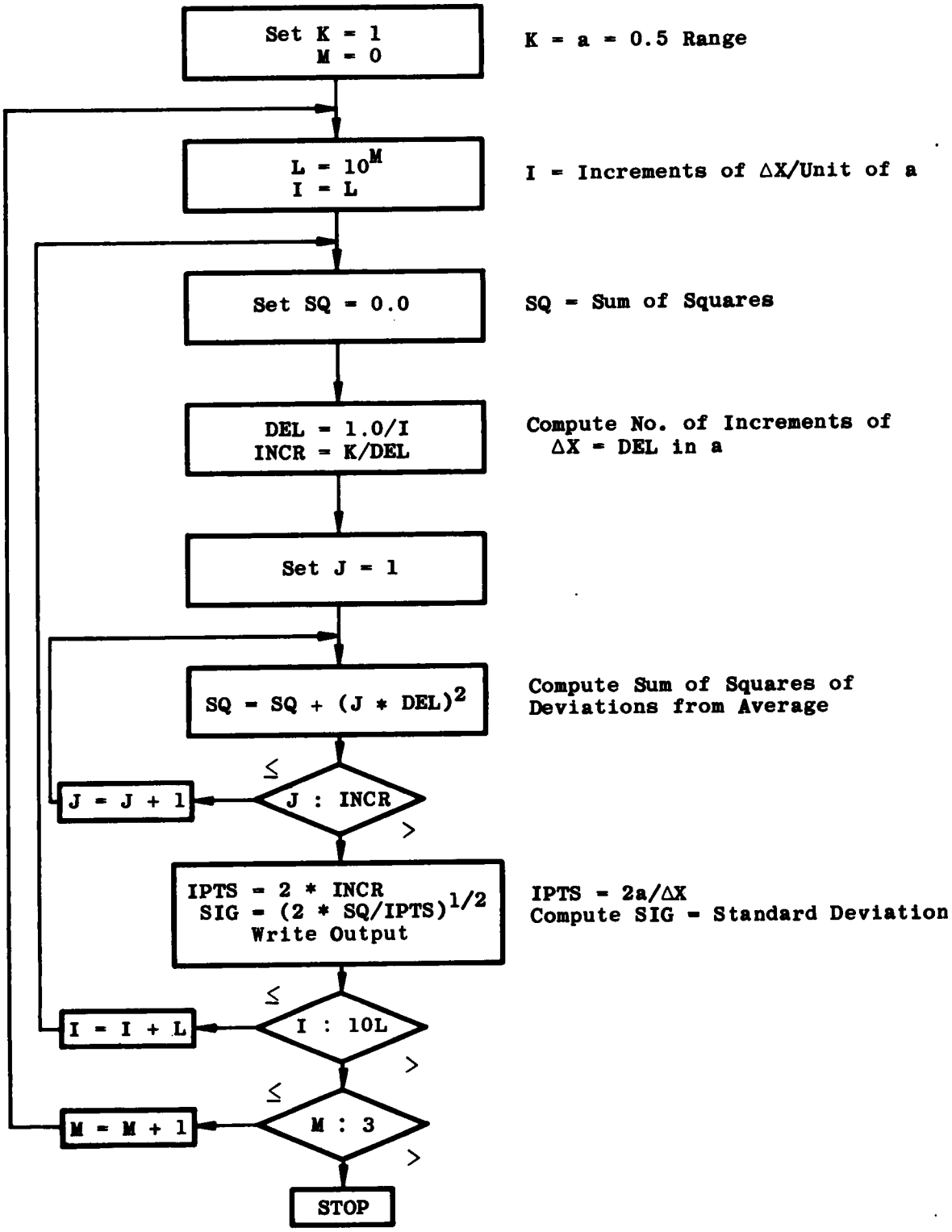


Figure 15. Calculation of standard deviation.

The remainder of the program performs the summation, completes the variance calculation, and extracts the square root to obtain the standard deviation.

A Fortran listing of this program for the Xerox 9300 computer is given on page 65.

```
 SIGMA COMPUTATION
▲FORTRAN LS, G0
     1: C *******  COMPUTATION OF STANDARD DEV.
     2:       K = 1
     3:       DO 150 M = 0,3
     4:       L = 10**M
     5:       DO 150 I = L,10*L,L
     6:       SQ = 0.0
     7:       DEL = 1.0/I
     8:       INCR = K/DEL
     9:       DO 100, J = 1,INCR
    10: 100   SQ = SQ + (J*DEL)**2
    11:       IPTS = 2 * INCR
    12:       SIG = SQRT(2**SQ/IPTS)
    13:       WRITE(108,10) DEL, SIG, IPTS
    14: 150   CONTINUE
    15: 10    FORMAT(1X,$DELTA X =$,F8.6,3X,$SIGMA =$,
    16:      1F8.6,3X,$NO. OF POINTS = $,3X,I6)
    17:       END
```

APPENDIX C

COMPUTER MODEL OF DATA ACQUISITION PROCESS

This is a description of a computer program written to model a data acquisition process and compute averages and standard deviations of increments of the simulated data. The artificial data points are generated by adding scaled random numbers to a base value.

A mixed-multiplicative congruential generator described in Reference (11) is used to produce uniformly distributed random numbers within the interval of 0 to 1. The congruential method is an arithmetic recurrence relation involving integers in which each new number is generated from the previous number by some deterministic approach.

Scaling is accomplished by multiplying the random number by a constant. The noise level (range) on the data can easily be controlled by adjustment of this constant.

A flow diagram of the program is given in Figure 16 with pertinent explanatory comments. Basically, the program generates data points in increments of eight and computes average values and standard deviations. Each succeeding increment is combined with those preceding it and the parameters are recalculated. Limits and increments of the Do Loops are used to control the number of data increments processed and the noise range.

To produce the data for Figure 6, page 27, the

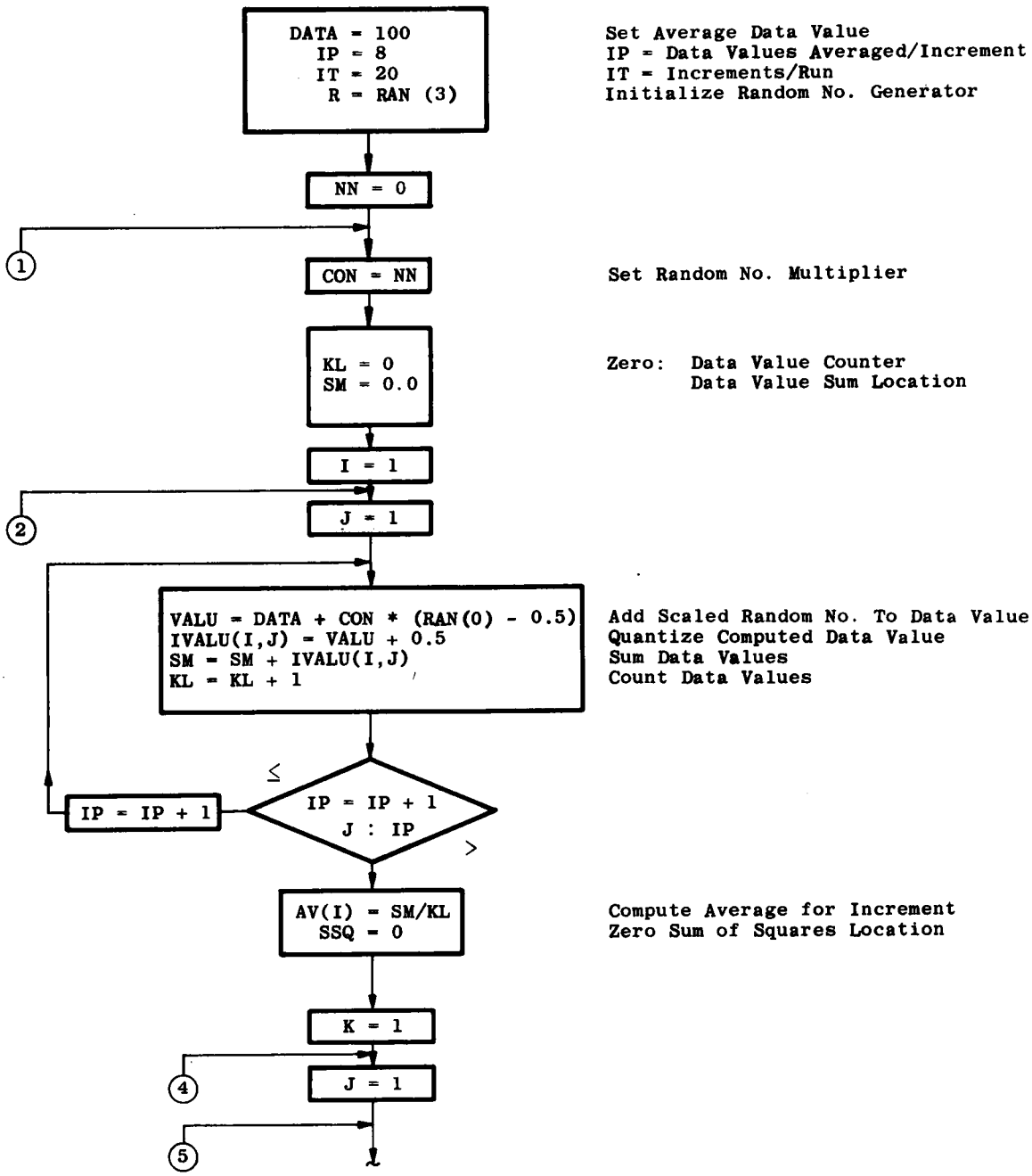


Figure 16. Modeling of data acquisition process.

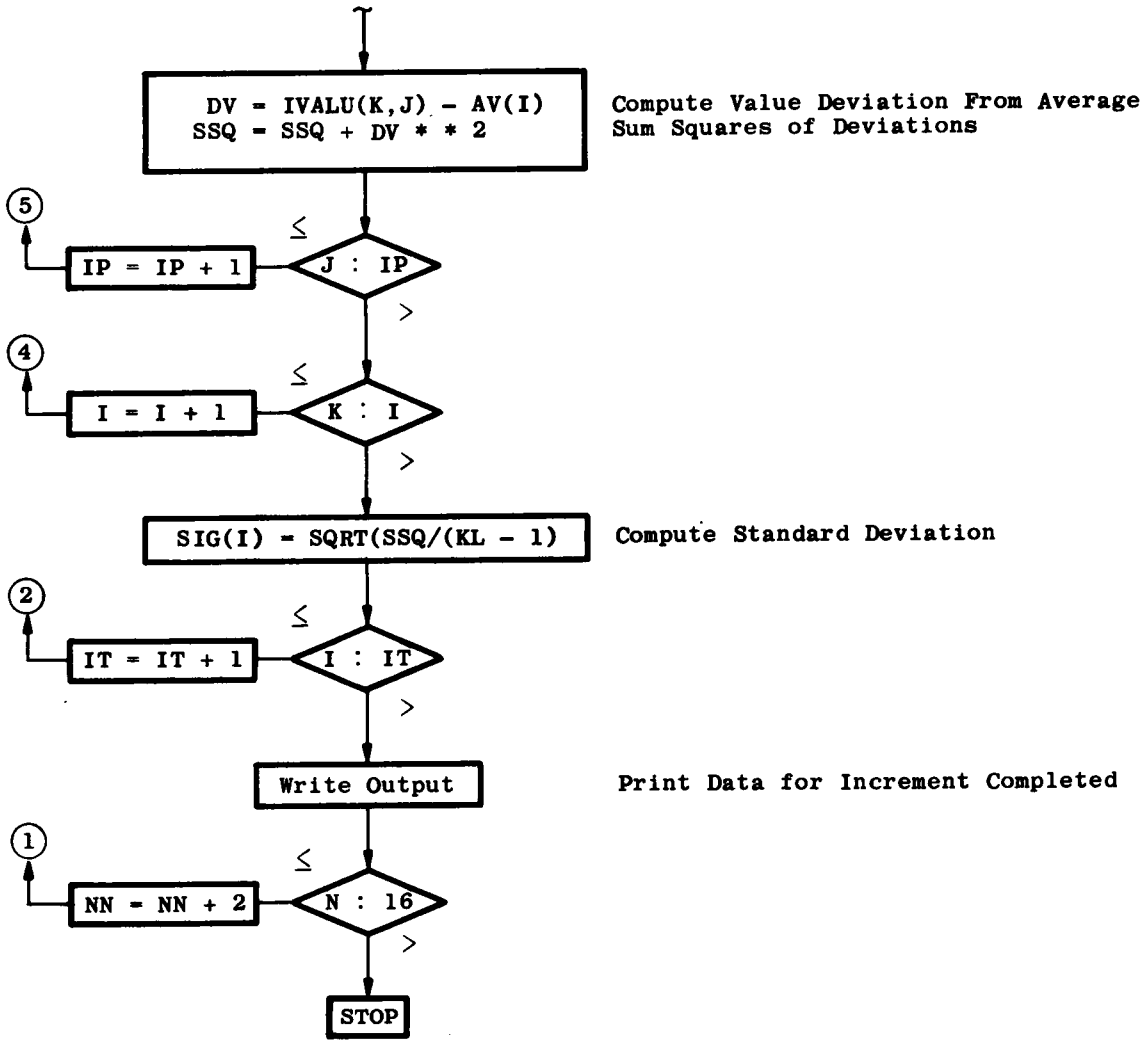


Figure 16. (continued)

program was looped to make repeated runs with the range fixed at 16.

Fortran listings for this program and the random number routine are given on pages 70 and 71. A Fortran listing for the program used to calculate the percentage of 1000 averages of simulated data falling within computed confidence limits are given on page 72. The application of this program is discussed in Chapter II and related to Table III, page 29.

```
DATA STUDY USING RANDOM NOISE
$FORTRAN LS,GO
     1: C *****DATA STUDY USING RANDOM NOISE
     2:       DIMENSION IVALU(20,20),AV(20),SIG(20)
     3:       DATA = 100
     4:       IP = 8
     5:       IT = 20
     6:       R = RAN(3)
     7:       DO 190 NN = 0,16,2
     8:       CON = NN
     9:       KL = 0
    10:       SM = 0.0
    11:       DO 150 I = 1,IT
    12:       DO 110 J = 1,IP
    13:       VALU = DATA + CON*(RAN(0)-0.5)
    14:       IVALU(I,J) = VALU + .5
    15:       SM = SM +IVALU(I,J)
    16:       KL = KL + 1
    17: 110   CONTINUE
    18:       AV(I) = SM/KL
    19:       SSQ = 0.0
    20:       DO 120 K = 1,I
    21:       DO 120 J = 1,IP
    22:       DV = IVALU(K,J) - AV(I)
    23:       SSQ = SSQ + DV**2
    24: 120    CONTINUE
    25:       SIG(I) = SQRT(SSQ/(KL-1))
    26: 150   CONTINUE
    27:       WRITE(108,500) DATA
    28:       DO 180 I = 1,IT
    29:       ICNT = I*IP
    30: 180   WRITE(108,510) ICNT,NN,AV(I),SIG(I)
    31: 190   CONTINUE
    32: 500   FORMAT(//1X,'NO. OF VALUES  PERCENT NOISE  AVERAGE   SIGMA,
    33:      1BASE READING =',I5,//)
    34: 510   FORMAT(6X,I4,12X,F4.1,7X,F7.2,2X,F8.5)
    35: 520   FORMAT(/,1X,(8F8.3))
    36:       END
```

```
NIFORM RANDOM NUMBER ROUTINE
*FORTRAN LS
     1:        FUNCTION RAN(JJ)
     2: C
     3: C      UNIFORM RANDOM NUMBER GENERATOR FOR THE XEROX 9300
     4: C
     5: C      FIRST CALL MUST BE OF THE FORM  X = RAN(J), WHERE J IS AN
     6: C      INITIAL INTEGER VALUE,  SUBSEQUENT CALLS MUST BE OF THE FORM
     7: C      X = RAN(0)
     8: C
     9:        DIMENSION N(128)
    10:        J = JJ
    11:        IF (J) 1,3,1
    12: 1      DO 2 I = 1,128,1
    13:        J = J * 65539
    14: 2      N(I) = J
    15:        L = JJ
    16:        M = JJ
    17:        K = JJ
    18: 3      L = L * 4357
    19:        M = M * 9197
    20:        I = 1 + IABS(L)/65536
    21:        RAN=0.5+FLOAT(N(I)+L+M)*0.59604644E-07
    22:        K = 5 * 10757
    23:        N(I) = K
    24:        RETURN
    25:        END
```

```
  DATA STUDY ON AVERAGE OUTLIERS • UNIFORM DIST•
▲FORTRAN LS,GO
     1: C *****DATA STUDY ON AVE• OUTLIERS • UNIFORM DIST.
     2:       DIMENSION IVALU(25,100)
     3:       DIMENSION AV(25),U(100),RD2(25)
     4:       DIMENSION TB(25), IMANY(25),PCT(25),IMORE(25)
     5:       READ(105,10)(TB(I),I • 1,25)
     6:       DATA • 100
     7:       NN • 1000
     8:       IP • 4
     9:       IT • 25
    10:       RANGE • 16•
    11:       M • 2468357
    12:       DO 95 I • 1,25
    13: 95    IMORE(I) • 0
    14:       DO 100 I • 1,25
    15: 100   IMANY(I) • 0
    16:       DO 160 LL • 1,NN
    17:       R • RAN(M)
    18:       M • M + 2
    19:       KL • 0
    20:       SM • 0.0
    21:       DO 150 I • 1,IT
    22:       DO 110 J • 1,IP
    23:       VALU • DATA + RANGE•(RAN(0)-0.5)
    24:       IVALU(I,J) • VALU + .5
    25:       SM • SM + IVALU(I,J)
    26:       KL • KL + 1
    27: 110    CONTINUE
    28:       AV(I) • SM/KL
    29:       DIFF • AV(I) • DATA
    30:       KK • KL/4
    31:       IF(ABS(DIFF)•LE• 0.9•RANGE•TB(KK)/16) GO TO 150
    32:       IMORE(KK) • IMORE(KK) + 1
    33:       IF(ABS(DIFF)•LE•RANGE•TB(KK)/16) GO TO 150
    34:       IMANY(KK) • IMANY(KK) + 1
    35: 150   CONTINUE
    36: 160   CONTINUE
    37:       IRANGE • RANGE
    38:       WRITE(108,2) NN,IRANGE
    39:       DO 180 I • 1,IT
    40:       PCT(I) • IMANY(I)/NN•100•
    41:       ICNT • I•IP
    42: 180   WRITE(108,4)ICNT,IMANY(I),PCT(I),IMORE(I)
    43: 2     FORMAT(/1X•TOTAL TRIALS PER SAMPLE • •,I5,3X,•RANGE • •,I5//•
    44:      11X•SAMPLE SIZE•,6X•NO• OF OUTLIERS•,6X•PERCENT•,6X,•-9 OUTLIERS•,
    45:      1/)
    46: 4     FORMAT(3X,I4,16X,I4,12X,F7.2,9X,I4)
    47: 10    FORMAT(10F5.3)
    48:        STOP
    49:       END
```

APPENDIX D

COMPUTER MODEL FOR DETERMINING STANDARD DEVIATION OF A UNIFORMLY DISTRIBUTED POPULATION

A simple computer model may be used for determining the approximate limiting value of the standard deviation of data readings uniformly distributed about their average. A flow diagram for a computer program to accomplish this modeling is given in Figure 17. The data values to be averaged are generated by adding a scaled random number (see Appendix C) to a pre-selected base. The average and standard deviation of the total population is then computed. By varying the scaling constant on the random numbers, different noise levels can be determined. The data for Figure 7, page 31, in Chapter II, was obtained by computing the standard deviation for 8,000 data points for noise levels ranging from 2 to 20 counts.

A Fortran listing for this program is given on page 76.

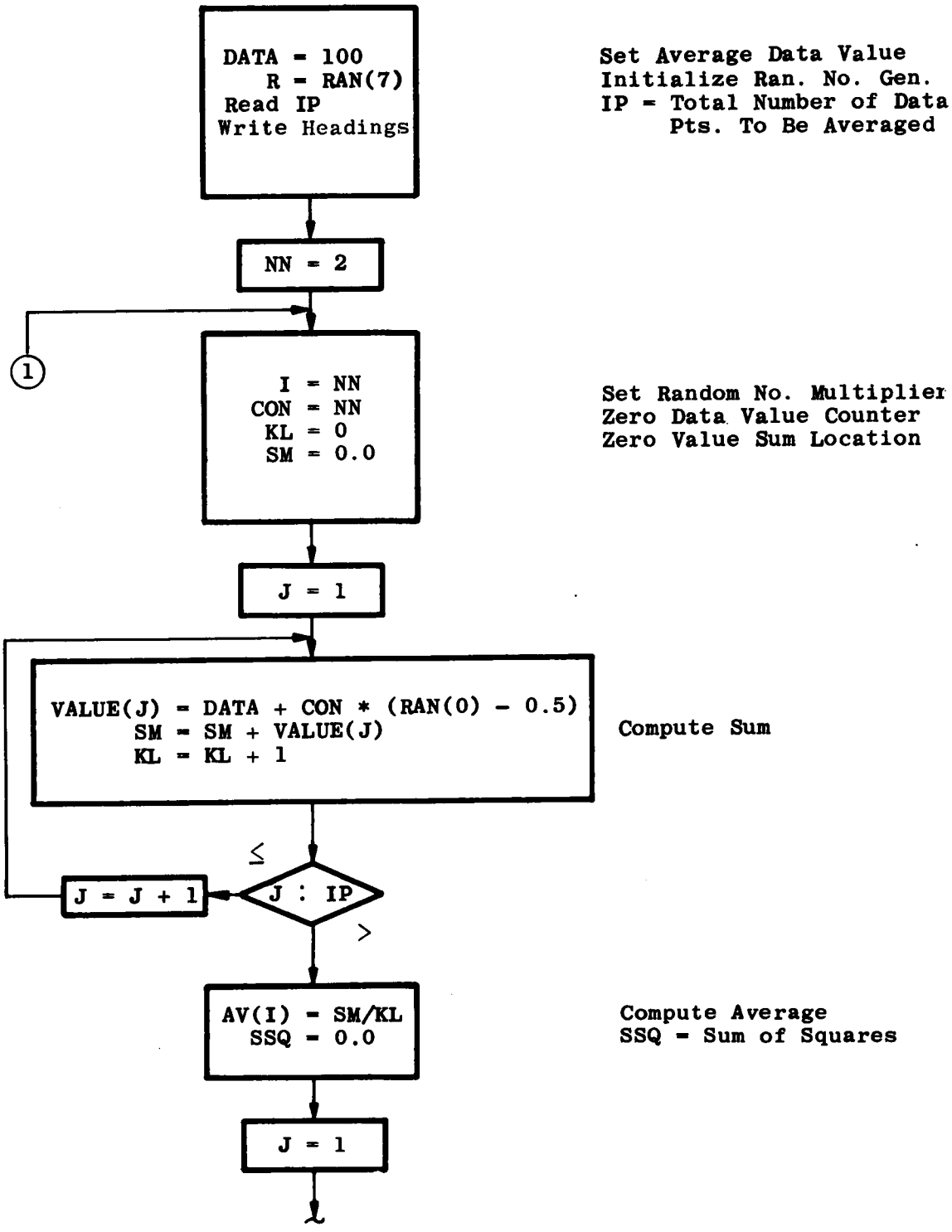


Figure 17. Model for determining standard deviation.

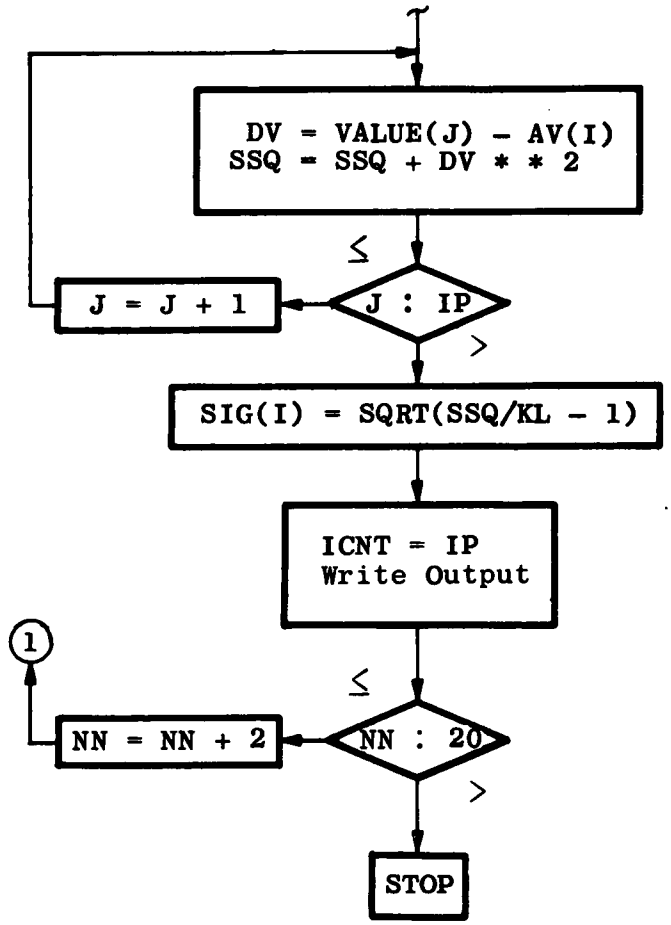


Figure 17. (continued)

```
DATA STUDY USING RANDOM NOISE
▲FORTRAN LS,GO
     1: C *****DATA STUDY USING RANDOM NOISE
     2:       DIMENSION VALUE(8000), AV(20), SIG(20)
     3:       DATA = 100
     4:       R = RAN(7)
     5: 100   CONTINUE
     6:       READ (51,550) IP
     7: 550   FORMAT(I6)
     8:       WRITE(108,500)
     9:       DO 190 NN = 2,20,2
    10:       I = NN
    11:       CON = NN
    12:       KL = 0
    13:       SM = 0.0
    14:       DO 110 J = 1,IP
    15:       VALUE(J) = DATA + CON*(RAN(0) - 0.5)
    16:       SM = SM + VALUE(J)
    17:       KL = KL + 1
    18: 110   CONTINUE
    19:       AV(I) = SM/KL
    20:       SSQ = 0.0
    21:       DO 120 J = 1,IP
    22:       DV = VALUE(J) - AV(I)
    23:       SSQ = SSQ + DV**2
    24: 120    CONTINUE
    25:       SIG(I) = SQRT(SSQ/(KL-1))
    26: 150   CONTINUE
    27:       ICNT = IP
    28: 180   WRITE(108,510) ICNT,NN,AV(I),SIG(I)
    29: 190   CONTINUE
    30:       IF(SENSE SWITCH 1)100,200
    31: 200   CONTINUE
    32: 500   FORMAT(//1X,'NO. OF VALUES  PERCENT NOISE  AVERAGE     SIGMA'/)
    33: 510   FORMAT(/6X,I4,12X,F4.1,7X,F7.2,2X,F8.5)
    34:       END
```

APPENDIX E

COMPUTER PROGRAM FOR DATA ANALYSIS

Data acquired on the VKF data system are recorded on digital magnetic tape in blocks of 375 channels. Each of these blocks, called a record, contains in addition to the data values, channel identifying information, one-second interval time recordings, and various flagging codes that identify conditions associated with the data. The initial record on the tape contains titling information. This is selected manually through a set of digital switches. Typical titling information might include a run number, scale factors, constants, reference values, the date, etc.

The initial portion of the program written to analyze the system data is devoted to retrieval of data records from the tape, sorting and counting flags.

A subroutine called MAGINP is used to call a complete record from the data tape. Then a test is made to determine if the record contained data, tape titles, or a tape code such as end of file or end of record.

As data records are encountered, each word is tested for flag codes and channel identifying information. Time words are decoded and stored for later use in the data printout. The various flag codes are also counted and stored. This information is useful for analyzing the system performance and identifying possible malfunctions or erroneous

data. The data words are sorted by channel identifying numbers, arranged in ascending numerical order for convenience in processing, and restored in the computer memory.

The program includes an option for proceeding with the statistical calculations at this point, or for calling in and storing all records in a data run before beginning the computations. For most of the tests conducted in this work, the scan lengths were selected to accommodate approximately 8 readings of a subcommutator data channel and 16 readings of a main commutator data channel in each record. The statistical calculations could then be made after each record. As each record is read, the data are added to that of the preceding records; hence, a convenient method is available for obtaining data for evaluating the averaging process as the number of values averaged is increased.

All readings of each individual data channel are summed and the average value and standard deviation are computed. The first run on each data tape was usually a zero run. This was made with the individual channel inputs shorted, and was used to identify any zero offset that might exist in each channel. The program tested for a code identifying this zero run and stored the averages obtained from this data as zero values. Each succeeding averaged data point was later corrected by subtracting the zero value associated with that channel number. If the input data originated from thermocouples, the corrected data point was compared to temperature versus millivolts table stored in

the computer memory. A linear interpolation was performed between the one degree values stored, to obtain 0.1 degree resolution.

The final portion of the program, prior to printout, performs a check for outliers. Each data reading is individually subtracted from the average value which was calculated for that channel. The absolute value of this difference is compared with the appropriate Thompson's τ (13) factor multiplied by the standard deviation. If it is larger than the product, the reading is considered an outlier and discarded. A new average value and standard deviation are computed after all data readings have been tested.

The program contains numerous system checking features and running options which are not detailed herein. It was modified a number of times to include or delete certain functions as necessary. A Fortran listing of the program in its present form is given on pages 80 to 86.

```
STATISTICAL EVALUATION OF UDS DATA
$FORTRAN LS, G0.
     1: C         STATISTICAL EVALUATION OF UDS DATA
     2: C       RE KLAUTSCH
     3:         DIMENSION IVALUE(4500),IDENT(100),KK(100)
     4:         DIMENSION CHAVE(100),SIG(100),IPT(1500)
     5:         DIMENSION ICH(375),IDATA(940),IDAY(5),IHOUR(5),
     6:        1IMIN(40),IHIST(43)
     7:         DIMENSION TAU(4),IOUT(10),SIZ(10),CHBIAS(100)
     8:         DIMENSION ZERO(100),V(10),TABLE(130),TEMP(10,20),INCH(10)
     9:         COMMON /RALPH/ TABLE
    10:         READ (105,999) (TABLE(I),I=1,130)
    11: 999     FORMAT(10F5.3)
    12: 50      CONTINUE
    13:         PAUSE 1
    14:         LIMIT = 4499
    15:         TAU(1) = 1.611
    16:         TAU(2) = 1.631
    17:         TAU(3) = 1.640
    18:         TAU(4) = 1.644
    19:         ITPERR = 0
    20: 58      ITTNO = 0
    21:         IREC = 0
    22:         IEOFER = 0
    23: 60      CALL MAGINP(ICH,IDATA,IDAY,IHOUR,IMIN,8,IND)
    24:         GO TO (97,80,70,65,50),IND
    25: 65      ITPERR = ITPERR + 1
    26:         GO TO 60
    27: 70      IF(IEOF.NE.0)GO TO 50
    28:         IEOF = 1
    29:         IF (SENSE SWITCH 1)76,249
    30: 76      IF (SENSE SWITCH 3) 50,58
    31: C         READ TAPE TITLES
    32: 80      IF(IREC.GT.0) GO TO 556
    33: 90      IDATE = ICH(1)
    34:         IRUN = ICH(2)
    35:         MREF = ICH(3)
    36:         ISIGN = ICH(4)
    37:         IF(ISIGN-1)95,93,95
    38: 93      MREF = MREF * (-1)
    39: 95      MM = ICH(5)
    40:         IGAIN = ICH(6)
    41:         IPCMPR = 0
    42:         IPROLS = 0
    43:         ICOMLS = 0
    44:         IALRM = 0
    45:         IOVRFL = 0
    46:         IBYPAS = 0
    47:         IFLGER = 0
    48:         ITIM = 0
    49:         ITITL = 0
    50:         ITTNO = 1
    51:         DO 96 I = 1,MM
    52: 96      KK(I) = 0
    53:         IDATCH = 0
    54:         LLL= 0
```

```
STATISTICAL EVALUATION OF UDS DATA
    55:          KKK = 0
    56:          GO TO 60
    57: 97       IF(ITTNO.NE.1) GO TO 552
    58:          KKD = 0
    59:          KKH = 0
    60:          KKM = 0
    61:          IEOF = 0
    62:          L = 0
    63: 98       DO 120 I = 1,375
    64: C          READ DATA RECORD, TEST FOR DATA CHANNEL, TIME, FLAGS
    65:          K=ICH(I)/10000 + 1
    66:          GO TO(100,101,102,103,104,105,106,120,108,109),K
    67: 100      L = L+1
    68:          ICH(L) = ICH(I)
    69:          IDATA(L) = IDATA(I)
    70:          GO TO 120
    71: 101      IPCMPR = IPCMPR + 1
    72:          GO TO 120
    73: 102      IPROLS = IPROLS + 1
    74:          GO TO 120
    75: 103      ICOMLS = ICOMLS + 1
    76:          GO TO 120
    77: 104      IALRM = IALRM + 1
    78:          GO TO 120
    79: 105      IOVRFL = IOVRFL + 1
    80:          GO TO 120
    81: 106      IBYPAS = IBYPAS + 1
    82:          ICH(I) = ICH(I)-60000
    83:          GO TO 100
    84: 108      ITITL = ITITL + 1
    85:          GO TO 120
    86: 109      ITIM = ITIM + 1
    87:          ISEC1 = ICH(I) - 90000
    88:          ISEC2 = ISEC1/100 + 1
    89:          GO TO (113,110,111,112),ISEC2
    90: 110      KKM = KKM + 1
    91:          GO TO 113
    92: 111      KKH = KKH + 1
    93:          GO TO 110
    94: 112      KKD = KKD + 1
    95:          GO TO 111
    96: 113      ISEC = ISEC1 - (ISEC2 - 1) * 100
    97:          IM = IMIN(KKM)
    98:          IH = IHOUR(KKH)
    99:          ID = IDAY(KKD)
   100:          GO TO 120
   101: 115      IFLGER = IFLGER + 1
   102: 120      CONTINUE
   103:          IDATCH = IDATCH + L
   104:          IOFLAG = IDATCH - IBYPAS
   105:          IF(LLL.GT.LIMIT)GO TO 60
   106: C         ESTABLISH CHANNEL NUMBER IDENTITIES
   107:          NN = 1
   108:          IDENT(1) = ICH(1)
   109:          IF (MM-1) 202,202,140
```

```
STATISTICAL EVALUATION OF UDS DATA
   110:  140    DO 185 I = 2,L
   111:         IF(ICH(I).EQ.0) GO TO 185
   112:         DO 180 J = 1,NN
   113:         IF (ICH(I) - IDENT(J)) 180,185,180
   114:  180    CONTINUE
   115:         NN = NN + 1
   116:         IDENT(NN) = ICH(I)
   117:  185    CONTINUE
   118:  C      ARRANGE CHANNEL NUMBERS IN ASCENDING ORDER
   119:  188    CONTINUE
   120:         DO 195 J = 1,NN
   121:         DO 195 K = J,NN
   122:         IF (IDENT(J).LE.IDENT(K)) GO TO 195
   123:  190    ISAVE = IDENT(J)
   124:         IDENT(J) = IDENT(K)
   125:         IDENT(K) = ISAVE
   126:  195    CONTINUE
   127:  C        ESTABLISH AN ARRAY OF VALUES
   128:  C          IDENTIFIED BY CHANNEL NUMBER
   129:  202    DO 240 I = 1,NN
   130:         KKK = KKK + 1
   131:         IPT(KKK) = LLL
   132:         DO 240 J = 1,L
   133:         IF(ICH(J).NE.IDENT(I)) GO TO 240
   134:  205    KK(I) = KK(I) + 1
   135:         LLL = LLL + 1
   136:         IVALUE(LLL) = IDATA(J)
   137:         IF(LLL.GT.LIMIT)GO TO 242
   138:  240    CONTINUE
   139:  242    CONTINUE
   140:         IPT(KKK+1) = LLL
   141:         KTOT = 0
   142:         DO 248 I = 1,NN
   143:  248    KTOT = KTOT + KK(I)
   144:         IREC = IREC + 1
   145:         IF(SENSE SWITCH 1)249,60
   146:  C        FOR INDIVIDUAL CHANNELS
   147:  C          DETERMINE AVE VALUES AND SIGMA
   148:  249    CONTINUE
   149:         DO 260 I = 1,NN
   150:         IK = I + NN * (IREC - 1)
   151:         SUM = 0.0
   152:         ISUM = 0
   153:         DO 254 K = I,IK,NN
   154:         JJJ = IPT(K+1)- IPT(K)
   155:         DO 253 J = 1,JJJ
   156:         JJ = J + IPT(K)
   157:         ISUM = ISUM + IVALUE(JJ)
   158:  253    CONTINUE
   159:  254    CONTINUE
   160:         SUM = ISUM
   161:         CHAVE(I) = SUM/KK(I)
   162:         CHBIAS(I) = CHAVE(I) - MREF
   163:         SSQDEV = 0.0
   164:         DO 258 K = I,IK,NN
```

```
STATISTICAL EVALUATION OF UDS DATA
    165:        JJJ = IPT(K+1) - IPT(K)
    166:        DO 256 J = 1,JJJ
    167:        JJ = J + IPT(K)
    168:        VALUE = IVALUE(JJ)
    169:        DEV = VALUE - CHAVE(I)
    170:        SSQDEV = SSQDEV + DEV**2
    171: 256    CONTINUE
    172: 258    CONTINUE
    173:        SIG(I) = SQRT(SSQDEV/KK(I))
    174: 260    CONTINUE
    175: 487    CONTINUE
    176: 488    CONTINUE
    177: 489    WRITE(108,14)IDATE,ID,IH,IM,ISEC,IRUN,IGAIN
    178:        NCH = 6
    179:        INCH(1) = 6
    180:        INCH(2) = 7
    181:        INCH(3) = 10
    182:        INCH(4) = 11
    183:        INCH(5) = 12
    184:        INCH(6) = 13
    185:        DO 495 IQ = 1,NCH
    186:        DO 493 I = 1,NN
    187:        IF(IDENT(I).EQ.INCH(IQ))GO TO 490
    188: 493    CONTINUE
    189: 490    IF(IRUN.NE.1)GO TO 492
    190:        ZERO(I) = CHAVE(I)
    191: 492    V(I) = CHAVE(I) - ZERO(I)
    192:        WRITE(108,46)IDENT(I),CHAVE(I),V(I),SIG(I)
    193: 495    CONTINUE
    194:        DO 960 I = 1,NCH
    195:        WRITE(108,4200) INCH(I)
    196:        DO 950  JP = 1,NN
    197:        IF(IDENT(JP).EQ.INCH(I)) GO TO 955
    198: 950    CONTINUE
    199: 955    IW = JP
    200:        IK = IW + NN*(IREC-1)
    201:        DO 960 K = IW,IK,NN
    202:        JJJ = IPT(K+1) - IPT(K)
    203:        J = 1
    204:        JJ = J + IPT(K)
    205:        WRITE(108,4000)(IVALUE(II), II = JJ,(JJ+JJJ-1))
    206: 960    CONTINUE
    207:        IF(SENSE SWITCH 2) 2200,504
    208: C ***  COMPUTE CHAVE AND SIG WITH OUTLIERS ON CHAVE
    209: 2200   CONTINUE
    210:        DIMENSION JL(50), ICHOUT(50), KOUT(50)
    211:        IZ = 0
    212:        DO 2210 III = 1,10
    213: 2210   JL(III) = 0
    214:        DO 2280 I = 1,NN
    215:        IF(SIG(I).LT.1) GO TO 2280
    216:        ISET = 0
    217:        IF(KK(I).LE.4) GO TO 2280
    218:        DO 2202 L = 1,NCH
    219:        IF(IDENT(I).EQ.INCH(L)) GO TO 2204
```

```
STATISTICAL EVALUATION OF UDS DATA
  220: 2202  CONTINUE
  221:       GO TO 2280
  222: 2204  CONTINUE
  223:       DO 2220 IM = 1,50
  224: 2220  KOUT(IM) = 0
  225:       IK = I + NN*(IREC-1)
  226:       DO 2254 K = I,IK,NN
  227:       JJJ = IPT(K+1) - IPT(K)
  228:       DO 2254 J = 1,JJJ
  229:       JJ = J + IPT(K)
  230:       VALUE = IVALUE(JJ)
  231:       DELT = ABS(CHAVE(I) - VALUE)
  232:       IF(KK(I).GT.8.0) GO TO 2250
  233:       IF(DELT.LE.TAU(KK(I) - 4)*SIG(I)) GO TO 2254
  234: 2240  IF(ISET.NE.0)GO TO 2252
  235:       IZ = IZ + 1
  236:       ICHOUT(IZ) = IDENT(I)
  237:       ISET = 1
  238: 2252  JL(IZ) = JL(IZ) + 1
  239:       KOUT(JL(IZ)) = JJ
  240:       GO TO 2254
  241: 2250  IF(DELT.GT.1.648*SIG(I)) GO TO 2240
  242: 2254  CONTINUE
  243:       IF(ISET.EQ. 0) GO TO 2280
  244:       ISUM = 0
  245:       DO 2264 K = I,IK,NN
  246:       JJJ = IPT(K+1) - IPT(K)
  247:       DO 2264 J = 1,JJJ
  248:       JJ = J + IPT(K)
  249:       DO 2260 MMM = 1,JL(IZ)
  250:       IF(JJ.EQ.KOUT(MMM)) GO TO 2262
  251: 2260  CONTINUE
  252:       ISUM = ISUM + IVALUE(JJ)
  253:       GO TO 2264
  254: 2262  KK(I) = KK(I) - 1
  255: 2264  CONTINUE
  256:       SUM = ISUM
  257:       CHAVE(I) = SUM/KK(I)
  258:       CHBIAS(I) = CHAVE(I) - MREF
  259:       SSQDEV = 0.0
  260:       DO 2274 K = I,IK,NN
  261:       JJJ = IPT(K+1) - IPT(K)
  262:       DO 2274 J = 1,JJJ
  263:       JJ = J + IPT(K)
  264:       DO 2270 MMM = 1,JL(IZ)
  265:       IF(JJ.EQ.KOUT(MMM)) GO TO 2274
  266: 2270  CONTINUE
  267:       VALUE = IVALUE(JJ)
  268:       DEV = VALUE - CHAVE(I)
  269:       SSQDEV = SSQDEV + DEV**2
  270: 2274  CONTINUE
  271:       SIG(I) = SQRT(SSQDEV/KK(I))
  272:       KK(I) = KK(I) + JL(IZ)
  273: 2280  CONTINUE
  274:       IF(IRUN.NE.1) GO TO 2316
```

```
STATISTICAL EVALUATION OF UDS DATA
   275:       DO 2310 I = 1,NN
   276: 2310  ZERO(I) = CHAVE(I)
   277: 2316  CONTINUE
   278:       WRITE(108,4500)
   279:       IF(IZ.EQ.0)GO TO 2320
   280:       DO 2368 IV = 1,NCH
   281:       DO 2368 I = 1,NN
   282:       IF(IDENT(I).EQ.INCH(IV))GO TO 2367
   283:       GO TO 2368
   284: 2367  V(I) = CHAVE(I) - ZERO(I)
   285:       WRITE(108,46)IDENT(I),CHAVE(I),V(I),SIG(I)
   286: 2368  CONTINUE
   287:       GO TO 504
   288: 2320  WRITE(108,4600)
   289: 504   IF(IEOFER.EQ.1) GO TO 558
   290:       IF (SENSE SWITCH 1)602,551
   291: 551   IF(SENSE SWITCH 3)50,58
   292: 552   WRITE(108,42)
   293:       PAUSE 2
   294:       ITTNO = 1
   295:       GO TO 60
   296: 556   IEOFER = 1
   297:       WRITE (108,44)
   298:       GO TO 249
   299: 558   IEOFER = 0
   300:       GO TO 90
   301: 600   GO TO 487
   302: 602   CONTINUE
   303: 620   GO TO 60
   304: 14    FORMAT($1TAPE DATE =I6,10X,$TIME=,I6,I4,I4,I4,
   305:      110X,$RUN NO. =$,I5,10X,$AMP. F.S. MV. =$,I5,/)
   306: 18    FORMAT(1X,$TOTAL NO. OF MEASUREMENTS,  BY CALCULATION =$,
   307:      116,6X,$BY COUNT =$,I6,///,1X,$REFERENCE COUNT VALUE =$,
   308:      216,///,
   309:      31X,$CHANNEL NUMBERS READ$,///,(10I10),///)
   310: 19    FORMAT(1X,$COMPUTED AVERAGE VALUE (ALL CHANNELS) =$,F8.2,///,
   311:      51X,$STANDARD DEVIATION =$,F8.2,///,
   312:      61X,$PERCENTAGE OF MEASUREMENTS WITHIN 1 SIGMA =$,F5.1,///,
   313:      71X,$PERCENTAGE OF MEASUREMENTS WITHIN 2 SIGMA =$,F5.1,///,
   314:      81X,$PERCENTAGE OF MEASUREMENTS WITHIN 3 SIGMA =$,F5.1,///,
   315:      91X,$BIAS (AVERAGE - REFERENCE) =$,F8.2,////)
   316: 20    FORMAT(1X,$CHANNEL NUMBER =$,I5,4X,$AVE =$,F8.2,4X,$BIAS =$,F8.2,
   317:      14X,$STAND. DEV. =$,F8.2)
   318: 22    FORMAT(///,$ TAPE FLAG COUNT$,///,
   319:      110X,$0 FLAG, DATA CHANNEL$,7X,I6,/,
   320:      210X,$1 FLAG, PCM PAR. ERROR$,5X,I6,/,
   321:      310X,$2 FLAG, PCM PROV. LOST SYNC$,I6,/,
   322:      410X,$3 FLAG, PCM COM. LOST SYNC$,1X,I6,/,
   323:      510X,$4 FLAG, ALARM$,14X,I6,/,
   324:      610X,$5 FLAG, OVERFLOW$,11X,I6,/,
   325:      710X,$6 FLAG, BYPASS$,13X,I6,/,
   326:      810X,$8 FLAG, TITLE (ERROR)$,6X,I6,/,
   327:      910X,$9 FLAG, TIME$,15X,I6,///,1X,$TAPE READ ERRORS =$,I5,////)
   328: 24    FORMAT($1DISTRIBUTION OF DEVIATIONS FROM REFERENCE VALUE, MREF =$,
   329:      115,///,10X,$COUNTS DEV.$,14X,$NO. OF MEAS.$,/)
```

```
$TATISTICAL EVALUATION OF UDS DATA
   330: 26    FORMAT(10X,$ UNDER -20 $,16X,I4)
   331: 28    FORMAT(17X,I3,17X,I4)
   332: 30    FORMAT(11X,$ OVER +20 $,16X,I4)
   333: 32    FORMAT($1DEVIATIONS FROM REFERENCE COUNT, MREF =$,I5,//)
   334: 34    FORMAT(1X,$CHANNEL NUMBER =$,I5,//,(10I10))
   335: 40    FORMAT($1DATA VALUES ON TAPE$,//,(10I10))
   336: 42    FORMAT($1DATA RECORD NOT PRECEDED BY TAPE TITLES$)
   337: 44    FORMAT($1DATA RECORD NOT CLOSED WITH EOF$)
   338: 46    FORMAT(1X,$CHANNEL NUMBER =$,I5,4X,$AVE =$,F8.2,4X,$VOUT =$,F8.2,
   339:      14X,$STAND. DEV. =$,F8.2)
   340: 1000  FORMAT(1X,$CHANNEL NUMBERS$,(10I10),//)
   341: 1002  FORMAT(/,1X,$TEMP IN DEG F$,3X,(10F10.2),/)
   342: 4000  FORMAT (10I10)
   343: 4200  FORMAT(/,1X,$CHANNEL NO$,I5,/)
   344: 4300  FORMAT(///,1X,$CHANNELS WITH OUTLIERS$,(10I5))
   345: 4400  FORMAT(//,1X,$    NUMBER OF OUTLIERS$,(10I5))
   346: 4500  FORMAT(/,$*** RECALCULATION ELIMINATING OUTLIERS$/)
   347: 4600  FORMAT(1X,$NO OUTLIERS$/)
   348: 800   CONTINUE
   349:       END
```

APPENDIX F

COMPUTER MODELS FOR NON-STATIC DATA

Linear Time Varying Data

The straight line equation $y = mt + b$ was used as the linear time varying input in this modeling exercise. The objective was to prepare a computer program to simulate a data acquisition process that periodically acquires a burst of closely spaced samples of the varying waveform. A noise component is added to each sample by adding a scaled random number from a uniformly distributed population. The scaling constant can be varied to simulate different noise levels.

The average value of each burst of data readings is computed and a time interpolation is made to determine the mid-point of the sample period. The true value of the data point is determined by substituting the mid-point time value into the original straight line equation. An error figure is computed for each data point by subtracting the true value from the computed average.

Pertinent parameters such as sample interval, number of samples/data point, total number of data points, interval between data points, slope of the time varying waveform, and noise level can be easily changed by altering the appropriate constants in the program.

Figure 12, page 52, shows the results of one computer run. The input followed the straight line equation

$y = 20t + 100$. Eight readings taken at the rate of 250 samples per second were averaged for each data point. Maximum noise range was ±5 counts and the interval between data points was one second.

Sinusoid Input

The program described above was modified to accommodata a sinusoid input signal $y = B \sin \omega t$. Values for sinusoid frequency and magnitude are substituted for slope and intercept in the parameter selections. The only other changes necessary are those involving the input equation.

Results of a typical run are given in Figure 13, page 53. Frequency and magnitude of the sinusoid was 0.05 Hz and 100 counts, respectively.

Fortran listings for both of these computer programs are given on pages 89 and 90.

```
DATA STUDY, NONSTATIC WITH RANDOM NUMBERS
*FORTRAN LS.
      C **** VARIABLES   SLOPE, INTERCEPT = B
      C                  SAMPLE INTERVAL = DELT (SECONDS)
      C                  SAMPLES/DATA POINT = JPTS + 1
      C                  NUMBER OF DATA POINTS = ITOT + 1
      C                  INTERVAL BETWEEN DATA POINTS = INT
      C                  MAX COUNTS NOISE = NN = CON
            DIMENSION VALU(0/9,0/99)
            SLOPE = 20.0
            B = 100.0
            DELT = .004
            JPTS = 79
            ITOT = 5
            INT = 1
            R = RAN(5)
            DO 150 NN = 0,10
            CON = NN
            WRITE (108,30) NN
            DO 130 I = 0,ITOT,INT
            SUM = 0.0
            DO 120 J = 0,JPTS
            T = I + J*DELT
            Y = SLOPE*T + B
            VALU(I,J) = Y + CON*(RAN(0)-0.5)
            SUM = SUM + VALU(I,J)
      120   CONTINUE
            AVE = SUM/(JPTS + 1)
            TIME = (JPTS*DELT)/2 + I
            Y = SLOPE*TIME + B
            ERR = AVE - Y
            SSQ = 0.0
            DO 125 K = 0,JPTS
            DEV = VALU(I,K) - AVE
            SSQ = SSQ + DEV**2
      125   CONTINUE
            SIG = SQRT(SSQ/(JPTS + 1))
            WRITE (108,40) TIME, AVE, SIG, ERR
      130   CONTINUE
            DO 140 I = 0,ITOT,INT
            WRITE(108,50) (VALU(I,J), J = 0,JPTS)
      140   CONTINUE
      150   CONTINUE
      30    FORMAT(//,'MAX COUNTS NOISE ','I4,///,
           11X,'  TIME        AVE         SIG         ERR',/)
      40    FORMAT(1X,F8.3,3X,F8.3,3X,F8.3,3X,F8.3)
      50    FORMAT(1X,(8F8.3))
            END
```

```
DATA STUDY, NONSTATIC WITH RANDOM NUMBERS
$FORTRAN LS.
     1: C **** VARIABLES
     2: C                    FREQUENCY = F
     3: C                    AMPLITUDE = B
     4: C                    SAMPLE INTERVAL = DELT (SECONDS)
     5: C                    SAMPLES/DATA POINT = JPTS + 1
     6: C                    NUMBER OF DATA POINTS = ITOT + 1
     7: C                    INTERVAL BETWEEN DATA POINTS = INT
     8: C .                  MAX COUNTS NOISE = NN = CON
     9:       DIMENSION VALU(0/9,0/99)
    10:       F = .05
    11:       W = 6.28 * F
    12:       B = 100.0
    13:       DELT = .004
    14:       JPTS = 31
    15:       ITOT = 20
    16:       INT = 1
    17:       R = RAN(5)
    18:       DO 150 NN = 0,10,2
    19:       CON = NN
    20:       WRITE (108,30) NN
    21:       DO 130 I = 0,ITOT,INT
    22:       SUM = 0.0
    23:       DO 120 J = 0,JPTS
    24:       T = I + J*DELT
    25:       Y = B * SIN(W*T)
    26:       VALU(I,J) = Y + CON*(RAN(0)-0.5)
    27:       SUM = SUM + VALU(I,J)
    28: 120   CONTINUE
    29:       AVE = SUM/(JPTS + 1)
    30:       TIME = (JPTS*DELT)/2 + I
    31:       Y = B * SIN(W*TIME)
    32:       ERR = AVE - Y
    33:       SSQ = 0.0
    34:       DO 125 K = 0,JPTS
    35:       DEV = VALU(I,K) - AVE
    36:       SSQ = SSQ + DEV**2
    37: 125   CONTINUE
    38:       SIG = SQRT(SSQ/(JPTS + 1))
    39:       WRITE (108,40) TIME, AVE, SIG, ERR
    40: 130   CONTINUE
    41:       DO 140 I = 0,ITOT,INT
    42: 140   CONTINUE
    43: 150   CONTINUE
    44: 30    FORMAT(///,'MAX COUNTS NOISE =',I4,///,
    45:      1 1X,'  TIME        AVE          SIG          ERR',/)
    46: 40    FORMAT(1X,F8.3,3X,F8.3,3X,F8.3,3X,F8.3)
    47: 50    FORMAT(1X,(8F8.3))
    48:       END
```

VITA

Ralph E. Klautsch was born on December 9, 1927 at Altamont, Illinois. He attended elementary school in that town and graduated from Altamont Community High School in 1945. He completed one year of college at Valparaiso University in Valparaiso, Indiana prior to being inducted into the United States Navy in March, 1946. After two years of service in the Navy, he continued his schooling at the University of Illinois, and obtained a Bachelor of Science degree in Electrical Engineering from that university in 1951. Immediately following graduation he was recalled to active duty in the United States Navy for 15 months, after which he accepted employment with ARO, Inc., at the Arnold Engineering Development Center, Arnold Air Force Station, Tennessee. He has held positions as instrumentation engineer and section supervisor, and is currently the manager of an aerospace instrumentation branch. During the period from 1966 to 1974, he attended The University of Tennessee Space Institute on a part time basis and graduated in June, 1974 with a Master of Science degree in Electrical Engineering.

He is married to the former Faye Dunn of Altamont, Illinois. They have three daughters, Teresa, Ruth Ann and Andrea. The Klautsch family resides at 205 Fleming, Tullahoma, Tennessee.