MEMORANDUM
RM-4526-PR
APRIL 1965

IMPULSE NOISE AND ERROR PERFORMANCE IN DATA TRANSMISSION

P. Mertz

PREPARED FOR:
UNITED STATES AIR FORCE PROJECT RAND
MEMORANDUM
RM-4526-PR
APRIL 1965

IMPULSE NOISE AND ERROR PERFORMANCE
IN DATA TRANSMISSION
P. Mertz

This research is sponsored by the United States Air Force under Project RAND—Con-
tract No. AF 49(638)-700 monitored by the Directorate of Development Plans, Deputy
Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained
in this Memorandum should not be interpreted as representing the official opinion or
policy of the United States Air Force.

DDC AVAILABILITY NOTICE
Qualified requesters may obtain copies of this report from the Defense Documentation
Center (DDC).

The RAND Corporation
1700 Main St. • Santa Monica • California • 90406
PREFACE

One of RAND's areas of research in communications has been the study of the statistical parameters that can properly describe the various types of noise on facilities designed primarily for telephone transmission and how these noise statistics can be best used to predict digital transmission characteristics over a wide range of transmission and error rates.

This Memorandum is an extension of earlier work of the author and summarizes the methods of approach and conclusions. The study shows that for the telephone-type transmission facilities for which appropriate experimental data are available that the noise and occurrence of errors do not follow the normal statistical laws, and therefore, that much of the conventional engineering of data transmission system performance for such systems has been academic and in some aspects misleading. The results, although not complete or final, even for telephone-type transmission facilities, do indicate simple formulas which are in reasonable agreement with available data. Also, the methods of analysis of the Memorandum should be extendable to other types of data transmission links.
SUMMARY

This report summarizes an extensive study of statistical parameters that can describe the amplitude and occurrence of impulse noise, and the occurrence of errors, on a data transmission facility. The results of the study are remarkably revealing. It has been found that the phenomena do not follow the conventional statistics of purely random events. Instead they can be described by equations of a hyperbola, of the first or higher or lower orders. There is some indication that the distributions change, between those valid for short periods of under a few hours, and those for longer periods of from a few hours up to a number of months. The study has included examination of the probability of "error chains" of extended duration which results from the hyperbolic laws. This turns out to be much more than the insignificant probability indicated by classical laws, and is in general agreement with experimental observations. Precautions are indicated for the measurement of error performance, which measurement can be illusory if carried out in a manner commonly used for quantities following conventional statistical laws. The subject as a whole is developed and described with the view of being most helpful to engineers working in this field, but it is probably adaptable to other fields also.
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREFACE</td>
<td>iii</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiii</td>
</tr>
</tbody>
</table>

Section

1. INTRODUCTION                                                        | 1    |
2. IMPULSE NOISE                                                        | 5    |
3. GEOMETRIC, BINOMIAL AND POISSON DISTRIBUTIONS                        | 16   |
4. HYPERBOLIC DISTRIBUTION                                              | 38   |
5. PARETO DISTRIBUTIONS                                                | 63   |
6. ERROR BURSTS AND CHAINS                                             | 91   |
7. MEASUREMENT OF CIRCUIT PERFORMANCE                                  | 112  |
8. CONCLUSIONS                                                         | 134  |

Appendix

- SPECIAL COORDINATE PAPERS                                             | 140  |
- REFERENCES                                                            | 150  |

Addendum

- ERROR PERFORMANCE DATA REPORTED BY BELL SYSTEM                       | 154  |
LIST OF FIGURES

1. Cumulated probability of amplitudes in Gaussian, Rayleigh, and hyperbolic distribution noise 11

2. Long-haul AT&T toll circuits (600 bits/sec) 13

3. Number of errors independent of level, as level is changed, and slope of variation for errors varying with level 15

4. Errors occurring in continuous time scale 20

5. Errors occurring in discrete time scale of bits 20

6. Event-free intervals in illustrative example 22

7. Probability of exactly c events in interval having long-time average a (Poisson) 23

8. Probability of at least c events in interval having long-time average a (Poisson) 24

9. Expected maximum and minimum error-free intervals in continuous and discrete distributions 28

10. Expected maximum error-free intervals in Gilbert Model, compared with case of errors occurring singly 32

11. Expected maximum and minimum burst-free intervals in Gilbert Model, compared with error-free intervals, and also with case of errors occurring singly 35

12. Relation between bias $h$ and long-time average $a$ of events in hyperbolic distribution 41

13. Probability of exactly $c$ events in interval having long-time average $a$ ($k = 1000$) 42

14. Probability of at least $c$ events in interval having long-time average $a$ (hyperbolic, $k = 1000$; and Poisson) 44

15. Relation between bias $h_0$ and long-time average $a$ of events in hyperbolic distribution of event-free stretches 49

16. Solution of implicit equation in $N(U)$ 50
17. Expected maximum and minimum error-free intervals in hyperbolic and Poisson distributions .......................... 54
18. Distribution of word errors per minute in long-haul private-line data circuit (Lincoln Laboratory data, 1300 bits/sec) comparison with hyperbolic and Poisson formulas ................................................. 55
19. Experimental error distributions (Lincoln Laboratory) .. 57
20. Experimental error distributions (Ialacable) .............. 58
21. Experimental error distribution (Philips) .................. 58
22. Cumulative probability of error-free stretches in switched telephone circuits used for data transmission ...... 59
23. Logarithmic plot of inter-error spacings u .................. 64
24. Linear plot of inter-error spacings u ....................... 65
25. Variation of slope with total number of errors N ........ 66
26. Incomplete zeta function S(N,s) ............................... 74
27. Total test time T in bits ......................................... 75
28. Slope versus test time and number of errors ................ 75
29. Maximum expected inter-error interval versus test time and number of errors .................................................. 77
30. Short-time error density probabilities versus average error density .......................................................... 79
31. Error records, German IBM tests ............................. 83
32. Inter-error distributions compared with theoretical curves ............................................................... 84
33. Slopes versus average error densities (experiment versus theoretical computations) ........................................ 85
34. Probability of an error versus bit position following initial error in burst ................................................. 93
35. Model multiform distribution of error burst durations .... 93
36. Model and experimental distributions of error burst durations .............................................................. 94
37. Density of errors within gate intervals .......................... 96
38. Experimental distribution and Gilbert approximation
   in error burst structure ............................................. 99
39. Summary Dimock results and Gilbert approximation ......... 102
40. Potential error chains and error-free gaps .................... 105
41. Assumed critical gaps for error chains of
   varying lengths ..................................................... 108
42. Expected numbers of error chains in illustrative tests ... 109
43. Expectation of ratios of maximum/average and
   maximum/median as functions of test duration
   (Poisson and simple hyperbolic distributions) ............... 114
44. Experimental ratios of maximum/average error incidence
   as functions of test duration ......................................... 116
45. Cumulated averages of error incidence ......................... 117
46. Cumulated medians of error incidence ......................... 119
47. Error distributions in long and segmented tests .......... 123
48. Experimental distributions in segmented test ............... 125
49. Experimental distributions in segmented test ............... 126
50. Error distributions in long test with original
    and merged data ..................................................... 130
51. Experimental effect of merging data ........................... 131
52. Experimental effect of merging data ........................... 132
53. Biased logarithmic scale to include zero in plot ........... 142
54. Biased logarithmic scale to expand grid just above 1 .... 142
55. Probability scale based on log scale .......................... 142

Addendum

A. Bit error rate distributions for: All calls, exchange,
   long-haul, and short-haul. After Fig.2 of Townsend
   and Watts ............................................................. 156
B. Distribution of error bit rates per one- and five-minute
   interval. After Fig.16 of Townsend and Watts ............ 157
C. Distribution of error spacings. After Fig. 7 of Townsend and Watts. Comparison with Mandelbrot model ............ 158

D. Error rate distributions measured in blocks of M bits. After Fig. 14 of Townsend and Watts. Comparison with computation from extension of Mandelbrot model .......... 161
LIST OF TABLES

1. Parameters of Bursts .................................................... 36
2. Error-Free Intervals versus Burst Rate ............................ 37
3. Parameter A(k) for Finite Averages ................................. 40
4. Parametric Function N(U) .................................................. 48
5. Incomplete Zeta Function S(N,s) ................................. 70
6. Parameters in 1-Minute Test Interval ............................ 80
7. Parameters of Illustrative Gilbert Model ......................... 87
8. Illustrative Chain Lengths T ............................................ 107
9. Number of Chains k'P(a,c) in Illustrative Tests ............... 110
10. Principal Grid Lines--Hyperbolic Paper ....................... 147
I. INTRODUCTION

A number of years ago it was recognized that the phenomena attending the introduction of errors in data transmission systems did not follow the established laws of purely random events. It was then thought that these phenomena did not follow any systematic laws at all. Further, it was thought that even if some type of law could be set up for one kind of transmission facility, the situation would be completely different for another kind of facility and hence would frustrate a general engineering philosophy. Thus, the established laws based on complete randomness tended to be used as a makeshift; and this tendency has continued because of mathematical simplicity, even where results appear unrealistic.

Over the last few years, these phenomena have been the subject of a study for The RAND Corporation; this Memorandum summarizes the findings of the study.

Noise is obviously the first item involved. The noise on the transmission facilities used generally seems to be made up of a mixture of Gaussian and impulsive noise. The latter tends to have occasional short periods of very high intensity and tends, therefore, to be more critical than the Gaussian noise in causing errors in the transmission. Very short duration line opens or "hits" also contribute to errors.

The precise causes of noise impulses and hits would form an interesting study, particularly if it led to suggestions for remedial measures. The causes, however, are found to be quite elusive, varied, and erratic. For the present a more constructive study describes the effects quantitatively and as exactly as possible, so that one
circuit may be compared with another in performance. A description in such cases is nearly always empirical, and this is true here. In spite of the erratic character and variety of the causes, and the wide range of the types of facilities, there is considerable statistical conformity in the effects. The variation from one case to another is generally expressed in terms of appropriate magnitude parameters, rather than in radical differences in the statistical laws.

The error occurrences, as a function of time, can be described in several ways. Campbell's description gives a distribution of the error rates, which are above and below the long-time average, for a small time interval. A second description gives the distribution of inter-error spaces. Both of these have been found commonly to follow new laws.

The results of the study lend remarkable insight into the time structure of the actual error occurrences in the experimental data system. They are presented in a form that will be directly applicable to the engineering problems of such systems, but they are also believed applicable to other fields that have eluded conventional statistical treatments.

In brief outline, the present summary proceeds as follows:

First, the amplitude and time distributions of the impulse noise are covered. Then a brief analysis is made of the evidence showing the extent to which short line opens or "hits" share with impulse noise the responsibility for data errors in transmission.

The various statistical formulations used, namely the Poisson distribution, the simple hyperbolic distribution, and the Pareto
distribution, are reviewed or developed. It is indicated that there is an apparent change in law of distribution between short-duration and long-duration tests. The maximum expected error-free gaps during short tests are studied for the respective distributions. These are then applied to the estimation of the probability of long error chains.

A study is finally presented of the philosophy of measurement of the average overall error: performance of a data transmission system, with a recognition that this average does not obey the laws of large numbers.

A brief analysis is then given of a variety of special graph scales that have been found useful in the study.

The results of the study give general information on the time-grouping of the noise peaks and the resulting data errors under a wide variety of conditions, and also on the numerical parameters that are needed to express this quantitatively, so that one can make engineering plans. Also, the departures from classical fully random events can be appreciated. These results permit far more precision of language than the usual statement that in experience the errors are found to cluster or "bunch" more than in a purely random distribution.

Nevertheless, the study will undoubtedly disappoint many mathematicians who would like to see all the facts explained and unified by a single formula. The facts simply refuse to fall into such a simple form. Several elegant mathematical models of error occurrence have been proposed and are described. But none of them holds over the wide range of test durations which it is realistically necessary to consider. The treatment is aimed, therefore, in the
direction of being most useful to the engineer, and most in con-
formance with reality. This is solved by recommending one model
for the shorter tests and another for the much longer ones.
II. IMPULSE NOISE

In the engineering of communications systems, Gaussian white noise has usually been assumed to be the one principal source of interference. In the last few years of experience with data transmission, however, there has been a gradual realization that, on many occasions, this approach is academic. Noise in fact tends to group itself into three categories. (1) Two of these are single-frequency and impulsive noise, concentrated respectively in frequency and time. Between these extremes, and capable of being formed from either if a sufficiently large number of random samples cumulate together, lies the category of Gaussian white noise.

The impulsive noise is more difficult than the other two forms to express simply in a quantitative form. This, of course, is due to its more erratic character. Experience with exact measurement of impulsive noise has been meager but does reveal a few characteristics common to all, or nearly all, the tests.

The nature of noise is being fairly continuously investigated. A recent fairly sophisticated study is that by Brown. (2) Other studies deal specifically with atmospheric noise in radio. One of these is by Hoff and Johnson, (3) who found that the atmospheric noise tends, for the lower probabilities, to much higher amplitudes than is the case for Rayleigh (or rectified Gaussian) noise. Their work at the National Bureau of Standards led to a very comprehensive and sophisticated analysis of atmospheric noise. (4-7) In particular, this latter analysis finds that the law changes from short-period distributions (a few minutes to an hour) to longer-period distributions
Further studies of atmospheric noise are by Horner and Harwood, (8) and Furutsu and Ishida. (9) A more general discussion is given by Fennick. (10)

The model outlined here was intended to be simple and readily usable by engineers. The aim was to introduce as few new parameters as possible, consistent with keeping the model reasonably realistic. This model was originally proposed independently of Refs. 2-8 and 9-10, but turns out to be very similar to some of those proposals. (11,12)

It is expected that by inserting appropriate values for the various parameters, the model can typify impulsive noise well enough to be used for general engineering purposes. As a part of this objective, the very values substituted numerically in any given situation serve as a quantitative description of the more significant characteristics of the impulsive noise experienced in that situation. It is not likely, however, that a model with this simplicity could replace the more detailed model described in Ref. 7 and used by the National Bureau of Standards for very specific reporting and engineering of radio noise conditions and radio circuits, respectively.

The present model involves specifications in two parameters. The first relates to the distribution of the impulses in amplitude, and the second to their distribution in time.

The distribution proposed for the amplitudes consists simply of using an empirical hyperbolic law. This is a variant of the Cauchy and Pareto distributions, and also of "Zipf's law." (13-16) The amplitude indication is not derived as simply from the rms as is the
Gaussian distribution. Rather, it is obtained from longer duration measurements on high amplitudes reached with infrequent probabilities of occurrence. These amplitudes are, incidentally, substantially larger than the Gaussian amplitudes having the same low frequencies of occurrence.

The specification of the distribution of impulses in time is more involved. It has been translated into a time distribution of errors. This is principally because experimental data on this distribution are at best quite meager, and in most cases have been measured on errors rather than on crossings of a given amplitude threshold. It is presumed that with any given data transmission system there is a certain degree of correlation between such a crossing of a threshold and the generation of an error. The amplitude value of the threshold, and the percentage of crossings that lead to errors, will, however, vary from system to system. This is not of immediate interest here. For the purpose of discussing a general impulse noise model, the description of the time distribution in terms of errors appears more convenient for comparison with such data as are available.

The matter of this time distribution of errors has been developed in some detail in the following parts of this Memorandum, and will not be further touched on here. A note of caution will, however, appear below against assuming too literally, in a critical case, that the time distribution of noise peaks can be determined from experience with errors only.
DISTRIBUTION OF AMPLITUDES

Experimentally, it has been found that, in at least many important instances and possibly even more generally, the distribution of amplitudes in impulsive noise is approximately hyperbolic. Reference will cover a specific case, but it agrees with references already cited and with other cases that have come to the author's attention. The distribution in nearly all these instances approximately follows a higher-order hyperbolic law.

To keep the frequency of occurrence of the smallest amplitudes finite, a small constant has to be added to the abscissa. Thus, the typical expression is

\[ p(x) = \frac{k}{(x + h)^m} \quad (1) \]

where \( p(x) \) = probability density of amplitude \( x \)

\( x \) = amplitude

\( k, h, m \) = constants

The cumulated probability \( P(x) \) of \( x \) or greater amplitude is*

\[ P(x) = k \int_x^\infty \frac{1}{(v + h)^m} \, dv \quad (2) \]

\[ = \frac{k}{m} (m - 1) (x + h)^{m-1} \quad (3) \]

The normalizing constant \( k \) must give 1 when the cumulation runs from 0 to \( \infty \); thus

*Note that \( P(x) \) is not the distribution function corresponding to \( p(x) \). See Ref. 18, pp. 168, 200.
\[ l = k/(m - 1)h^{m-1}, \quad (4) \]

or

\[ k = (m - 1)h^{m-1} \]

When dealing with the integrated probability, it is convenient to set \( n = m - 1 \). With \( k \) substituted in Eqs. (1) and (3), these become, respectively,

\[ p(x) = (m - 1)h^{m-1}/(x + h)^m \quad (5) \]
\[ P(x) = h^n/(x + h)^n \quad (6) \]

In Refs. 8, 9 and 11 the hyperbolic formula corresponding to Eq. (6) is given in the form

\[ P(x) = c/(x^n + c) \quad (7) \]

This differs from Eq. (6) only in the regions where the noise amplitude \( x \) is small, which are not the regions of particular interest for engineering purposes.

It is now of interest to compute the rms value of \( x \) in the distribution of Eq. (6). Let \( x^2 \) be the mean square. Hence,

\[ \overline{x^2} = \int_0^\infty x^2 p(x) \, dx \]

\[ = (m - 1)h^{m-1} \int_0^\infty \frac{x^2}{(x + h)^m} \, dx \quad (8) \]
Let
\[ v = x + h \]
\[ x^2 = (m - 1) h^{m-1} \int_0^\infty \frac{(v-h)^2}{v^{m-1}} \, dv \]  
\[ = (m - 1) \left[ \frac{1}{m-3} - \frac{2}{m-2} + \frac{1}{m-1} \right] h^2 \]
\[ = 2h^2 / (m-2)(m-3) \]
\[ = 2h^2 / (n-1)(n-2) \]  
\[ (10) \]

The rms value then is
\[ \text{rms}(x) = h \sqrt{2 / (n-1)(n-2)} \]  
\[ (11) \]

It is of interest to plot some distributions. In actual experience it is found that \( n \) runs in the order of a little over 2 to something like 5. It is clear from Eq. (11) that a distribution with \( n = 2 \) does not yield a finite rms, and that a little more than 2 is necessary for this.

It is also clear that the constant \( h \), which was introduced to keep the probability density finite for near-zero \( x \), achieves importance as a scale factor. It does not affect the distribution shape at large amplitudes, nor the amplitude ratio to the rms when \( n \) is fixed and known, but it does influence the distribution for small amplitudes.

A plot of the cumulative distributions for \( n = 3, 4, \text{ and } 5 \) is shown in Fig. 1. Here \( x \) is plotted as an amplitude ratio, in \( \text{dB} \), to the rms value. For these, \( h \) has been arbitrarily set equal to 1 in Eqs. (6) and (11). For comparison, Gaussian and Rayleigh...
Fig. 1 — Cumulated probability of amplitudes in Gaussian, Rayleigh, and hyperbolic distribution noise.

distributions are also plotted.

At low frequencies of occurrence, the hyperbolic distributions show much larger amplitudes than do the Gaussian or Rayleigh distributions.

With a Gaussian distribution, a measurement of amplitude can be obtained in a relatively short time by measuring the rms. With the hyperbolic distribution, the rms is meaningless unless other parameters \( n \) and \( h \) are known with some accuracy. For a reliable figure it is necessary to take a fairly long-time measurement and observe the high amplitudes actually reached at low frequencies of occurrence.
The point is also found to hold for the time distributions of errors, and the matter is discussed again in Section VII.

**MULTIPLICATIVE NOISE**

The correlation touched upon above, between impulsive noise peaks and errors, is impaired by another form of disturbance that causes errors.\(^{(17)}\) Such disturbance has been called "hits" or "dropouts" to indicate circuit interruptions that might be shorter than one bit in duration, but that also might be longer, and might even extend up to minutes.

The dropout is, of course, a special case of a more general circuit disturbance, which has occasionally been called "multiplicative noise." This is to distinguish it from the more usual "additive" noise considered above, and which merely adds to the signal wave. The multiplicative noise multiplies the instantaneous value of the signal wave amplitude by an erratic function. It has also been called "modulation noise." In the simple case of dropouts the multiplying function is either 1 or 0; most of the time 1, but occasionally and for short intervals of time, 0. Generally, the multiplying function varies irregularly over a range of amplitude values.

In the case of additive noise an increase in the signal amplitude reduces the effect of the noise. The distinguishing feature of multiplicative noise is, however, that a rise in signal amplitude does not change the effect of the noise.

In Ref. 17 tests are reported of the change in error rate resulting from a change in signal level. Tests of this kind are extremely difficult to make, because--as will appear in Section VII--the clustering
of the errors generally prevents ascertaining the typical performance of a circuit in a reasonably short time. However, the results—smoother than might casually be expected—are depicted in Fig. 28 of Ref. 17. The figure has been cross-plotted to bring out the present point better, as Fig. 2 herewith.

![Graph of Bit Errors vs. Transmitted Level](image)

*Fig. 2—Long-haul AT&T toll circuits (600 bits/sec). Percentage of circuits with rates of error better than that shown on ordinate as transmitting level is reduced. Dotted lines show typical slopes.*

If only additive noise triggered the errors, then the curves would all show the same slope as the order $n$ of the hyperbola in Eq. (6). This is added in dotted lines in Fig. 2, for various orders. The flattening of the curves for the higher signal levels indicates
the presence of an effect that is independent of signal level.

It is possible, from the figure, to gain some insight into the relative proportion of errors that come from the additive noise and the multiplicative noise (assuming this to be the effect independent of signal level) under the various conditions of the test of Fig. 2.

If the overlapping of errors from the two sources is ignored, then the probabilities of the two are numerically additive. Let $P$ be the cumulated probability of error independent of level, and $K_y^{-n}$ the probability of error at amplitude level $y$ from the additive impulse noise. Then the total probability of error at level $y$, $P(y)$ is

$$P(y) = P + K_y^{-n}$$

or

$$K_y^{-n} = P(y) - P$$

If then the further assumption is made that $n$ stays constant over the range of $y$, a logarithmic plot of Eq. (13), with various trial values of $P$, should show a straight line for one of these values. This has been done, with the resulting values of $P$ and $n$ as shown in Fig. 3.

The figures indicate that for the median circuits (50 per cent), at the "normal" operating level of -6 dbm, the multiplicative noise accounts for 0.23 errors per $10^5$ bits, out of a total of 0.27 errors per $10^5$ bits, or a fraction of 85 per cent. This is quite large.
Fig. 3 — Number of errors independent of level, as level is changed, and slope of variation for errors varying with level. Plotted versus percentage of circuits, from Fig. 2.

For the marginal circuits at the boundary between the 80 per cent "better" circuits and the 20 per cent "poorer" ones, the multiplicative noise accounts for 1.3 errors per $10^5$ bits, out of a total of 2.33 errors per $10^5$ bits. This amounts to 58 per cent.

Because of the many uncertainties involved in the determination, these figures are not to be taken too literally. But they do show that the multiplicative noise is certainly important under the conditions of the AT&T tests (particularly, the use of switched circuits taken out of telephone traffic, and not private line circuits).

The results also indicate that if a critically exact model is needed to illustrate the timing of the peaks of impulsive noise, a further experimental examination of this noise is necessary.
III. GEOMETRIC, BINOMIAL AND POISSON DISTRIBUTIONS

GEOMETRIC AND BINOMIAL DISTRIBUTIONS

The statistical grouping of errors in a data transmission can be briefly reviewed for the case where these errors occur completely at random. This is a conventional treatment for the general case of fully random events. (18)

The data system sends bits of information in successive equal intervals of time. Assume that a fraction \( \alpha \) of these are in error. Then the probability of error in the very next bit is

\[ p(\alpha, 1) = \alpha \]  \( (14) \)

and the probability of no error in the very next bit is

\[ q(\alpha, 1) = 1 - \alpha \]  \( (15) \)

The probability of no error in the next succeeding \( K \) bits is

\[ q(\alpha, K) = (1 - \alpha)^K \]  \( (16) \)

This is the probability that in a long test no error will be found among intervals at least as long as \( K \) bits. It is known as a "geometric" distribution of the events (or errors).

It is noted that the limit of the quantity \( (1 - (\alpha K/K))^K \), as \( K \) becomes indefinitely large (see Eq. (5.4), Ref. 18) is

\[ \lim (1 - (\alpha K/K))^K = \exp(-\alpha K) \text{, as } K \to \infty \]  \( (17) \)
That is
\[ q(\alpha, K) = \exp(-\alpha K), \text{ as } K \to \infty \]  \hspace{1cm} (18)

In a time interval of \( K \) bits duration there can be \( c \) errors, where
\[ c \leq K \]  \hspace{1cm} (19)

The probability of any given grouping of the \( c \) errors is
\[ p = \alpha^c (1 - \alpha)^{K-c} \]  \hspace{1cm} (20)

But there are \( \binom{K}{c} \) such groupings, so that the total probability of exactly \( c \) errors in the time interval of \( K \) bits is
\[ p(\alpha, c) = \binom{K}{c} \alpha^c (1 - \alpha)^{K-c} \]
\[ = \left( \frac{K!}{c!(K-c)!} \right) \alpha^c (1 - \alpha)^{K-c} \]  \hspace{1cm} (21)

This is known as a "binomial" distribution of the events.

As \( K \) becomes indefinitely large, it approaches (see Eq. (5.6) Ref. 18)
\[ p'(\alpha, c) = \left( \frac{(\alpha K)^c}{c!} \right) \exp(-\alpha K) \]  \hspace{1cm} (22)

**POISSON DISTRIBUTION**

So far the discussion has covered the occurrence of errors in discrete bits or intervals of time. One can also consider time as a continuous variable, and the possibility of errors, or events,
occurring at any instant. This can be approached by assuming the bit intervals to become indefinitely small, and correspondingly the number $K$ of them in any given time interval to become indefinitely large. The results have already been foreshadowed in Eqs. (18) and (22), which lead to the Poisson equations. (A slightly different approach to the development of the latter is given by Davenport and Root\(^{(19)}\).)

For the probability of finding no errors in intervals of at least a given duration (in the course of a long test) as given by Eq. (18), one sets this given duration as say $u$, and puts

$$au = \gamma K$$  \hspace{1cm} (23)

where $a$ is the probability of an error per unit time, in the same units of time that $u$ is measured. Then $K$ is allowed to increase indefinitely, with $\gamma$ decreasing in proportion, to keep $au$ constant. Also, since the interval in question is at least $u$, the probability $Q$ no longer really refers to a density but to a cumulative probability with respect to $u$, and can be written as $Q$. Then

$$Q(a, u) = \exp (-au)$$  \hspace{1cm} (24)

For the probability of finding exactly $c$ errors in intervals of duration $u$, one can substitute in the same way in Eq. (22). Then, by increasing $K$ indefinitely, with $\gamma$ correspondingly decreasing by the limitation of Eq. (23), one gets

$$p(a, c) = \left(\frac{(au)^c}{c!}\right) \exp (-au)$$  \hspace{1cm} (25)
Here $a_1$ is used, because in practice the equation is often written with $u$ in mind as the unit of time, and $a = a_1 u$. Then

$$p(a, c) = (a^c/c!) \exp(-a)$$  \hspace{1cm} (26)

It is often desired to cumulate the probability to that for at least $c$ errors in the interval during which the long-time average is $a$. Consider the following probabilities in this interval. By "exactly $c$" it is meant that the number of errors is not smaller than $c$, nor is it equal to or greater than $c + 1$.

$$p(a, c-1) = \text{probability of exactly } c-1 \text{ errors}$$

$$p(a, 2) = \text{" } \text{" } \text{" } 2 \text{ errors}$$

$$p(a, 1) = \text{" } \text{" } \text{" } 1 \text{ error}$$

$$p(a, 0) = \text{" } \text{" } \text{" } \text{no errors}$$  \hspace{1cm} (27)

The probability of $c$ events or more is that of all the numbers of events not bracketed in Eqs. (27). Namely, it is one less the sum of all the probabilities bracketed, or

$$P(a, c) = 1 - \left(1 + a/1! + a^2/2! + \ldots + a^{c-1}/(c-1)!ight) \exp(-a)$$  \hspace{1cm} (28)

This has been tabulated and plotted by Campbell.  \hspace{1cm} (20)

**ILLUSTRATION AND CURVES**

In order to exemplify the equations which have been discussed, a specific example has been selected from a table of nearly random numbers. From this table a sequence of 2000 decimal digits was selected. The gaps between successive zeros were noted, and their distribution tabulated and plotted.
At this point it is important to note two possible conventions in designating this gap. In Fig. 4 appears a continuous scale of

![Diagram of continuous time scale with events A, B, C, D and time intervals AB, BC, CD.](image)

**Fig. 4**—Errors occurring in continuous time scale. Time units are called "bits" only for comparison with Fig. 5.

time which is divided into successive equal intervals of time. These are called "bits" merely for convenience in a comparison to be made below. Events appear at A, B, C and D. Events C and D are separated by a fraction of a "bit," and this is, of course, possible in a continuous distribution of events.

In Fig. 5 the intervals of time refer to actual bits in a

![Diagram of discrete time scale with events A, B, C.](image)

**Fig. 5**—Errors occurring in discrete time scale of bits.
transmission system, and are therefore discretely quantized. Here events A and B are contiguous, and according to the convention one uses, may be described as separated by one bit or by zero bits. Similarly B and C may be considered as separated by one or by two bits. Here the C and D of Fig. 4 cannot exist as separate events, and are shown as merged into the single event C.

For the illustrative distribution, which is of course discrete, we have the choice of conventions, and will select that illustrated in Fig. 4, which can apply to both figures. This is the opposite convention to that used for K in Eq. (16). If we take u to be measured in bit units of time in Fig. 4, we therefore have

$$K = u - 1$$

(29)

and Eqs. (16) and (24) become respectively

$$q(\alpha, u) = (1 - \alpha)^{u-1}$$

(30)

$$Q(\alpha, u) = \exp -\alpha(u - 1)$$

(31)

For the illustration with the 2000 digits,

$$\alpha = a = 0.1$$

(32)

The experimental points and Eqs. (30) and (31) are shown
plotted in Fig. 6. Equation (30), of course, is meant to be used for integral values of \( u \), but it can also be computed for fractional values, and a few have been shown for \( 1 < u < 2 \), to judge the comparison with Eq. (31).

Examination of Fig. 6 shows that the limit designated in Eq. (17) indicates a fair approximation in this illustration, even for \( K \) as low as plotted.
Curves of Eq. (26) are plotted in Fig. 7, and of Eq. (28).

**Fig. 7**—Probability of exactly $c$ events in interval having long-time average $a$ (Poisson)

in Fig. 8. This last is a re-plot of a portion of the curves
Fig. 8—Probability of at least \( c \) events in interval having long-time average \( a \) (Poisson)

presented by Campbell,\(^{(20)}\) to somewhat different scales.

Because of the simplicity and often reduced number of parameters involved it is frequently convenient to use the Poisson laws rather than the geometric or binomial laws, even when the variable is discrete. It is then necessary to check that the granularity error is not too large and that the time variable is used correctly. Both of these were done in the case of the numerical illustration which was described above.
MAXIMUM AND MINIMUM INTER-ERROR SPACINGS

A distribution of inter-error spacings, such as illustrated in Fig. 6, does not extend indefinitely each way when it represents an experimental and therefore finite test. If the test contained \( N \) errors, then there are \( N \) points in the curve, and \( N + 1 \) spaces.

The first point to the extreme left indicates the largest inter-error spacing in the finite test. Its probability in the test is \( 1/N \). In a test duration of \( T \) time units, and average error occurrence of \( a \) errors per time unit

\[
N = aT
\]  

(33)

For convenience \( T \) is often measured in bit durations.

Then the probability for the first point at the left is

\[
Q(a, u) = 1/N = 1/(aT)
\]  

(34)

The duration of the interval \( u \), having this probability, assuming the Poisson law of Eq. (24), is given by

\[
1/(aT) = \exp(-au)
\]  

(35)

The solution in terms of \( au \), which is the potential number of errors in the interval \( u \), as a long-term average, is

\[
a u = \log_e aT
\]  

(36)

The direct solution in terms of \( u \) is

\*See Ref. 14, p. 32.
\[ u_0 = (1/a) \log_e aT \]  

Equation (37) shows an interesting course if we assume the total testing time \( T \) as fixed, and vary \( a \). In a general way one can say that as \( a \) increases, \( 1/a \) will diminish faster than \( \log_e aT \) increases, so that \( u_0 \) will diminish. This is reasonable, for one can expect that as the error rate increases, the maximum error-free gap will diminish.

However, when the error rate \( a \) is very small, so that the total number of errors \( N = aT \) is little more than 1, say \( 1 + \varepsilon \), then

\[ aT = 1 + \varepsilon \]  
\[ a = (1 + \varepsilon)/T \]  
\[ \log_e aT = \varepsilon - \varepsilon^2/2 + \ldots \]  
\[ u_0 = \varepsilon T/(1 + \varepsilon) - \varepsilon T \]  

Thus in this region, the maximum error-free gap appears to increase as the error rate increases. This paradox arises from our assuming that the number of error-free gaps is equal to the number of errors, and not providing an exact definition of a gap at the very beginning and very end of the test (where it cannot be bounded by two errors). The critical point of maximum gap can be investigated, and is found to appear at \( aT = \varepsilon \). This region is generally of minor practical importance and the paradox will be ignored.

The last point at the extreme right of the distribution of Fig. 6, in an experimental test containing \( N \) errors, indicates
the smallest inter-error spacing in the test. Its cumulated probability in the test is \( 1 - \frac{1}{N} \). The duration of the interval \( u_1 \) having this probability, again assuming the Poisson law of Eq. (24), is given by

\[
1 - \frac{1}{N} = 1 - \frac{1}{aT} = \exp (-au_1)
\]  

(42)

The solution in terms of \( au_1 \) is

\[
-au_1 = \log_e \left( 1 - \left[ \frac{1}{aT} \right] \right)
\]

(43)

which can be expanded to

\[
a u_1 = \frac{1}{aT} + \frac{1}{(2aT)^2} + \frac{1}{(3aT)^3} + \ldots
\]

(44)

The direct solution in terms of \( u_1 \) is

\[
u_1 = \frac{1}{(aT)} + \frac{1}{(2aT)^2} + \frac{1}{(3aT)^3} + \ldots
\]

(45)

These are generally rapidly converging series, except in the immediate neighborhood of \( aT = 1 \), a value which also causes trouble in Eq. (37), and can similarly usually be ignored.

It is to be noted that the above treatment is simpler and otherwise differs somewhat from one appearing earlier, (22) although the results, at least for the maximum spacing, are not very much different. A logical flaw was pointed out in the earlier treatment.

**ILLUSTRATION**

To give reality to the equations discussed above, one can
consider their application to a specific example. We will assume a one-minute test on a data system, specified as follows:

\[ \text{Bit rate} = 1,000 \text{ bits/sec} \]

\[ T = 60,000 \text{ bits} \]

It is desired to determine the expected maximum \((u_o)\) and minimum \((u_1)\) error-free intervals (assuming the Poisson distribution) as the total number of errors \(N = aT\) changes from 1 to 60,000. This is merely a question of applying Eqs. (37) and (43). The results are plotted as the dotted lines (marked "continuous") in Fig. 9. Because of the difficulties with the formulas below.

\[ \text{Fig. 9—Expected maximum and minimum error-free intervals in continuous and discrete distributions} \]
N = e, it is noted that the minimum interval becomes greater than the maximum. This can be disregarded in line with the previous discussion.

The solid lines in Fig. 9 are discussed immediately below.

**MAXIMUM AND MINIMUM GAPS IN DISCRETE DISTRIBUTIONS**

It is to be noted that the maximum gap shown by the dotted lines in Fig. 9 is still about 10 bits in duration for aT = 60,000. This seems paradoxical if the number of errors equals the number of bits transmitted.

The reason, of course, is that Eq. (41) is derived from a Poisson law, which assumes a continuous distribution of errors. That is, the "bit" is merely a convenient designation of unit time interval, and more than one error can occur in it (as illustrated, for example, in Fig. 4). Thus, if some bits contain more than one error, there must be gaps of more than one "bit" duration.

The seeming paradox may be resolved by deriving the formulas from the geometric law of Eq. (30), for a discrete distribution. Then, instead of Eqs. (35) and (42), we have

\[ \frac{1}{(aT)} = (1 - a)^{u_0 - 1} \]  \hspace{1cm} (46)

\[ 1 - \frac{1}{(aT)} = (1 - a)^{u_1 - 1} \]  \hspace{1cm} (47)

The solution of Eq. (46) for the maximum is obtained by taking logarithms of each side.
\[ \log aT = (u_o - 1) \log \left(1/(1 - a)\right) \] (48)

\[ u_o = \frac{\log aT}{-\log(1 - a)} + 1 \] (49)

For the minimum, correspondingly

\[ u_1 = \frac{-\log \left(1 - (1/aT)\right)}{-\log(1 - a)} + 1 \] (50)

The computations of Eqs. (49) and (50) for the illustration of the previous section have been plotted as the solid lines ("discrete") of Fig. 9. It is to be observed that there is no longer any paradox and that the maximum inter-error spacing approaches 1 bit as \( N \) becomes equal to \( T \) at 60,000 bits. The departures from the continuous to the discrete laws become significant only as the spacing gets small. In the illustration this occurs as the spacing becomes smaller than 10 to 100 bits.

**ERRORS APPEARING IN BURSTS**

The discussions above have been based on errors occurring singly. It will appear in Section VI that a somewhat closer approach to actual experience can often be obtained by assuming errors to occur in bursts. If this theory is to be followed, then events which have been discussed above are not to be considered as simple errors, but as the initiations of bursts of errors. The bursts can be assumed in a variety of forms, but in general the number of errors in a burst, and their spacings in the burst, are erratic.
Thus, the maximum and minimum inter-error spacings become the corresponding spacings between burst initiations. The determinations, which have previously been affected by the paradox of error events overlapping within a bit, now become further affected by the paradox of burst initiations occurring as closely together as successive bits, if the bursts are to contain a multiplicity of errors.

The mathematics of a theory to fit such hypotheses can become very complicated. However, a neat model has been presented, together with the solutions, by Gilbert. \(^{(23)}\)

The details of this model, based on a Markov chain, are given in Section VI, and only a few highlights will be sketched here. The circuit can exist in two states, G, with zero error probability; and B, with a probability of error of 1 - h (and therefore a burst); and changes randomly between these. \(P\) is the probability of a transition from G to B, and \(p\) of a transition back from B to G. Three other parameters, related to these (and designated respectively A, J and L) describe the inter-error spacings. The maximum expected inter-error spacing, \(u_0\), is given implicitly by the equation

\[
\frac{1}{\langle \tau \rangle} = AJu_0^{-1} + (1 - A)Lu_0^{-1} \tag{51}
\]

Here \(\alpha\) (as in Eq. (14)) indicates the probability of bit errors. For a typical situation a burst contains an average of four errors, so that the number of burst initiations in such a case is \(\alpha/4\). To make the solution comparable to the plots of Fig. 9 (although the maximum spacings continue to be between errors), it will be plotted against numbers of burst initiations whose
probability is $a$, which is taken as

$$a = \sigma/4$$  \hspace{1cm} (52)$$

The solution of Eq. (51) for $u_o$ can be greatly simplified by noting that for a practical case (as noted in Ref. 23) for large $u_o$ the second term on the right is negligible. Thus

$$(u_o - 1) \log 1/J = \log 4\sigma T + \log A$$  \hspace{1cm} (53)$$

$$u_o - 1 = (\log 4\sigma T + \log A)/\log 1/J$$  \hspace{1cm} (54)$$

The parameters $h$, $p$, $P$, $A$, $J$ and $L$, the values $\sigma T$, and the result $u_o$, appear in Tables 1 and 2. The determination of the parameters appears in Section VI. The results, $u_o$ versus $\sigma T$, are plotted as the solid line marked "Gilbert" in Fig. 10.

Fig. 10—Expected maximum error-free intervals in Gilbert Model, compared with case of errors occurring singly.
They are there compared with the results for errors occurring singly. These are shown dotted and are called "1-bit bursts."

Examination of Fig. 10 shows that there is a rather significant departure for high error burst incidence, as might be expected.

A corresponding minimum expected inter-error spacing could be obtained by setting up an equation analogous to Eq. (51). It would not, however, be of much significance since for all the practical cases it would come out as 1 bit. This is because the minimum spacing would always be between errors in the same burst, and the concentration of errors in the burst practically guarantees the occasional presence of immediately successive errors.

The presence of the data circuit in state B represents, of course, the occasion of a burst. One can compute the expected maximum and minimum spacings between bursts (as distinguished from spacings between errors, discussed above). In this case the expected minimum has significance; the maximum should be substantially the same as the corresponding inter-error spacing.

The proportion of time spent in condition G is

\[ P(G) = 1 - P(B) = p/(p + P) \]  \hspace{1cm} (55)

Thus, in a test of duration \( T \) the time \( T' \) remaining, after the total cumulated duration of the bursts has been subtracted, is

\[ T' = Tp/(p + P) \]  \hspace{1cm} (56)

The probability of a burst initiation is \( P \), hence in time \( T' \) there will be \( X = PT' \) burst initiations, which is
\[ X = \frac{P_T p}{(p + P)} \]  

(57)

Thus, if the distribution of burst initiations (subtracting out the burst durations themselves) is geometric, then by Eqs. (49) and (50) the maximum and minimum expected intervals between bursts are

\[ u_o = \frac{\log X}{-\log (1 - p)} + 1 \]  

(58)

\[ u_1 = \frac{-\log \left(1 - \left(\frac{1}{X}\right)\right)}{-\log (1 - p)} + 1 \]  

(59)

These have been evaluated for the illustration which we have been considering, and the results have been plotted as the solid lines and solid dots in Fig. 11 (see next page). Comparison is made with the dotted lines for errors occurring singly and called "1-bit bursts," as taken from Fig. 9, without any allowance for burst durations.

The curving back that occurs at the right-hand end of the solid curve for maxima in Fig. 11 is a consequence of the drop that occurs for \( \alpha = 0.473 \) in Table 1. It comes from a mediocre accuracy in the successive approximations used in calculating the relations between parameters and is not physically significant.
Fig. 11 — Expected maximum and minimum burst-free intervals in Gilbert Model, compared with error-free intervals, and also with case of errors occurring singly.
Table 1

PARAMETERS OF BURSTS

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>L</th>
<th>J</th>
<th>A</th>
<th>h</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1.08 \times 10^{-4}$</td>
<td>.340</td>
<td>$1.25 \times 10^{-5}$</td>
<td>.229</td>
<td>.401</td>
<td>$2.75 \times 10^{-5}$</td>
<td>.153</td>
</tr>
<tr>
<td>$1.08 \times 10^{-3}$</td>
<td>.340</td>
<td>$1.25 \times 10^{-4}$</td>
<td>.229</td>
<td>.401</td>
<td>$2.75 \times 10^{-4}$</td>
<td>.153</td>
</tr>
<tr>
<td>$1.08 \times 10^{-2}$</td>
<td>.340</td>
<td>.9975</td>
<td>.229</td>
<td>.401</td>
<td>$2.75 \times 10^{-3}$</td>
<td>.150</td>
</tr>
<tr>
<td>$1.08 \times 10^{-1}$</td>
<td>.337</td>
<td>.97</td>
<td>.243</td>
<td>.401</td>
<td>.0332</td>
<td>.151</td>
</tr>
<tr>
<td>.214</td>
<td>.332</td>
<td>.925</td>
<td>.270</td>
<td>.401</td>
<td>.0838</td>
<td>.151</td>
</tr>
<tr>
<td>.318</td>
<td>.321</td>
<td>.85</td>
<td>.322</td>
<td>.401</td>
<td>.170</td>
<td>.150</td>
</tr>
<tr>
<td>.422</td>
<td>.289</td>
<td>.70</td>
<td>.500</td>
<td>.407</td>
<td>.361</td>
<td>.145</td>
</tr>
<tr>
<td>.473</td>
<td>.20</td>
<td>.55</td>
<td>.89</td>
<td>.460</td>
<td>.666</td>
<td>.094</td>
</tr>
</tbody>
</table>
Table 2

ERROR-FREE INTERVALS VERSUS BURST RATE

<table>
<thead>
<tr>
<th>α</th>
<th>aT</th>
<th>( \nu_o )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.08 \times 10^{-4}</td>
<td>1.62</td>
<td>1.57 \times 10^4</td>
</tr>
<tr>
<td>1.08 \times 10^{-3}</td>
<td>1.62 \times 10^2</td>
<td>1.08 \times 10^4</td>
</tr>
<tr>
<td>1.08 \times 10^{-2}</td>
<td>1.62 \times 10^3</td>
<td>2.00 \times 10^3</td>
</tr>
<tr>
<td>1.08 \times 10^{-1}</td>
<td>2.43</td>
<td>243</td>
</tr>
<tr>
<td>0.214</td>
<td>0.32 \times 10^4</td>
<td>106</td>
</tr>
<tr>
<td>0.318</td>
<td>0.48 \times 10^4</td>
<td>53.8</td>
</tr>
<tr>
<td>0.422</td>
<td>0.63 \times 10^4</td>
<td>27.5</td>
</tr>
<tr>
<td>0.473</td>
<td>0.91 \times 10^4</td>
<td>17.9</td>
</tr>
</tbody>
</table>
IV. HYPERBOLIC DISTRIBUTION

OUTLINE OF DISTRIBUTION

Analysis of errors is really a form of extremal statistics, since in a well operating system errors occur only seldom, and one is dealing with the extreme "skirts" of the probability curve. Various distributions have been described by Gumbel.* In general, in these the extremal values are independent, and therefore they do not reflect the clustering that has been found in data transmission errors. A simpler empirical approach has led to the use of hyperbolic distributions to describe these errors. (12, 21)

The hyperbolic distribution probability density \( p(x) \) of the variable \( x \) is

\[
p(x) \, dx = nh^n \frac{dx}{(x + h)^{n+1}} \quad x \geq 0, \quad h > 0
\]

(60)

The cumulated probability \( P(x) \) of \( x \) or greater is

\[
P(x) = h^n/(x + h)^n \quad x \geq 0, \quad h > 0
\]

(61)

where

\[
h = \text{bias constant, and} \]

\[
n = \text{order of hyperbola}
\]

Normalizing constants have been introduced to integrate to a total probability of 1. The average value, \( \overline{x} \), is

\[
\overline{x} = h/(n - 1)
\]

(62)

*See Ref. 14, p. 32.
and the rms is

\[ \text{rms}(x) = \frac{h}{\sqrt{2/(n - 1)(n - 2)}} \]  

(63)

It is clear that \( \bar{x} \) is not finite unless \( n > 1 \), and \( \text{rms}(x) \)
is not finite unless \( n > 2 \) (see Eq. (11)).

RELATION BETWEEN LONG-TIME AVERAGE AND BIAS

Error burst occurrences can be described by a first-order hyperbola, i.e., where \( n \) in Eqs. (60) and (61) equals 1. Attention from here on will be centered on this case. From Eq. (62), \( x \) has no finite long-time average over its range from zero to infinity.

Experimental averages are finite and significant only because the experiments themselves are finite and have in fact often run only to about or somewhat beyond the centile boundary.

By running the integration to a finite boundary it is possible to derive a long-time average \( a \) that corresponds to the experimental one.

\[ a = \int_{o}^{2} xp(x) \, dx = \int_{o}^{2} xh/(x + h)^{2} \, dx \]  

(64)

Let

\[ v = x + h \quad \text{and} \quad y = z + h \]

Then

\[ a = h \int_{h}^{y} (1/v) \, dv - h^{2} \int_{h}^{y} (1/v^{2}) \, dv \]  

(65)

\[ a = h \left[ \log_{e} \left( y/h \right) - 1 + (h/y) \right] \]  

(66)
Let \( k \) express a quantile boundary; i.e., \( k = 10 \) for decile, 100 for centile, etc., boundaries. At the \( k \)-ile boundary, the cumulated probability is

\[
P(x) = \frac{h}{x + h} = \frac{h}{y} = \frac{1}{k} \tag{67}
\]

\[
a = h \left[ \log_e k - 1 + \frac{1}{k} \right] \tag{68}
\]

\[
a = h A(k) \tag{69}
\]

Here \( A(k) \) is a simple numeric. It may be computed as in Table 3. The terms "milile" and "micrile" have been coined to fit.

Plots of the relationships between \( a \) and \( h \) are shown in the dotted lines of Fig. 12 for the four quantile boundaries of Table 3. These are compared with some experimental values to be taken up later in this Section.

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PARAMETER A(k) FOR FINITE AVERAGES</strong></td>
</tr>
<tr>
<td>Interval</td>
</tr>
<tr>
<td>( k )</td>
</tr>
<tr>
<td>( \log_e k )</td>
</tr>
<tr>
<td>( A(k) )</td>
</tr>
</tbody>
</table>

**PROBABILITY OF EXACTLY \( c \) EVENTS**

This probability is obtainable by integrating Eq. (60) for \( n = 1 \), using the meaning for "exactly \( c \)" discussed for Eqs. (27). It
It is advantageous in practice to interpret the bias $h$ in terms of the long-time average, integrated up to the quantile designated by $k$. Thus Eq. (73) becomes

$$p(a,c) = \frac{aA}{(cA + a)(cA + A + a)}$$  \hspace{1cm} (74)

and

$$p(a,0) = \frac{A}{(A + a)}$$  \hspace{1cm} (75)

$$p(a,1) = \frac{aA}{(A + a)(2A + a)}$$  \hspace{1cm} (76)

$$p(a,2) = \frac{aA}{(2A + a)(3A + a)}$$  \hspace{1cm} (77)

$$p(a,c) = \frac{aA}{(3A + a)(4A + a)}, \text{ etc.}$$  \hspace{1cm} (78)

These values have been plotted in Fig. 13 for $k = 1000$ and $a$.

Fig. 13—Probability of exactly $c$ events in interval having long-time average $a$ ($k = 1000$)
corresponding \( A(k) = 5.91 \). They can be compared with the similar set of probability densities for the Poisson distribution plotted in Fig. 7. Both of these are plotted on a special logarithmic probability paper described in the Appendix.

Examination of Figs. 7 and 13 shows:

1. The probability of zero events remains much higher for the hyperbolic than for the Poisson laws as \( c \) increases.

2. The probabilities for various numbers of events peak much more sharply, and at lower values of \( c \), for the Poisson than for the hyperbolic laws.

3. In the hyperbolic law the probabilities for finite numbers of occurrences always stay below the probability for zero occurrence. For the Poisson law they cross and are above the probability for zero occurrence for larger \( c \).

**Cumulative Probabilities of Events**

From Eq. (70), with an appropriate change in limits, it is possible to compute the cumulative probabilities of events for the hyperbolic distribution. These correspond to the exponential summation of Eq. (28) for the Poisson distribution. They are in general more practical to use than the densities.

The cumulated probability is

\[
P(a, c) = \left[-\frac{h}{(x + h)}\right]_c^\infty
\]

\[= \frac{h}{c + h}
\]

(79)
It is sometimes preferable to express $h$ in terms of $a$. This gives

$$P(a, c) = \frac{a}{(Ac + a)}$$ (81)

which has been plotted in Fig. 14 for $k = 1000$ and which is compared with Campbell's Poisson curves in dotted lines. These last are copied

Fig. 14—Probability of at least $c$ events in interval having long-time average $a$
(hyperbolic, $k = 1000$; and Poisson)
from Fig. 8. The comparison shows:

1. Except for the single value \( c = 1 \), the curves for one system cross those of the other. The Poisson probabilities are higher than the hyperbolic probabilities toward larger \( a \).

2. Very large values of \( c \) compared with \( a \) show much greater hyperbolic than Poisson probabilities.

**EVENT-FREE STRETCHES**

With some modification from the Poisson statistics, it is possible to derive from Eq. (73) the distribution of event-free intervals. For \( c = 0 \), the equation becomes

\[
p(a,0) = \frac{1}{(1 + h)}
\]  

(82)

Here \( p(a,0) \) represents the fraction of the unit intervals (chosen for Eq. (70)) in the test that have an expectation of containing no events. As the interval duration \( u \) is changed, the bias \( h \) is proportional to it, and we can take

\[
h = h_1 u
\]  

(83)

Then

\[
p(a,0,u) = \frac{1}{(1 + h_1 u)}
\]  

(84)

If we now set

\[
h_0 = \frac{1}{h_1}
\]  

(85)

we get

\[
p(a,0,u) = \frac{h_0}{(h_0 + u)}
\]  

(86)
Here \( p(a,0,u) \) represents the fraction of the total number of intervals of duration \( u \) in the test that contains no events.

One can, incidentally, check this in the following way. If there are no events in duration \( u \), then the density of events at that time in the test is not greater than 1 per \( u \) units of time, or \( 1/u \). If we call the probability of this \( R(1/u) \), then

\[
R(1/u) = h_0/(h_0 + u)
\]  \hspace{1cm} (87)

or

\[
R(1/u) = (1/u)/\left(\frac{1}{h_0} + \frac{1}{h_0}\right)
\]  \hspace{1cm} (88)

\[
1 - R(1/u) = \frac{1}{h_0}/\left(\frac{1}{l/u} + \frac{1}{1/h_0}\right)
\]  \hspace{1cm} (89)

We now notice that \( 1 - R(1/u) \) represents the fraction of intervals of duration \( u \) in the test when the density of events is at least \( 1/u \). If we then place

\[
c = \frac{1}{u}
\]  \hspace{1cm} (90)

and

\[
h = \frac{1}{h_0}
\]  \hspace{1cm} (91)

we obtain

\[
1 - R(c) = P(a,c) = h/(c + h)
\]  \hspace{1cm} (92)

This is the same Eq. (80) that we have started with.

The quantity \( p(a,0,u) \) in Eq. (86) is not strictly analogous to the quantity \( Q(a,u) \) developed for the Poisson distribution in Eq. (24). To obtain the quantity analogous to this latter, it is necessary to change the basis of the probability fraction from the proportion of the total time of the test, to the proportion of the number of
inter-event intervals. This is done as follows.

The quantity \( p(a, 0, u) \) is in a sense a cumulative probability because it includes parts of all longer intervals. The first step, therefore, is to remove the portions that form part of the longer intervals in the conventional manner, namely by a differentiation. The total test time \( \tau du \) devoted to event-free intervals of length between \( u \) and \( u + du \) is

\[
\tau du = -T \frac{d}{du} p(a, 0, u) \, du
\]  

(93)

The negative sign occurs because \( p \) diminishes as \( u \) increases, as a result of the direction of cumulation. Here

\[
T = \text{total test time (in same units as } u) \\
\tau du = \frac{Th_o}{(h_o + u)^2} \, du
\]  

(94)

The number \( m(u) \) of such intervals is equal to the time divided by the duration \( u \) of each interval (ignoring the \( du \) increment), namely

\[
m(u) \, du = \frac{\tau du}{u} = \frac{T(h_o/u)}{(h_o + u)^2} \, du
\]  

(95)

The cumulated number \( M(u) \) of such intervals is

\[
M(u) = T \int_u^\infty \frac{h_o/u}{(h_o + u)^2} \, du
\]  

(96)

\[
= T \left\{ \frac{1}{h_o + u} - \frac{1}{h_o} \log \frac{h_o + u}{u} \right\}_u
\]  

(97)

\[
= T \left\{ - \frac{1}{h_o + u} + \frac{1}{h_o} \log \frac{h_o + u}{u} \right\}
\]  

(98)
That is

\[ M(u) = \left( \frac{T}{h_o} \right) N(U) \]  

(99)

where

\[ N(U) = \log_e \left( \frac{1 + U}{U} \right) - \frac{1}{1 + U} \]  

(100)

\[ U = \frac{u}{h_o} \]  

(101)

The quantity analogous to \( QA(a,u) \) in Eq. (24) is then

\[ Q(a,u) = \left( \frac{1}{a T} \right) M(u) \]  

(102)

\[ = \left( \frac{1}{a h_o} \right) N(U) \]

For convenience, Table 4 is an abbreviated table of \( N(U) \).

<table>
<thead>
<tr>
<th>U</th>
<th>N(U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.0001</td>
<td>8.21</td>
</tr>
<tr>
<td>.001</td>
<td>5.91</td>
</tr>
<tr>
<td>.01</td>
<td>3.63</td>
</tr>
<tr>
<td>1</td>
<td>1.489</td>
</tr>
<tr>
<td>10</td>
<td>.1932</td>
</tr>
<tr>
<td>100</td>
<td>4.40 \times 10^{-3}</td>
</tr>
<tr>
<td>1000</td>
<td>4.93 \times 10^{-5}</td>
</tr>
<tr>
<td></td>
<td>4.99 \times 10^{-7}</td>
</tr>
</tbody>
</table>

A plot is shown in Fig. 15 of the relationship between \( h_o \) and \( a \) for the four quantile boundaries (in dotted lines) listed in
Fig. 15—Relation between bias $h_0$ and long-time average $a$ of events in hyperbolic distribution of event-free stretches

Table 3. It is compared with some experimental values to be discussed later.

MAXIMUM AND MINIMUM INTER-EVENT SPACINGS

In a manner analogous to that of Section III it is possible to find the maximum and minimum expected gaps between events in a hyperbolic distribution.

The maximum is given implicitly by

$$Q(a, u_o) = 1/aT$$  \hspace{1cm} (103)$$

and the minimum by

$$Q(a, u_1) = 1 - (1/aT)$$  \hspace{1cm} (104)$$
Equation (103) is transcendental and may be solved graphically. First one can set up a parameter $V$ as follows:

$$V = \frac{u_o}{T}$$

(105)

Here $V$ represents the ratio of the maximum gap to the total period of the test. Then one can write, from Eqs. (101-103)

$$Q = \frac{N(U_o)}{(ah_o)} = \frac{1}{(aT)}$$

$$1/Q = ah_o/N = aT, \quad N = h_o/T = \frac{u_o}{T} \left( \frac{h_o}{u_o} \right)$$

(106)

$$N(U_o) = V/U_o$$

Thus if $N(U)$ is plotted on log-log paper, as in Fig. 16, and then

![Graph](image-url)
V/U is plotted for a specific V as a diagonal straight line sloping down at 45°, the intersections of these two will give solutions of Eq. (106), and also therefore of Eq. (103) for that value of V. The U value at the intersection is then U_o. In Fig. 16 plots of V/U are shown for a wide range of values of V.

It is to be noted that for each value of V there are two values for U_o, one large and one small. This has the same cause as the two values for solutions of Eq. (35), namely the ignoring of a definition to include the very first and very last gaps in the test. As in the previous case, only the larger value of U_o is significant, and the other will be ignored. It is to be noted, too, that there is no solution for

\[ V > 0.20 \]  

which is also analogous to the case for Eq. (35).

The solution of Eq. (104) can be quickly approximated by ignoring the 1/\(aT\) term on the right-hand side. Then

\[ Q(a,u) = 1 = \left(1/\left( \frac{a}{\Delta} \right) \right) N(U_1) \]  

or, from Eqs. (69) and (91)

\[ ah_o = A(k) = N(U_1) \]  

From this one can compute \( u_1 \) as

\[ u_1 = U_1 h_o \]
or

\[ u_1 = \frac{U_1 \Delta t}{aT} \]  \hspace{1cm} (111)

In studying this with numerical values, it is found that \( u_1 \) is generally small. In fact it is generally much smaller than one bit (assuming this being used as a unit of time), except where \( aT \) itself is quite small. This last area is where the computations become unreliable, as noted for Eq. (106). The minimum spacing between events being less than the one bit, the fact that \( u_1 \) is smaller than one bit means that more than one event occurs within one bit in times of high event concentration during the test. This problem, discussed in Section III, is due to the fact that the hyperbolic law of Eq. (80) is expressed in continuous and not discrete terms. However, the hyperbolic law is purely empirical, and an adaptation of it to discrete terms would, everything else remaining consistent, require a compounding of the fabrication.

It will be found below that it is not necessary to pursue the matter further, because the simple hyperbolic law is not too well adapted to describing error occurrences in very short time intervals. In using the law for long time intervals the convenient time unit would include many discrete bits, and the time variable would therefore be essentially continuous.

**ILLUSTRATION**

At this point one can return to the illustration of error occurrences which was presented in Section III and show how the expected maximum and minimum inter-error spacings vary over a test of fixed
duration as the total number of errors changes from a very small to a very large value. As before, the parameters assumed are

\[
\begin{align*}
\text{Test time} & = 1 \text{ min} \\
\text{Bit rate} & = 1,000 \text{ bits/sec} \\
T & = 60,000 \text{ bits} \\
A(k) & = A(T) = 10.00
\end{align*}
\]

The computation is merely a question of applying these parameters to Eqs. (106) and (111). The resulting \( u_0 \) and \( u_1 \), in bits duration, are plotted as the solid lines in Fig. 17. They are there compared with the Poisson Eqs. (37) and (43) (also plotted in Fig. 9 as dotted lines).

Examination of Fig. 17 can start with the expected maxima (scale A). The general trend for the hyperbolic law is similar to that for the Poisson law. However, in the process of reflecting the error-clustering characteristics of the hyperbolic law, the solid line nearly always shows greater maximum intervals than the dotted lines. And it exaggerates the "paradox" of showing large intervals when the total number of errors is equal to the total number of bits duration of the test. Also, the hyperbolic law, in common with the Poisson law, shows a breakdown for very small numbers of errors. Indeed, it shows this breakdown commencing at a disquietingly large number of errors, and the matter should be further examined if critical designs or experiments are envisaged in this region.

With regard to the minima for the hyperbolic law (scale B), Fig. 17 shows what has already been noted. This is that these minima are very
small—much smaller than one bit where the computations are reliable. This is in great contrast to the Poisson law, where the expected minima can at least be plotted on the same scale (scale A) as the expected maxima.

These observations generally show that the simple hyperbolic law could not realistically describe error occurrence in a test having the parameters of the illustration. These error occurrences would be so extremely clustered in the description that during the high error density periods thousands of errors would occur in the same bit.
COMPARISONS WITH EXPERIENCE

Experimental information on the occurrence of errors in data transmission systems is still fairly scarce. A few examples are presented herewith for comparison with the formulas.

The author is indebted to E. J. Hofmann for the opportunity to examine basic data in some Lincoln Laboratory tests.\((24, 25)\) Fig. 18 shows some early plots of the relationship observed between the

![Graph](image-url)

**Fig. 18**—Distribution of word errors per minute in long-haul private-line data circuit (Lincoln Laboratory data,\((24)\) 1300 bits/sec) comparison with hyperbolic and Poisson formulas
short-time high incidence of word errors and the long-time incidence average. The experimental data are compared with plots of the hyperbolic formula and of the Poisson law. The approximation for the former is seen to be much better than for the latter. The abscissa scale follows suggestions advanced in the Appendix for what are called "nearly worst" intervals. For the "nearly worst decile" this interval lies between 0.05 and 0.15 in the cumulative probability. For the "nearly worst centile" it lies between 0.005 and 0.015 in the cumulative probability, etc.

More comprehensive measurements on the error distributions over a fair variety of circuits were later reported by the Lincoln Laboratory. (26-29) A plot of these is shown in Fig. 19. To condense the plot they are given various ordinate scales, as indicated. Plot H was not of errors but of records of noise impulses over a given threshold on a specific voice-frequency cable circuit. For simplicity the number of records per day is plotted rather than the number of impulses. For comparison a dotted straight line representing Eq. (80) is plotted as S. With the probability scale used this equation appears as a straight line with a slope of -1. The ordinate intercept of this line with the 50 per cent probability abscissa is a measure of the bias h. Two illustrations are also shown of specific Poisson distributions.

Additional measurements have more recently been reported to the CCITT. (30,31) One of these is a voice-frequency telegraph facility of Italcable between Rome and Milan which is plotted in Fig. 20. The other is a 1000-baud airline reservation wire circuit between Paris
Fig. 19—Experimental error distributions (Lincoln Laboratory (24-29))

A. Kingston-Canaveral (1960, 162 cumulated hours)
B. Kingston-Canaveral (1958, line A, 72 cumulated hours)
C. Kingston-Canaveral (1958, line B, 113 cumulated hours)
D. Lexington-South Truro (line A, 509 cumulated hours)
E. Lexington-South Truro (line B, 702 cumulated hours)
F. Milgo tests (623 cumulated hours)
G. Hawaiian cable (641 cumulated hours)
H. Analoger tape records (208 cumulated hours)
Fig. 20—Experimental error distributions (Italcable(30))

and Hilversum reported by Philips, and plotted in Fig. 21. The

Fig. 21—Experimental error distribution (Philips(31))
ordinates of these plots, incidentally, go to zero. This is accomplished on the otherwise logarithmic scale by an arrangement described in the Appendix. The effect of the zeros on the hyperbolic law plot is discussed below.

Other data have been reported by the Bell System.\(^{17}\) Figure 22 shows a plot of the incidence of error-free stretches for two cases in terms of cumulative probability. These are compared with the hyperbolic and Poisson formulas. The latter are significantly too low.

![Graph](image)

**Fig. 22**—Cumulative probability of error-free stretches in switched telephone circuits used for data transmission (Bell system data, \(^{17}\) 1200 bits/sec)
In the hyperbolic law the one adjusted constant is the bias parameter, and it is therefore significant to compare its experimental relation to the long-time average probability of the events with that from the theoretical derivations presented above.

The bias \( b \) in Eq. (80) and its relation to the long-time average \( a \) in Eq. (69) has already been noted in Fig. 12. Equation (80) covers the short-time incidence of events for given long-time averages. In the figure comparison is made with a variety of data from a number of the Lincoln Laboratory\(^{(24-29)}\) and Bell System\(^{(17)}\) experiments. Most of these data, at the extremes, are reached somewhere between the centile and milile limits. There appears to be a good general agreement, with some scattering.

The bias \( b_0 \) is used in the description in Eq. (86) of the distribution of event-free intervals. The plot in Fig. 15 shows its relation to \( a \) and also compares this with data from Lincoln Laboratory and Bell System sources. Few of these data, at the extremes, reached beyond the milile limit. As in Fig. 12, there is a general agreement. The scattering in this case, however, tends toward somewhat higher values than indicated theoretically.

These results and others, some published and some not, indicate the generality of the hyperbolic laws to describe long-time error distributions that have been experienced in practice. They give some feeling of security on the broad reliability of the results, where the tests are not of too short a duration.
LIMITATIONS OF SIMPLE HYPERBOLIC LAW

It is obvious that the law of Eq. (80) extends in one direction to unlimitedly large \( c \), with correspondingly low probability \( P(a,c) \); and in the other direction, to unlimitedly small \( c \), with a corresponding probability approaching one. On a logarithmic probability scale each goes to infinity—one in one direction and the other in the other.

The consequence of the first of these facts has already been touched on. It means that the law includes a finite (though small) probability that the errors per bit can run in the thousands or to any number that one chooses. This is, of course, embarrassing if one is dealing with a digital data system, and the test is short enough to make one bit a convenient unit of time. Under such conditions, as has already been noticed, the simple hyperbolic law cannot describe error occurrences realistically. Where, however, one is studying tests lasting say a month, the appropriate unit of time may be an hour, and this feature of the law presents no difficulty unless one goes to extremely low probabilities.

Although the matter has not as yet been studied altogether systematically, there is a strong indication that this limitation of the form of the law is not the only point involved. It also appears that the actual distribution in practical transmission systems is changing so that the law is really different over 1 or 10 minutes, or even somewhat longer, from what it is over a whole day, or a month, or a year. In the case of some radio links, the distributions of atmospheric noise which would cause the errors have been studied
systematically, and the changes in the law have been described in some detail.\(^{4,7,32}\) Thus, it is not too surprising to find this change in the distribution for data facilities, and it may even have a physical basis. The next two Sections will describe how one can treat the matter of two distributions.

At the other end of the scale in the probability law, one gets the case where sizable periods of time yield no errors. This is, of course, no hardship on the user of the system, but when such tests are plotted the curves truncate as was illustrated in Figs. 20 and 21. If there are sufficient points to the curve before the truncation, it can be established easily enough in spite of the zeros. If there are not enough points it merely means that the test has not proceeded long enough to plot a satisfactory curve, or that too short a time interval is taken as unit. This is not, therefore, strictly speaking, a limitation of the hyperbolic law, but of the physical phenomenon being measured. The point will be discussed again in Section VII.
V. PARETO DISTRIBUTIONS

In a recent paper Berger and Mandelbrot (33) comment on certain aspects of the one-dimensional random walk problem. In particular they have noted the distribution of times at which the random walk returned to or through the origin. Writers on the random walk problem have insistently pointed out how these times vary from a Poisson distribution. The times tend to come in irregular close clusters, between which long periods occur that are free of zero crossings. These properties immediately suggested a model for the times of occurrence of errors in a data transmission system.

The probability of the different time lapses from one to the next crossing has been studied by various authors, and some treatment of this is given by Feller. Berger and Mandelbrot give a solution, which after a brief discussion they simplify into a Pareto distribution.

OUTLINE OF PARETO DISTRIBUTION

This distribution is noted by Gumbel, and has also been studied by Miniruzzaman. (35,36) The law has been used in economic statistics to describe income distributions in a population. Gumbel gives it very simply as

\[ P(u) = 1 - u^{-k} \]  \hspace{1cm} (112)

where \( P(u) \) is the cumulative probability of a variable \( u \), and \( k \) is a constant. One can obviously cumulate in the opposite direction and write

---

* See Ref. 18, Chapter III.
** See Ref. 14, p. 151.
\[ Q(u) = 1 - P(u) = u^{-k} \]  \hspace{1cm} (113)

This is a hyperbolic distribution without bias and of order \( k \).

Figure 23 shows the distribution as a plot in logarithmic coordinates, as presented by Mandelbrot, for the probability \( Q(u) \) of an interval of at least \( u \) from one error to the next.

![Logarithmic plot of inter-error spacings](image)

Fig. 23—Logarithmic plot of inter-error spacings \( u \)

**FINITE EXTENT OF DISTRIBUTION**

The assumption is made in Fig. 23 that the test is of finite duration and that the maximum inter-error interval in it is \( u_0 \).

The number of intervals in the test is equal to the number of errors, if we assume that immediately successive errors are one bit apart (i.e., \( u = 1 \)), and if we ignore the details of how the very end intervals are bounded (which is permissible if there are enough errors in the test).
Consequently, if the number of errors in the test is \( N \), the probability of the single longest one \( u_0 \) is

\[
Q(u_0) = \frac{1}{N}
\]  

This is so indicated in Fig. 23.

It is possible to plot Fig. 23 in linear coordinates. If, as ordinate, one plots \( NQ(u) \) rather than \( Q(u) \), then each unit step in the scale represents one interval between errors in the test, and they are arrayed in order of duration \( u \) of this interval. This has been done in Fig. 24 (which is not to scale). The maximum interval \( u_0 \) comes at \( NQ(u_0) = 1 \), and the minimum, which is usually 1, comes at \( NQ(1) = N \).

![Figure 24 - Linear plot of inter-error spacings](image)

It is clear that the sum total of all the intervals \( u \) is equal to the total duration of the test (assumed as \( T \)). This is indicated in Fig. 24 by designating the area under the curve as \( T \). It is similarly designated as \( (T/N) \) in Fig. 23, but it must be understood
that the area there is meant to be taken with both the coordinates linear, and this is the significance of the parentheses.

It is convenient here to measure time in bit durations. Thus both \( u \) and \( T \) are measured in numbers of bits. If \( a \) is taken as the average number of bit errors during the test \( T \), then

\[
a = N/T, \text{ or } N = aT \tag{115}
\]

**Slope of Pareto Distribution versus Other Parameters**

Consider now an ensemble of tests, each of duration \( T \) bits, but having different numbers of errors \( N_i \). Then, for the tests having large numbers of errors \( 1/N_i \) is small, and for those with few errors \( 1/N_i \) is large. The quantity \( 1/N \), it is recalled, is the smallest ordinate in the plot of Fig. 23. Also, it is generally expected (though not necessarily true in individual cases) that the longest spacing \( u_0 \) in each test will be shorter when there are many errors and longer when there are fewer errors. These relations are indicated in Fig. 25, where \( i = 1 \) is the case for the fewest errors, and \( i = 3 \) for the most errors.

![Fig. 25 — Variation of slope with total number of errors \( N \)]
The consequence of the relations is that the slope \(-k\) cannot remain constant in the ensemble as the number of errors changes. One can expect that \(k\) increases numerically, in some fashion, with the number of errors.

One gets, from Eqs. (113) and (114)

\[
N = u^k_0
\]  

(116)

By inverting Eq. (113)

\[
u(Q) = Q^{-1/k}
\]  

(117)

One is first tempted to integrate under the curve of Fig. 24, namely

\[
T = \int_{1/N}^{1} Q^{-1/k} dQ
\]  

(118)

\[
T = \left[ \frac{NQ^{1-1/k}}{1 - 1/k} \right]_{1/N}^{1}, \quad k \neq 1
\]  

(119)

\[
T = Nk \left[ N^{1/k-1} - \right] / (1 - k), \quad k \neq 1
\]  

(120)

However, it is to be noted that the curve in Fig. 24 is really stepped, and Eq. (118) should be a summation and not an integral. Some exploration has shown that the integral, while easier to handle than the summation, occasionally leads to a sizable error.

In Fig. 24 the \(n\)th interval in order of length has the cumulative probability

\[
Q(u_n) = n/N = u_n^{-k}
\]  

(121)
\[ u_n = (N/n)^{1/k} \]  

(122)

The total test duration \( T \), to be substituted for Eq. (118) therefore is

\[ T = u_1 + u_2 + \ldots + u_n + \ldots + u_o \]  

(123)

\[ = N^{1/k} \left[ (1/1)^{1/k} + (1/2)^{1/k} + \ldots + (1/n)^{1/k} + \ldots + (1/N)^{1/k} \right] \]  

(124)

The expression in the large brackets of Eq. (124) comes from the hyper-harmonic series.

**INCOMPLETE RIEMANN ZETA FUNCTION**

The Riemann Zeta function is defined as:

\[ \zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \]  

(125)

where \( n \) = a positive integer, running progressively from one to infinity

\( s \) = a constant in the summation, argument of the function.

It is seen to be a hyper-harmonic series summed to infinity.

The function which is really needed for the present purpose is

\[ S(N, s) = \sum_{n=1}^{N} \frac{1}{n^s} \]  

(126)
where \( N \) is a finite integer, upper limit to the summation. This is the "incomplete" Zeta function. It represents the term in the large brackets of Eq. (124), and here \( s \) is real and positive, and equal to \( 1/k \). Mathematicians have in the past been interested in the Zeta function of a complex argument, and have given little attention to the incomplete form. Thus, no tables of this latter have been found.

A computation of this function in the regions of present greatest interest is presented as Table 5, which also includes a calculation of \( N^s \) for the same parameters. A plot is shown of the function in Fig. 26, over the major part of its range. The function does not plot well because of the large range of values, which also go very close to 1. Thus the figure, even with a special scale, gives only the very general trend of the function.

**APPLICATION TO COMPUTATIONS**

Once computed, the incomplete Zeta function of Eq. (126) can be used to compute the test duration \( T \) of Eq. (124). This becomes

\[
T = N^{1/k} S(N,1/k) \tag{127}
\]

However, there is still a problem, because in most practical cases \( T \) and \( N \) will be the independent variables, and \( k \) will be the resultant or dependent variable. That is, Eq. (127) would need to be solved for \( k \). Since it is transcendental, the solution is to be obtained by successive approximations, which is a lengthy and tedious process.

*Courtesy of the 7044/7040 complex at The RAND Corporation.*
<table>
<thead>
<tr>
<th>$s = 0.100000$, $k = 10$</th>
<th>$s = 0.500000$, $k = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S(N, s)$</td>
<td>$N^8$</td>
</tr>
<tr>
<td>1.933033</td>
<td>1.071773</td>
</tr>
<tr>
<td>4.550882</td>
<td>1.174619</td>
</tr>
<tr>
<td>8.619334</td>
<td>1.258925</td>
</tr>
<tr>
<td>1.62368x10</td>
<td>1.349283</td>
</tr>
<tr>
<td>3.73040x10</td>
<td>1.478758</td>
</tr>
<tr>
<td>6.981874x10</td>
<td>1.584893</td>
</tr>
<tr>
<td>1.305143x10</td>
<td>1.698646</td>
</tr>
<tr>
<td>2.980869x10</td>
<td>1.861646</td>
</tr>
<tr>
<td>4.036227x10</td>
<td>1.925350</td>
</tr>
<tr>
<td>5.565207x10</td>
<td>1.995262</td>
</tr>
<tr>
<td>1.038811x10</td>
<td>2.138469</td>
</tr>
<tr>
<td>2.370994x10</td>
<td>2.343673</td>
</tr>
<tr>
<td>4.423059x10</td>
<td>2.511886</td>
</tr>
<tr>
<td>8.254026x10</td>
<td>2.692173</td>
</tr>
<tr>
<td>1.882878x10</td>
<td>2.950509</td>
</tr>
<tr>
<td>3.513606x10</td>
<td>3.162278</td>
</tr>
<tr>
<td>6.556652x10</td>
<td>3.389245</td>
</tr>
<tr>
<td>1.495648x10</td>
<td>3.714471</td>
</tr>
<tr>
<td>2.790981x10</td>
<td>3.981072</td>
</tr>
<tr>
<td>5.208158x10</td>
<td>4.266807</td>
</tr>
<tr>
<td>1.188038x10</td>
<td>4.676242</td>
</tr>
<tr>
<td>2.216958x10</td>
<td>5.011872</td>
</tr>
<tr>
<td>4.13698x10</td>
<td>5.371592</td>
</tr>
<tr>
<td>9.436923x10</td>
<td>5.887040</td>
</tr>
<tr>
<td>1.760992x10</td>
<td>6.309573</td>
</tr>
<tr>
<td>3.286128x10</td>
<td>6.762433</td>
</tr>
<tr>
<td>7.496015x10</td>
<td>7.411344</td>
</tr>
<tr>
<td>1.398806x10</td>
<td>7.943282</td>
</tr>
<tr>
<td>$s(N,s)$</td>
<td>$N^s$</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>1.500000</td>
<td>2.000000</td>
</tr>
<tr>
<td>2.283333</td>
<td>5.000000</td>
</tr>
<tr>
<td>2.928968</td>
<td>1.000000x10^2</td>
</tr>
<tr>
<td>3.597739</td>
<td>2.000000x10^2</td>
</tr>
<tr>
<td>4.492055</td>
<td>5.000000x10^2</td>
</tr>
<tr>
<td>5.187376</td>
<td>1.000000x10^3</td>
</tr>
<tr>
<td>5.878027</td>
<td>2.000000x10^3</td>
</tr>
<tr>
<td>6.792811</td>
<td>5.000000x10^3</td>
</tr>
<tr>
<td>7.128992</td>
<td>7.000000x10^3</td>
</tr>
<tr>
<td>7.485444</td>
<td>1.000000x10^4</td>
</tr>
<tr>
<td>8.1784</td>
<td>2.000000x10^4</td>
</tr>
<tr>
<td>9.0945</td>
<td>5.000000x10^4</td>
</tr>
<tr>
<td>9.7876</td>
<td>1.000000x10^5</td>
</tr>
<tr>
<td>10.4806</td>
<td>2.000000x10^5</td>
</tr>
<tr>
<td>11.3967</td>
<td>5.000000x10^5</td>
</tr>
<tr>
<td>12.0896</td>
<td>1.000000x10^6</td>
</tr>
<tr>
<td>2.000000x10^6</td>
<td>2x10^5</td>
</tr>
<tr>
<td>5.000000x10^6</td>
<td>5x10^5</td>
</tr>
<tr>
<td>1.000000x10^6</td>
<td>10^6</td>
</tr>
<tr>
<td>2.000000x10^6</td>
<td>2x10^6</td>
</tr>
<tr>
<td>5.000000x10^6</td>
<td>5x10^6</td>
</tr>
<tr>
<td>1.000000x10^7</td>
<td>10^7</td>
</tr>
<tr>
<td>2.000000x10^7</td>
<td>2x10^7</td>
</tr>
<tr>
<td>5.000000x10^7</td>
<td>5x10^7</td>
</tr>
<tr>
<td>1.000000x10^8</td>
<td>10^8</td>
</tr>
<tr>
<td>2.000000x10^8</td>
<td>2x10^8</td>
</tr>
<tr>
<td>5.000000x10^8</td>
<td>5x10^8</td>
</tr>
<tr>
<td>1.000000x10^9</td>
<td>10^9</td>
</tr>
<tr>
<td>S(N,s)</td>
<td>n^6</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
</tr>
<tr>
<td>1.250000</td>
<td></td>
</tr>
<tr>
<td>1.456111</td>
<td></td>
</tr>
<tr>
<td>1.549768</td>
<td></td>
</tr>
<tr>
<td>1.596663</td>
<td></td>
</tr>
<tr>
<td>1.631332</td>
<td></td>
</tr>
<tr>
<td>1.63383</td>
<td></td>
</tr>
<tr>
<td>1.642932</td>
<td></td>
</tr>
<tr>
<td>1.643501</td>
<td></td>
</tr>
<tr>
<td>1.643927</td>
<td></td>
</tr>
<tr>
<td>1.644427</td>
<td></td>
</tr>
<tr>
<td>1.644727</td>
<td></td>
</tr>
<tr>
<td>1.644827</td>
<td></td>
</tr>
<tr>
<td>1.644967</td>
<td></td>
</tr>
<tr>
<td>1.644922</td>
<td></td>
</tr>
<tr>
<td>1.644925</td>
<td></td>
</tr>
<tr>
<td>1.644926</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
<tr>
<td>1.644927</td>
<td></td>
</tr>
</tbody>
</table>

Table 5 (cont'd)
### Table 5 (cont'd)

| \( s = 3.333333 \), \( k = 0.3 \) | \( s = 5.000000 \), \( k = 0.2 \) |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| \( S(N,s) \) | \( N^s \) | \( N \) | \( S(N,s) \) | \( N^s \) |
| 1.099213 | 1.007937x10 | 2.00 | 1.031250 | 3.200000x10 |
|           |               | 3.00 | 1.035365 | 2.43 x10^3  |
|           |               | 4.00 | 1.036342 | 1.024 x10^3  |
| 1.139414 | 2.137470x10^2 | 5.00 | 1.036662 | 3.125 x10^3  |
|           |               | 6.00 | 1.036790 | 7.776 x10^3  |
|           |               | 7.00 | 1.036850 | 1.6807 x10^3 |
|           |               | 8.00 | 1.036880 | 3.2768 x10^4 |
| 1.145586 | 2.154435x10^3 | 10 | 1.036907 | 1.000000x10^5 |
| 1.146984 | 2.171534x10^4 | 2x10 | 1.036926 | 3.200000x10^6 |
| 1.147310 | 4.660539x10^5 | 5x10^2 | 1.036927 | 3.125000x10^8 |
| 1.147346 | 4.664588x10^6 | 10^3 | 1.036927 | 1.000000x10^10 |
| 1.147353 | 4.678427x10^7 | 2x10^4 | 1.036927 | 3.200000x10^11 |
| 1.147353 | 9.921255x10^9 | 5x10^5 | 1.036927 | 3.125000x10^13 |
| 1.147353 | 3.045510x10^9 | 7x10^7 | 1.036927 | 1.680700x10^14 |
| 1.147353 | 9.99998x10^9 | 10^9 | 1.036927 | 1.000000x10^15 |
| 1.147353 | 1.007937x10^11 | 2x10^11 | 1.036927 | 3.200000x10^16 |
| 1.147353 | 2.137469x10^12 | 5x10^12 | 1.036927 | 3.125000x10^18 |
| 1.147353 | 2.154434x10^13 | 10^14 | 1.036927 | 1.000000x10^20 |
| 1.147353 | 2.171533x10^14 | 2x10^14 | 1.036927 | 3.200000x10^21 |
| 1.147353 | 4.660539x10^15 | 5x10^14 | 1.036927 | 3.125000x10^23 |
| 1.147353 | 4.664588x10^16 | 10^15 | 1.036927 | 1.000000x10^25 |
| 1.147353 | 4.678427x10^17 | 2x10^17 | 1.036927 | 3.200000x10^26 |
| 1.147353 | 9.921253x10^18 | 5x10^17 | 1.036927 | 3.125000x10^28 |
| 1.147353 | 9.99997x10^19 | 10^19 | 1.036927 | 1.000000x10^30 |
| 1.147353 | 1.007937x10^21 | 2x10^21 | 1.036927 | 3.200000x10^31 |
| 1.147353 | 2.137469x10^22 | 5x10^22 | 1.036927 | 3.125000x10^33 |
| 1.147353 | 2.154434x10^23 | 10^23 | 1.036927 | 1.000000x10^35 |

| \( s = 10 \), \( k = 0.1 \) |
|-----------------|-----------------|
| \( S(N,s) \) | \( N^s \) | \( N \) |
| 1.000977 | 1.024 x10^3 | 2.00 |
| 1.000993 | 5.9049 x10^4 | 3.00 |
| 1.000994 | 1.048576x10^6 | 4.00 |
| 1.000995 | 9.765625x10^6 | 5.00 |
| 1.000995 | 6.046619x10^7 | 6.00 |
| 1.000995 | 2.824752x10^8 | 7.00 |
| 1.000995 | 1.073742x10^9 | 8.00 |
A plot of Eq. (127) is given in Fig. 27. This can only be used for a gross graphical solution of the equation. Somewhat finer solutions can be obtained by reploting in the specific area desired, to a larger coordinate scale. Since the lines over most of both Figs. 26 and 27 are very nearly straight, fairly good interpolations can be obtained on a logarithmic basis.

A plot of \( k \) as a function of \( T \) and \( N \) is shown in Fig. 28 for three specific values of \( T \). The central value is chosen as \( T = 60,000 \) bits, the same as used in illustrative cases in Sections III and IV. For a 1000 bit/sec system it corresponds to a test of one-
minute duration. The two other values of $T$ were selected above and below this figure, to give a reasonable range. Again, the scale of $k$ is very coarse, and the plot can be used only for rough approximations. (The scale used is a reciprocal square root plotted on K & E paper No. D-180.)

**MAXIMUM INTER-ERROR INTERVAL VERSUS NUMBER OF ERRORS**

By inverting Eq. (116) one gets

$$u_o = N^{1/k}$$ (128)

From this it is possible to plot the maximum expected inter-error interval $u_o$ as a function of the total number of errors $N$, in various test times $T$ (measured in bit durations). First it is necessary to find $k$ (as in Fig. 28) as a function of $N$ for the various test times $T$. Then the quantities are merely substituted in Eq. (128). This has been done and the results plotted in Fig. 29.

The curves are very similar to those which have already been plotted for other distributions in Figs. 9, 10, 11 and 17. The most noteworthy difference is that the curves show a much more rounded character than the previous ones. The maximum interval shows a much faster drop, over a narrow range of errors down from the case of all bits in error, than for the other distributions. This means that if one sets a critical expected maximum interval (say to mark the limit of an error chain, as will appear in Section VI), the number of errors in test duration $T$ that corresponds to this will not be too sensitive to the very precise value of the critical interval chosen.
Fig. 29—Maximum expected inter-error interval versus test time and number of errors

COMPUTATION OF ERROR DENSITIES

So far the discussion has been in terms of inter-error intervals, and nothing has been said of the distribution of the short-period error densities during the test $T$. This distribution was computed by Campbell for the Poisson case and illustrated in Fig. 8, and it was also illustrated in Fig. 14 for the simple hyperbolic distribution.

It can be computed from the distribution of inter-error intervals by the general technique that was described in developing Eqs. (87) to (92) in Section IV.

From Eq. (124) above, the cumulative value $T_n$ of the sum of the $n$th longest intervals in the test is given by

$$T_n = N^{1/k} \left[ (1/1)^{1/k} + (1/2)^{1/k} \ldots + (1/n)^{1/k} \right]$$  (129)
\[ T_n = N^{1/k} S(n, 1/k) \quad (130) \]

The ratio of \( T_n \) to \( T \) is the fraction of the total test time that the inter-error interval is at least as long as the \( n \)th in this sequence. That is the cumulative probability of this interval duration, or

\[ R_1(N, u_n) = \frac{T_n}{T} \quad (131) \]

Since there is one error corresponding to each inter-error interval (except at the ends, on which comment was made in Section III), the error rate over the interval is the reciprocal of its duration. That is

\[ c_n = 1/u_n = (n/N)^{1/k} \quad (132) \]

where \( c_n \) is the error rate during the \( n \)th longest interval.

The cumulative probability of the error rate being no greater than \( c_n \) is then

\[ R(a,c_n) = \frac{T_n}{T} \quad (133) \]

Here \( a \) means the average error rate during the whole test \( T \), and

\[ a = \frac{N}{T} \quad (134) \]

It is usual to cumulate this probability in the reverse direction, i.e., to consider \( P(a, c_n) \) as the cumulative probability of the error rate being at least as great as \( c_n \). Then

\[ P(a, c_n) = 1 - R(a, c_n) = 1 - \left( \frac{T_n}{T} \right) \quad (135) \]

or

\[ P(a, c_n) = \frac{T - T_n}{T} \quad (136) \]

From Eqs. (127) and (130), one gets

\[ P(a, c_n) = \frac{(S(n, 1/k) - S(n, 1/k))}{S(N, 1/k)} \quad (137) \]

The computations have been carried through for the central case of Fig. 28. The results have been plotted as the solid lines in Fig. 30.

*These computations were provided by C. R. Lindholm of The RAND Corporation.
There are some granularity problems in these computations. The solution of Eq. (127) for fixed $T$ and $k$, in general comes out as a non-integer for $N$. However, for computing $P(a,c_n)$ in Eq. (137) the nearest integer has been used. Similarly, the integer values for $c_n$ should be plotted in Fig. 30, but instead the curves have for simplicity been plotted as continuous. The result is that the curves involve some approximation, particularly in the regions where $N$ and $c_n$ are small. This does not influence their general trend.

The calculations for the curves were carried out for a range of values of $k$. Since this quantity is a function of the total number of errors $N$, it is also a function of the average error rate $a$ (in errors per bit) during the test. The correspondence is shown in Table 6.
Table 6
PARAMETERS IN 1-MINUTE TEST INTERVAL

<table>
<thead>
<tr>
<th>k</th>
<th>N</th>
<th>a = N/T</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>3</td>
<td>5.0 x10^{-5}</td>
</tr>
<tr>
<td>.2</td>
<td>9</td>
<td>1.5 x10^{-4}</td>
</tr>
<tr>
<td>.3</td>
<td>26</td>
<td>4.33 x10^{-4}</td>
</tr>
<tr>
<td>.4</td>
<td>73</td>
<td>1.217 x10^{-3}</td>
</tr>
<tr>
<td>.5</td>
<td>191</td>
<td>3.18 x10^{-3}</td>
</tr>
<tr>
<td>.7</td>
<td>1,069</td>
<td>1.782 x10^{-2}</td>
</tr>
<tr>
<td>1</td>
<td>6,421</td>
<td>.1070</td>
</tr>
<tr>
<td>2</td>
<td>30,130</td>
<td>.502</td>
</tr>
<tr>
<td>5</td>
<td>48,011</td>
<td>.800</td>
</tr>
<tr>
<td>7</td>
<td>51,439</td>
<td>.857</td>
</tr>
<tr>
<td>10</td>
<td>54,010</td>
<td>.900</td>
</tr>
</tbody>
</table>

There is also a dotted line shown in Fig. 30 to illustrate the typical trend of a distribution of the simple hyperbolic type, as described by Eq. (80).

The comparison between the solid and the dotted lines brings out the following points:

1. The solid Pareto curves show the realistic saturation to \( c = 1 \). This therefore makes them utilizable for describing high error-density short-period distributions. It has been observed in Section IV that the simple hyperbolic law breaks down in such cases because it leads to a plurality (sometimes large) of errors per bit—which is obviously unrealistic.

2. The general slope of the solid Pareto curves as they recede from the saturation point varies rapidly from a low value for high average error rates (and values of \( k \) from 2 to 10) to a fairly steep
value, well beyond the -1 of the simple hyperbolic dotted line, for
the lower average error rates (and values of k smaller than 1).
The actual values of slope approached are of the order of -2.

3. All of the curves plotted correspond to an average error rate
of well over the 1 error per $10^5$ bits that conventionally indicates
a data system of fairly good quality. Of course, to consider systems
of this kind or better it would be necessary to take tests of con-
siderably over one minute's time if a reasonably large sample of
ersors were to be included. In other words, the one-minute period
selected for the calculations would represent one of the poorer than
average performance periods of a system of reasonably good quality.

4. One might wonder if the Mandelbrot model could not be used
for long-time error distributions as well as for the short-time distri-
butions. As just noted above, the more interesting long-time distribu-
tions would generally be for an average error rate significantly
lower than the lowest one of Fig. 30. The general slope which has
been observed on long-time distributions (as, for example, plotted in
Fig. 19) is often a little steeper than the strict -1 of the simple
hyperbolic formula, and usually runs from -1.0 to -1.2. (However,
it is noted that an extreme case in Fig. 19, for the Hawaiian cable
tests, went to an average slope of -1.5 to -1.6.) Thus, the problem
would be, with the lower average error rate, to avoid too steep a
slope. Berger and Mandelbrot themselves say:

"By comparing successive very long stretches of data concerning
errors (that is, stretches of several hours duration), we found further
effects that one model seems unable to explain. Therefore, even if our
model represents fully the data on shorter sequences of errors, it may turn out to be necessary after all to assume that the fundamental parameters vary in time."

5. The net result therefore confirms the finding, already noted in Section IV, that the error distribution seems to change from one form at short-time periods to a different one at long-time periods.

COMPARISONS WITH EXPERIENCE

The Berger and Mandelbrot paper (33) cites experimental data obtained by W. Hoffman of the IBM German Laboratories, together with the German Postal Administration. This was communicated to CCITT by the Federal Republic of Germany. (34)

Four dialed connections were set up (IV, V, VI and VII—the tests I to III are not mentioned), each at four different signal transmitting levels (output to the line). Each connection was tested over 15-minute periods for a total of 19 hours. The system was operated at 1200 bits/sec so that this gave a duration of 2.05 x 10^7 bits transmitted, per level; or a total of 8.2 x 10^7 bits per connection. For some reason it appears that the results were discarded for 15-minute tests in which 5 or fewer bit errors occurred. Fortunately, this hardly affects the results, but a remark on it appears later.

Average error data, and inter-error interval data, are given in the paper. Unfortunately these two are not exactly for the same tests. The paper says:

"The degree of variation (in error rate) is indicated in the published results of the phase modulation tests which are repeated in Table 1 [our Fig. 31]."
"The similar results for the frequency modulation data that we have used (for the interval analysis) exhibit an even larger variation from level to level and between connections. It is, therefore, of considerable interest to find that the same empirical distribution as obtained from the total data fits equally well as a first approximation, independent of the transmit level and connection, and even for the limited samples of a single 15-minute test."

We can hope that this can also hold for the comparisons to be made here.

In Fig. 31 the various error records are plotted for the several tests at the different levels, together with an average performance.

![Graph showing error records](image_url)

**Fig. 31**—Error records, German IBM tests
The solid lines are plotted in Fig. 32 to show the various inter-error interval distributions (the dotted lines are to be discussed later). These represent consolidations of the four tests, at each of four levels. The lower case letters correlate the levels with Fig. 31.

![Graph showing inter-error distributions compared with theoretical curves](image)

*Fig. 32—Inter-error distributions compared with theoretical curves*

If the two figures represented the same tests, one could then check Eq. (127) which is solved for the slope \( k \) in Fig. 28, by estimating the slopes for the four solid curves in Fig. 32, and comparing them with the slopes as indicated in Fig. 28. This has been done. The slopes have been estimated from Fig. 32 (giving little weight to the more sudden drop appearing for \( u \) intervals greater than \( 10^4 \) bits, for reasons to appear shortly). The test duration \( T \) is \( 8.2 \times 10^7 \) bits.
The results for the experimental values are plotted in Fig. 33 as the dots joined by solid lines. The computed, or "theoretical" values appear as the dotted lines. Considering all of the handicaps involved in the comparison, the correspondence is not bad. The general trend of the slopes is roughly correct, though the absolute values are somewhat too low.

![Graph showing theoretical and experimental slopes versus average error densities.](image)

Fig. 33—Slopes versus average error densities (experiment versus theoretical computations)

A comparison may also be made, on the experimental curves of Fig. 32, between the Mandelbrot model and the Gilbert model of Section III. In this last the probability \( Q(a, u) \) of an inter-error spacing \( u \) for a long-term average error rate \( a \) is

\[
Q(a, u) = A^{u-1} + (1 - A) L^{u-1}
\]  

(138)
Here A, J and L are secondary parameters which depend on the primary parameters

\[ P = \text{probability of transition from state } G \text{ to } B \]
\[ p = \text{probability of transition from state } B \text{ to } G \]
\[ 1-h = \text{probability of error in state } B \]
\[ 0 = \text{probability of error in state } G \]

The primary parameters are related to the secondary parameters as follows:

\[ h = \frac{LJ}{J - A(J - L)} \] (139)

\[ P = \frac{(1 - L)(1 - J)}{1 - h} \] (140)

\[ p = A(J - L) + (1 - J) \left[ \frac{L - h}{1 - h} \right] \] (141)

\[ a = (1 - h) \cdot \frac{P}{p + P} \] (142)

It will be noticed that these are all in the inverse form to that desired. Thus, it is necessary to compute the secondary from the primary parameters by processes of successive approximation. It is first necessary, however, to establish "typical" primary parameters. Two more or less extreme cases were assumed, and the results appear in Table 7. Some further discussion of the assumptions will appear in Section VI.

The results of the computations, for the two values of error incidence, have been plotted as the dotted lines in Fig. 32. They
Table 7

PARAMETERS OF ILLUSTRATIVE GILBERT MODEL

<table>
<thead>
<tr>
<th>a</th>
<th>L</th>
<th>J</th>
<th>A</th>
<th>h</th>
<th>P</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.08\times10^{-4}</td>
<td>.340</td>
<td>1.2.5\times10^{-5}</td>
<td>.229</td>
<td>.401</td>
<td>2.75\times10^{-5}</td>
<td>.153</td>
</tr>
<tr>
<td>1.08\times10^{-1}</td>
<td>.337</td>
<td>.970</td>
<td>.243</td>
<td>.401</td>
<td>.0332</td>
<td>.151</td>
</tr>
</tbody>
</table>

are for

x: \ a = 10.8 \times 10^{-5} \ bit \ errors/bit

y: \ a = 10,800. \times 10^{-5} \ bit \ errors/bit

The shape of the characteristic for the Gilbert model is distinctive. It consists first in a rapid drop of the $L^{u-1}$ term, when the characteristic reaches a plateau. This is followed by a slower drop in the $J^{u-1}$ term. The proportions depend entirely on the relative values of the parameters.

The Mandelbrot approximation is a straight line. Only one for the average slope is shown, in dot and dash line, marked $z$. Berger and Mandelbrot suggest that the downward turn of the experimental curves, toward very large $u$, is possibly due to the discarding of the data for very low error incidence. This would certainly have such an effect. However, it is also possible that the distribution law changes, as already hinted. For both reasons this downward turn has generally been ignored in estimating slopes.

The Mandelbrot approximation is much better than the Gilbert model; however, one might further modify the parameters. Also, the original first-order hyperbolic approximation could not be made to
fit for the short values of inter-error spacing considered here.

While the Gilbert model does not fit very well, it certainly does show some of the swing of the experimental data. In a rough way it does follow the correct trend. It can be seen how Gilbert was able to make it fit some experimental data reasonably well, so long as a short period test was involved.\(^{(23)}\)

There have been further comparisons of an elaboration of the Mandelbrot model, to be discussed now.

**CORRELATIONS BETWEEN SUCCESSIVE ERROR GAPS**

In addition to the data which have been discussed above, Berger and Mandelbrot\(^{(33)}\) also have analyzed to some extent the correlations between successive inter-error intervals. These involve the short-time distribution, with the interval involved never going above some $10^4$ bits.

They have, for example, compared the distributions of distance from an error to the next error for varying distances to the last previous error, the variation ranging from 1 bit to 2000 bits. They have also studied the distributions from one to the next error for groups of as many as three last preceding errors, spaced in various ways—but in such cases, individually spaced only up to 2 or 3 bits apart. The correlations do not seem striking, and the reader must be referred to the original paper for a complete report.

In part using these correlations, Sussman\(^{(38)}\) has derived a number of internal relationships that can be expected in error distributions if the Mandelbrot model is followed. Among these are the performances of some error-correcting codes and error occurrences using
certain block lengths of codes. Also involved are sums of two or three, or more, successive inter-error spacings; and the probability that, starting from any arbitrary instant of time, a given number of successive bits are received error-free, which he terms "probability of error-free transmission." The mathematics required are fairly sophisticated, and again the reader must be referred to the original paper for a complete report.

Sussman then checks certain of these deductions with some of the experimental data of Alexander, Cryb and Nast\(^{(17)}\) and of Fontaine and Gallager.\(^{(39)}\)

For the Alexander, Cryb and Nast data, he finds:

<table>
<thead>
<tr>
<th>Original plot, slope ( k )</th>
<th>( k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of error-free transmission,</td>
<td>0.115</td>
</tr>
<tr>
<td>no error correction, ( 1 - k = 0.88 )</td>
<td>0.12</td>
</tr>
<tr>
<td>Same, single error correction, ( 1 - k = 0.89 )</td>
<td>0.11</td>
</tr>
<tr>
<td>Ratio of effectiveness, single to no error correction, 10-bit block ( 1 - 10^{-k} = 0.24 )</td>
<td>0.115</td>
</tr>
</tbody>
</table>

The agreement among the different values of \( k \), as obtained directly from the slope or computed indirectly, appears as a confirmation of Sussman's deductions.

For the Fontaine and Gallager data, he finds:
Original plot, $k = 0.25$

Ratio of error probabilities for two block lengths, one double other, \( 2^{1-k} = 1.68 \)
\( 4^{1-k} = 2.79 \)
Same, one four times other, \( P(1,u) = 0.21 \)
Fraction of code blocks with exactly one error, \( P(2,u) = 0.17 \)
Same, with exactly two errors, \( P(2,u) = 0.30 \)

In this case, the agreement is not quite as good, particularly for the last item. However, it is not really bad. Sussman's final conclusion is:

"The application of the Pareto distribution to the error statistics on the telephone channel has been highly successful. The distribution's single parameter $k$ is sufficient to predict the measured statistics with surprising accuracy."
VI. ERROR BURSTS AND CHAINS

FIXED DURATION BURSTS

It was discovered early in data transmission experience that errors tended to group in clusters or bursts. This was generally indicated only qualitatively, but in 1960 a very simple model was proposed (12) to describe some observations of error occurrences in several Lincoln Laboratory tests (24, 25).

In this model the errors come during relatively short error-sensitive periods, and not at all outside these periods. In the very simple form of the model, the periods are all of the same duration and come irregularly. Within the periods the errors have a certain probability of occurrence and follow a Poisson distribution in time. Because this model has been superseded, it will not be analyzed here in detail, but the reader can find it in the original paper.

It is interesting, however, to note what could be done with the model. In the Lincoln Laboratory tests it was found, first, that the average number of bit errors per word in error was greater than would be expected from a completely random distribution. Second, the ratio of cases of two successive words in error to single words in error was also larger than would be expected.

(Note: A "word," according to the IRE Dictionary (40) is "an ordered set of characters which is the normal unit in which information may be stored, transmitted, or operated upon within a computer." In these tests the words were of 16 bits for the cases studied.)
It was found that the use of the model led to "possible reasonable average numbers of errors per word in error, and also to possible reasonable frequencies of occurrence of two consecutive words in error, as compared with single words in error." Here, "possible reasonable" means "consistent with the experimental tests."

In the complete model the bursts were distributed in time according to the simple hyperbolic distribution which was described in Section IV.

**VARIABLE DURATION BURSTS**

It is obvious that the above model was too simple to go very far in describing error occurrences. A closer examination of the errors in the Lincoln Laboratory tests. The author is indebted to Mr. E. J. Hofmann of Lincoln Laboratory for the opportunity to study the basic data on the tests.

From the records the conditional probability of errors was determined for the bits following the initial error of a burst. The cumulated results are plotted in Fig. 34. They show a conditional probability which has a large drop after the first bit (which is necessarily 1); after that it diminishes more or less steadily to 20 or 25 bits, when it becomes very small and irregular. The plotted value W indicates the word length of 16 bits for the particular case.

A simple model to fit this is indicated in Fig. 35. The first point to note is that it consists of two parts, to allow for the high proportion of one-bit bursts observed. The abscissa w is
the number of bits following the initial error, and $p(w)$ is the conditional probability density of succeeding errors. $V$ indicates the maximum burst duration in this model.

The fit of the model is indicated in the cumulative probability plot of Fig. 36. Here the parameter $K$ measures the ratio of the bursts in the continuous distribution (i.e., not the one-bit bursts) to total bursts. It is seen that a fair fit is obtained for $K = 0.75$. 

Fig. 36—Model and experimental distributions of error burst durations
The results as shown in the figure.

For the equation of the graph was plotted.

On the x-axis, the error was closed, this was the result of the graph that was plotted.

Subsequently, the graph was read, to read the number of errors within the graph. After that, the graph was plotted.

The result is 600/600 and that is the error of 1.200/sec. This is possible to show in the figure.

A second system was observed, and this was possible to show in the figure.

A second system of errors was expressed in percentage with the error of 190/1200.

The results of the first system was 190/1200.

To section 1. Indeed, implies the graph as also was expressed in the figure and the results of the first system was expressed in the figure.

In addition, also was expressed in the figure and the results of the first system was expressed in the figure.

This implies that the graph as also was expressed in the figure and the results of the first system was expressed in the figure.

With this modified, the graph as also was expressed in the figure and the results of the first system was expressed in the figure.

where K = 0.5 and L2 = 1.2. Is there any possible to get numbers of errors within the same.

The averages for the second curves are shown with the median.
Fig. 37—Density of errors within gate intervals (Bell System data—all calls in test, 600 bits/sec)

In spite of the difficulty of direct comparison, the curves reflect much the same characteristics as for the Lincoln Laboratory data. Principally, they suggest a somewhat higher proportion of 1-bit bursts. This is roughly indicated by the proportion of 20-bit gate openings that show only one bit (the triggering bit) in error.

This general type of burst structure for the short-period error occurrence, together with the burst initiations distributed according to the simple hyperbolic law outlined in Section IV, form a fairly realistic description of actual error occurrences (21) which does, nevertheless have some weaknesses. One of these is the occasional overlapping of error bursts, which is apt to be ignored; and the other is the saturation during short periods of high error density, which
is not well described (as noted in Sections IV and V). These will be covered further below.

**EXPERIMENTAL DEFINITION OF BURSTS**

When one is confronted with a succession of errors, the individual bursts are not always well resolved. In order to treat them statistically it is necessary to adopt some rule by which to distinguish whether a given gap between errors is merely a spacing within a burst, or whether it is really a separation between two bursts. The rule adopted in Ref. 41 was that groups of errors were called different bursts if separated by 5 or more correct bits. An essentially similar definition has been used by Fennick, (10) except that his critical gap is 0.6 millisecond.

**GILBERT MODEL OF BURSTS**

This (23) is a refinement of the variable duration bursts, and the essentials of the model have already been described in Sections III and V. The parameters have been listed in Eqs. (138) to (142). The "primary" parameters $P$, $p$ and $h$, basically describe the structure and frequency of occurrence of the bursts. "Secondary" parameters $A$, $J$ and $L$, are used to build up Eq. (138), which describes inter-error spacings. In an inverse manner, the primary parameters are related to the secondary parameters via three equations, Eqs. (139-141). A fourth equation, (142), gives the long-term average error rate $a$ in terms of the primary parameters.

The fundamental problem has been to establish the parameters for the cases to be illustrated. A basic assumption which has been
made is that the burst structure statistics are relatively constant over the range of conditions assumed, especially over the range of burst frequencies of occurrence that are covered. This was generally indicated, at least as a first-order approximation, in Ref. 41. The structure to be assumed is that shown in Fig. 34. It will be used to compute the parameters \( p \) and \( h \), which then stay constant throughout the remainder of the discussion. These parameters are:

\[
p = \text{probability of ending burst period}
\]
\[
h = \text{probability of no error during burst period, or}
\]
\[
l - h = \text{probability of error during burst period.}
\]

The conditional probability \( E(K) \) of error, after the first error, for the \( K \)th succeeding bit, is

\[
E(K) = (1 - h)(1 - p)^K
\]

One can now take two key points in Fig. 34, and with these solve for \( h \) and \( p \); then adjust them so that the curve best fits the overall trend of the experimental data. This has been done, and the result plotted as a fine dotted line in Fig. 38, with the parameter values

\[
h = 0.600
\]
\[
p = 0.150
\]

It is seen that the fit is fairly good. It must be clear, of course, that in the experimental data the burst always starts with an
error, which is not necessary in the Gilbert model. This distinction will be ignored for the present study.

Values of $P_0$, the probability of initiating a burst period, which therefore controls the frequency of occurrence of bursts, have then been assumed ranging from around $10^{-3}$ to $2/3$, to establish a gamut of the order of what might be encountered in practice.

The second step in the basic problem is to determine the values of the secondary parameters $A$, $J$ and $L$ that will yield the assumed primary parameters $h$, $p$ and $P$ from the Eqs. (139-141). This is a problem of inversion of simultaneous equations. Rather than attempt to solve it formally, successive approximations have been used.
The approach is simplified by noting a few approximations that hold when \( P \ll 1 \), which is true for most of its range. The first of these is that \( J \) is very nearly 1. Although it is not immediately obvious, it is obtained from Eq. (140) once some substitutions are tried. The equation is reproduced here for convenience.

\[
P = \frac{(1 - L)(1 - J)}{1 - h}
\]  

(145)

The remaining approximations then follow fairly quickly:

\[
L \approx h
\]

(146)

\[
1-J \approx P
\]

(147)

\[
A \approx p/(1 - h)
\]

(148)

One can start with the lowest values of \( P \) and work back with the exact equations to suggest successive approximations. The values of \( P \) nearer to 1 then require more approximations to achieve exactness in the basic parameters. The final values achieved were already listed in Tables 3 and 6, and it is to be observed that they did not always reach perfect accuracy. (Note: It will be seen that sometimes \( a \) and sometimes \( \alpha \) are used to designate bit error rate. This is because occasionally \( a \) has been reserved for burst initiation rate. In the present context \( a \) refers to bit error rate, and it is the same as the \( \alpha \) of Table 3.)
With these parameters the cumulative probability \( Q(a,u) \) of inter-error spacing \( u \) was given in Eq. (138), and it is reproduced here for convenience.

\[
Q(a,u) = Aj^{u-1} + (1 - A) L^{u-1}
\]

(149)

This was compared in Fig. 32 with some experimental data and with the Mandelbrot model. It gave a fair fit with the data, though it was not as good as the Mandelbrot model.

So far the comparisons have been made on a relatively short-time basis that shows the burst structure and its spacings from immediate neighbors. The Gilbert model also purports to describe long-time distributions of bursts, say over days and months. In this aspect, however, it is not realistic, as will be shown below.

**ERROR BURST CHAINS**

In 1960, when studying magnetic tape records of impulse noise in cable, Pfeiffer and Yudkin\(^{(29)}\) discovered that there was a significant tendency for noise bursts to chain together in time, over substantial periods.

In the same year Dimock\(^{(42)}\) studied the outage times of data transmission circuits over a variety of long-haul Bell System communication facilities. These included Type K carrier on multi-pair cables, Type L carrier on coaxial cables, and TD-2 microwave relay. The results for the three were so similar that he summarized all of them on one diagram, which has been re-plotted in Fig. 39 herewith. The frequency of occurrence is normalized to outages
per 100 hours per 1000 miles (Dimock uses the term "hits" instead of "outages"). The two curves marked "Dimock" show the upper and lower limits respectively of his summary plot. The curve marked "Gilbert" will be discussed later.

DEFINITION OF ERROR CHAINS

An error burst chain (or simply an error chain if it is found that the intermediate clustering structure is not necessary) consists of a group of error bursts, or errors, that succeed each other closely in time. The same problem exists as in the definition of "burst"—namely, how is "closely" to be interpreted?

This has been resolved in the same way as for "burst." That is, when the number of successive good bits in a given stretch
exceeds a critical value, this is considered to cut the chain in
two at this point. This definition was also used by Pfeiffer and
Yudkin. (29) There is a difference with the case of bursts, however,
in that the range of lengths can be very much greater. This argues
for the possibility of a variable critical gap, according to the
duration of the immediately preceding chain. Both fixed and
variable critical gaps have been used, but in this present summary
only the variable gap will be considered.

ANALYSIS OF CHAINS

The Gilbert burst model is not of finite maximum length like
the triangular model of Fig. 35 and thus lends itself to the
possibility of describing an extended chain of errors. The
probability of such an extended chain is given by the geometrical
distribution formula of Eq. (143). When \( K \) is large, this
becomes an exponential (see Section III) and thus diminishes very
fast with \( K \). The trend of such a function is plotted as the curve
marked "Gilbert" in Fig. 39. This shows the probability of extended
error chains to be insignificant compared with the occurrence
experimentally found.

The occurrence of bursts in quick succession may cause them
to appear to merge, raising the probability of longer-duration
chains. However, the probabilities for these events will also be
of the geometric or exponential form, and thus show the same trend
as for the single bursts. Hence, they cannot describe extended
chains realistically.
An attack was made\(^{(22)}\) on this problem in 1962, using Poisson statistics for short-time burst distribution and simple hyperbolic statistics for the long-time distributions. However, an error was discovered in the use of some of the Poisson formulas. Since that time new light has been thrown on short-time error distributions, so that instead of simply making a correction, a new approach is made, using the Mandelbrot model for the short-time periods and the simple hyperbolic model for the long-time periods. It will be found, incidentally, that since the latter are those really critical to the theory, the new results come out much the same as in the 1962 study.

NEW CHAIN CALCULATIONS

The broad objective is to find out how many error chains of various given lengths may be expected in a long overall test under different conditions. For the example chosen the longest overall test is taken as lasting \(T' = 10^{10}\) bits. At 1000 bits/sec this is 2778 hours or 116 days. Fractions of this test are also taken as follows:

<table>
<thead>
<tr>
<th>(T'), bits</th>
<th>hours</th>
<th>days</th>
<th>(a) (per (10^5) bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10^{10})</td>
<td>2778</td>
<td>116</td>
<td>1.77</td>
</tr>
<tr>
<td>(10^9)</td>
<td>278</td>
<td>11.6</td>
<td>1.39</td>
</tr>
<tr>
<td>(10^8)</td>
<td>27.8</td>
<td>1.16</td>
<td>1.00</td>
</tr>
<tr>
<td>(10^7)</td>
<td>2.78</td>
<td>0.12</td>
<td>0.614</td>
</tr>
</tbody>
</table>
The average assumed error incidence a diminishes somewhat as the long-time tests are shortened as shown in the last column. This is a consequence of the simple hyperbolic law as discussed in Section IV. The average assumed, for any of the test durations, represents that generally obtainable over a current fairly high-grade data transmission facility.

The technique used for the computation consists in subdividing the test duration into a number of short-time intervals. A fraction of those intervals can be expected to contain error chains over their whole lengths. This fraction, when determined and multiplied by the number of intervals in the long-time test, gives the number of error chains, of at least the short-time interval length, that can be expected in the long-time test.

The criterion as to when the short-time interval contains an error chain over its whole length (see Fig. 40) is that the longest

![Diagram]

Fig. 40 — Potential error chains and error-free gaps
error-free gap in it should be less than the critical assumed value. The correlation of such a maximum error-free gap \( u_0 \) with the number of errors \( N \) in the short-time interval \( T \) was computed in Section V for the Mandelbrot model. This is

\[
u_0 = N^{1/k}
\]  

(150)

A decision must now be reached as to what will be set as a critical gap. The Mandelbrot model offers the possibility of a very simple choice. This is the gap for which \( k = 1 \), or

\[
u_0 = N
\]  

(151)

The short-time interval for which this occurs is (following Eq. (127) which is reproduced here):

\[
T = N \delta(N, 1)
\]  

(152)

A plot of what this means for the illustration which is being presented is shown in Fig. 41, and the specific numbers are listed in Table 8. As was discussed in connection with Fig. 29, in the Mandelbrot model the results are not too sensitive to the precise assumptions made for the critical gap.

In the practical situation, because of the way the equations appear, it is convenient to decide first on \( N \), then on \( u_0 \), and finally on \( T \). A variety of values are shown in Fig. 41 and Table 8. The formulas give results for \( N \) very low, such as equal to 2, but for reasons which have already been noted several times these are not to be taken too seriously.
Table 8

ILLUSTRATIVE CHAIN LENGTHS T

<table>
<thead>
<tr>
<th>N</th>
<th>bits</th>
<th>seconds</th>
<th>minutes</th>
<th>hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.14×10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2.93×10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$10^2$</td>
<td>5.19×10²</td>
<td>5.19×10⁻¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$10^3$</td>
<td>7.49×10³</td>
<td>7.49</td>
<td>0.125</td>
<td></td>
</tr>
<tr>
<td>$10^4$</td>
<td>9.79×10⁴</td>
<td>9.79×10</td>
<td>1.63</td>
<td></td>
</tr>
<tr>
<td>$10^5$</td>
<td>1.21×10⁶</td>
<td>1.21×10³</td>
<td>2.01×10</td>
<td>0.335</td>
</tr>
<tr>
<td>$10^6$</td>
<td>1.44×10⁷</td>
<td>1.44×10⁴</td>
<td>2.40×10²</td>
<td>4.00</td>
</tr>
</tbody>
</table>
Fig. 41—Assumed critical gaps for error chains of varying lengths.

The next step in the process consists in finding out the probability that the short-time interval $T$ contains a chain over its whole length. This is obviously the probability that it contains at least $N$ errors, or that it has an error density of at least $c$, where

$$c = \frac{N}{T}$$

The probability measures the fraction of the intervals $T$ in the test time $T'$ that shows an error density $c$ where the average density over $T'$ is $a$. For this figure, over the long-time period, we use the simple hyperbolic law of Eq. (81), or

$$P(a, c) = \frac{a}{(Ac + a)}$$

where

$$A = A(k') = (\log_e k') - 1 + (1/k')$$
Here $k'$ means the number of intervals $T$ in $T'$. The prime is used to distinguish the quantity from the Pareto slope $-k$.

The computations have been carried out and the results are listed in Table 9, and plotted as the solid lines in Fig. 42.

*Fig. 42—Expected numbers of error chains in illustrative tests*

(The various values of $a$ for corresponding values of $T'$, that were listed above, were computed using the $A(k')$ values for $T = 9.79 \times 10^4$ bits.)

There is a scale of seconds and minutes at the top of Fig. 42, for a bit rate of 1000/sec.)
<table>
<thead>
<tr>
<th>$k'$</th>
<th>$A(k')$</th>
<th>$k'P(a,c)$</th>
<th>$N$</th>
<th>$k'$</th>
<th>$A(k')$</th>
<th>$k'P(a,c)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$3.33 \times 10^6$</td>
<td>14.0</td>
<td>2.19</td>
<td>2</td>
<td>$3.33 \times 10^7$</td>
<td>16.3</td>
<td>3.05x10</td>
</tr>
<tr>
<td>$8.76 \times 10^5$</td>
<td>12.7</td>
<td>9.66x10^-1</td>
<td>5</td>
<td>$8.76 \times 10^6$</td>
<td>15.0</td>
<td>1.31x10</td>
</tr>
<tr>
<td>$3.42 \times 10^5$</td>
<td>11.7</td>
<td>5.22x10^-1</td>
<td>10</td>
<td>$3.42 \times 10^6$</td>
<td>14.0</td>
<td>7.12</td>
</tr>
<tr>
<td>$1.93 \times 10^4$</td>
<td>8.87</td>
<td>6.91x10^-2</td>
<td>10^2</td>
<td>$1.93 \times 10^5$</td>
<td>11.2</td>
<td>8.95x10^-1</td>
</tr>
<tr>
<td>$1.34 \times 10^3$</td>
<td>6.20</td>
<td>9.90x10^-3</td>
<td>10^3</td>
<td>$1.34 \times 10^4$</td>
<td>8.50</td>
<td>1.78x10^-1</td>
</tr>
<tr>
<td>$1.02 \times 10^2$</td>
<td>3.64</td>
<td>1.68x10^-3</td>
<td>10^4</td>
<td>$1.02 \times 10^3$</td>
<td>5.93</td>
<td>1.69x10^-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10^5</td>
<td>8.27x10</td>
<td>3.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10^6</td>
<td>6.95</td>
<td>1.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$T' = 10^9$ bits</th>
<th>$T' = 10^{10}$ bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k'$</td>
<td>$A(k')$</td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>$3.33 \times 10^8$</td>
<td>18.6</td>
</tr>
<tr>
<td>$8.76 \times 10^7$</td>
<td>17.3</td>
</tr>
<tr>
<td>$3.42 \times 10^7$</td>
<td>16.3</td>
</tr>
<tr>
<td>$1.93 \times 10^6$</td>
<td>13.5</td>
</tr>
<tr>
<td>$1.34 \times 10^5$</td>
<td>10.8</td>
</tr>
<tr>
<td>$1.02 \times 10^4$</td>
<td>8.23</td>
</tr>
<tr>
<td>$8.27 \times 10^2$</td>
<td>5.72</td>
</tr>
<tr>
<td>$6.95 \times 10$</td>
<td>3.26</td>
</tr>
</tbody>
</table>
COMPARISON WITH EXPERIMENT

The Dimock curves in Fig. 39 have been transferred to Fig. 42 for comparison. This comparison can only be rough, first because the bit rate for Dimock's data was 750 instead of 1000/sec, and second because he did not define error chain (which he called "hit") very precisely, and we cannot tell how closely our own definition corresponds.

Nevertheless, the drift of the curves is very similar, and undoubtedly the theory is on the right track and far better than the fit in Fig. 39. The Dimock data are normalized to 100 hours of test, and their center of gravity oscillates roughly about the 100-hour region (between 27.8 and 278 hours) of the figure.
VII. MEASUREMENT OF CIRCUIT PERFORMANCE

INDEX OF PERFORMANCE

One of the most frustrating aspects in the study of data circuit performance lies in the specifications which can be made for its numerical measurement.

Obviously, the most important index of performance is the extent to which the system introduces accidental errors in the transmission. The errors come haphazardly, and it is a serious problem to express quantitatively and in a single number just the extent to which they do occur. If the error incidence were completely random, then conventional statistical laws can be used to give an exact specification. However, it has been found above that the errors are not completely random, that they do not obey these statistical laws, and also that the departures from the laws can be very serious.

The usual measurement procedure is to determine the average number of errors which occurs in a given period of time. It is recognized that the measurement must cover a long enough period of time to obtain a reproducible result. With a conventional Poisson distribution, one can estimate the variance in successive measurements of the average error rate, and from this estimate, the size of sample needed to obtain the confidence interval desired. A test long enough for such a sample then gives a correspondingly adequate accuracy in the determination of the performance.

However, it is found in the case of the data errors that
reproducible averages are not readily obtained, even over fairly long tests. This correlates with the fact that neither the simple hyperbolic nor the Pareto law have finite averages over an infinite sample. Thus, neither of these formulas obeys the laws of large numbers. In addition, there is a question as to whether the average, or some other index, is the most expressive quantity to use.

If the difficulty were merely a matter of awkward mathematics, one could easily search for more satisfactory methods. But the fact that actual experience brings out the same problems as the mathematics emphasizes the basic applicability of the latter.

**PEAK ERROR INCIDENCE**

A general notion of the circumstances involved may be glimpsed from Fig. 43. In general if a measurement is repeated, peak values will occasionally be obtained -- specifically, higher values than any of those preceding. These peak values can be expected to grow as the test proceeds. If the quantity measured follows Poisson statistics, the probability \( P(a, c) \) of obtaining a specific value of at least \( c \), when the long-time average is \( a \) (which, for present purposes equals one), was previously given by Eq. (28) and plotted in Fig. 8. The expectation of a given maximum \( c_m \), in a number of observations equal to \( T \), is that value of \( c \) for which

\[
P(a, c_m) = \frac{1}{T}
\]

(156)

The ratio \( c_m/a \) is plotted as the lower solid line in Fig. 43.

For the simple hyperbolic distribution the same probability
Fig. 43—Expectation of ratios of maximum, average and maximum median as functions of test duration (Poisson and simple hyperbolic distributions)

was given in Eq. (81). The result is simple and becomes, by application of Eq. (156):

$$c_m/a = (T - 1)/A(T)$$

(157)

This is plotted as the upper solid line in Fig. 43. Dotted lines indicate ratios of expected maxima to long-time median values instead of averages.

A consideration of the curves in this figure gives a measure of what can be expected while carrying out a test. It is quite true that in a Poisson distribution the expected peak values continually rise as the test progresses. However this rise is
slow, and it gets slower as the test continues. With the simple hyperbolic distribution, however, the rise is practically linear, and stays so indefinitely. In an actual experiment, of course, it would finally saturate to the point where half the bits would be in error (all bits in error would mean an exactly "upset" signal, with marks and spaces interchanged).

COMPARISON WITH EXPERIENCE

Tests with durations as long as those included in the range of Fig. 43 are really very long and not usual. Figure 19 included some fairly long practical tests, and the peak data for these are shown in Fig. 44. To simplify comparison they are indicated with the same kind of lines as in Fig. 19 and are labelled with the same letters. Test H is omitted. In these T is measured in test days (not necessarily 24-hour days), which are the units in which the data were reported. Even the longest of these tests are so short (in the measurement units) that they do not reach to the end of the first decade plotted in the abscissae of Fig. 43.

In spite of their limited duration the tests show distinctly a trend toward the simple hyperbolic expected curve. For the much shorter durations there is a trend toward the Poisson curve. This illustrates the change in distribution between short-time and long-time tests, which has been noted in earlier Sections—though the short-time distribution is now recognized to be Pareto rather than Poisson.
COURSE OF EXPERIMENTAL AVERAGE

One can, as an experiment proceeds, compute the average of all the readings to date, and plot the value as a function of the duration of the experiment. If the total duration is long enough, and if a large enough sample of similar tests is considered, such a plot will give a notion of how long it takes for the average to stabilize to a reasonably reproducible value.

The plot is shown in Fig. 45 for those tests which were
Fig. 45—Cumulated averages of error incidence (Tests of Fig. 19)

included in Fig. 19. The same letters and types of line identify the tests. The curve for test H has been displaced downward by one log cycle to bring it into a clearer part of the plot. The test duration here is $T_1$, measured in hours.

A certain recurrent pattern may be detected in the figure.
As the test proceeds, at some given hour a large burst of errors occurs. This suddenly raises the average. For a number of hours following, the error rate may be near usual or below. The new contributions do not increase the cumulated total much, but the test time grows. Thus, the cumulated average error drops along a hyperbola (which plots as a straight line on the log-log paper). The pattern formed is somewhat that of a ripsaw tooth, with one side steep and the other sloping more gently.

The conclusion reached from Fig. 45 confirms the statement that stabilization of the average does not readily occur. Some traces suggest a possibility of stabilization, but examination of the other traces shows that this is not dependable.

Essentially the same conclusions, in other terms, have been noted elsewhere.\textsuperscript{(43, 44)} It has been observed that in Fig. 19 the range between maximum and minimum error rates reported for each given test runs from a little over one to more than five decades. Enticknap indicates day-to-day variations in average error rate frequently running over two decades. He indicates in a detailed chart, 20-minute averages in error rate over three decades apart, from one 20-minute period to the next (excluding zero error rate cases, which could give infinity decades between the zero and any finite error rate).

\textbf{COURSE OF EXPERIMENTAL MEDIAN}

It is clear that if one wants to attenuate the effect of very large bursts of error on a "typical" or "index" value, one
should use the median rather than the average. Here the advent
during the test of a large value merely moves the index value one
(presumably small) step in the middle of the range.

The cumulated medians for the same tests are plotted in Fig. 46.

Fig. 46—Cumulated medians of error incidence
(Tests of Fig. 19)
(To simplify the computation, the tests involve merely daily periods
of equal durations, i.e., less than 24 hours, and the actual moderate
variations in these are ignored. Thus, the final values in Fig. 46
do not always exactly accord with values which could be scaled off
from Fig. 19.)

At first glance, the instability in Fig. 46 seems to be about
as bad as in Fig. 45. However, if one ignores the early values,
which would be expected to be unstable, one finds that tests D
and E are made somewhat worse, but that tests F and G come
out much better. Thus, the overall stability is somewhat, though not
spectacularly, improved.

OTHER QUANTILE INDICES

From Fig. 19 one gathers that the most interesting and regular
part of the distributions lies toward the rarer large values rather
than the median. These can be the quartile, the decile, centile, or
even further, boundaries.

By application of the simple hyperbolic law of Eq. (80),
namely

\[ P(a, c) = h/(c + h) \]  \hspace{1cm} (158)

it is found that when \( P \) is respectively \( 1/2, 1/10, \) and \( 1/n \), the
following relations hold between \( h \) and \( c \):

\[ c + h = 2h, \quad h = c \]  \hspace{1cm} (159)

\[ c + h = 10h, \quad h = c/9 \]  \hspace{1cm} (160)
\[ c + h = nh, \quad h = \frac{c}{(n - 1)} \]  \hspace{1cm} (161)

The bias \( h \) can thus be computed from any of the quantile boundaries in the tests, and the results are comparable one with another. The principal problem encountered in this process is that the test may not be long enough to secure meaningful boundaries for the higher-order quantiles.

Some exploration of this, up to the decile, was carried out with the tests covered by Fig. 19, but it was necessarily crude because the data were not extensive enough. However, it indicated possibly a little better stability as an index than the median.

Another index which is possible is a composite bias \( h \) obtained by averaging the values obtained from a series of quantile boundaries, starting with the median. The exploration showed this to be slightly more stable than the decile alone or than the median alone.

The detailed results of these explorations will not be reproduced here, as they were much too crude; but they did indicate this to be a promising line of approach in the search for a stable index of significance.

The best index has also been studied for a Pareto distribution by Muniruzzaman.\(^{(36)}\) The Pareto distribution which he considers seems to be rather more complicated than that referred to by Mandelbrot or Gumbel. Also, his style is excessively condensed and not always too lucid. However, he says:

"From a comparison of the variances of the estimate of these
three measures of average it appears that the harmonic mean allows itself to be most accurately estimated... The sample geometric mean, however, estimates its population counterpart with much less precision. The median stands in between the two."

Then, at another place, and more or less as a conclusion, he states:

"The above results reveal the unique position of the geometric mean as a measure of location in a Pareto population. Moreover, the sample geometric mean itself is a 'best' estimate of its population counterpart. The geometric mean is as natural in this distribution as the arithmetic mean in a normal distribution."

In spite of these comments it would seem likely that the composite index which has been described above would turn out the most generally satisfactory. Incidentally, it would not appear that the evasion constituted by expressing the performance result in terms of the reciprocal of the error rate, or mean (or other index) time between errors, really escapes any of the instability problems just discussed.

SEGMENTATION OF A LONG TEST

It is possible to learn something of the characteristics of error distribution measurements in a long test by segmenting the test and comparing the measurements on the segments with each other and with those on the total test.

The first consideration will be theoretical, on the assumption that the errors follow the simple hyperbolic law. For this purpose
one can use straight-line plots on the special hyperbolic probability paper (such as used in Fig. 19, and described in the Appendix).

In Fig. 47 the heavy line marked "Total Test" represents an idealized distribution of error densities in a 30-day test with daily measurements. This is then segmented into five consecutive 6-day tests. It is obvious that each of these tests cannot be described by the same segment of the central line, cut off at the upper end at five times the lowest probability of the consolidated test and at the lower end in an analogous manner.
This is because such a description would leave out, among other test results, the single worst observation in the long test (shown at 29 errors and 3.3 per cent). This is obviously part of one of the segmented tests.

These tests must clearly be described by five parallel lines about the central line, and the highest of these must end at the same worst observation as for the long test (namely, at 29 errors and 16.5 per cent). In an actual test the highest line would represent the worst segment, and half the observations in it would be expected to lie above the line and half below it. The observations above the line would comprise 1/2 x 1/5 or 1/10th of all the observations. Thus, together with a corresponding lowest line, they would constitute 80 per cent confidence limits for the five actual segmented tests. (The lowest line can be constructed in a similar manner from the lower end of the chart.) The intermediate lines can be drawn by the same general procedure for 40 per cent confidence limits.

Experimental plots closely approximating this case can be constructed from the longer tests F and G of Fig. 19. They are shown respectively in Figs. 48 and 49. Because of the comparatively small number of observations, the traces are rather irregular. But, nevertheless, they show considerable resemblance to those of Fig. 47. As noted in Section V, they show overall a slightly steeper slope than Fig. 47.

All this has a certain application to the estimated duration of test which will give a reasonably reliable measure of the error performance of the system. It is clear that, in the plot of Fig. 47,
Fig. 4. — Experimental distributions in segmented test
(F of Fig. 19)
Fig. 49. - Experimental distributions in segmented test (G of Fig. 19)
if the test duration on the system were increased 10 times, the end points on the idealized curve would merely be extended one decade each way; and the confidence intervals on the five new segmented tests (each test also extended one decade) would remain exactly the same if measured logarithmically. This would continue even if the tests were extended more.

This experience is frustrating, because the theoretical confidence intervals cannot be narrowed down (in contrast with the case of conventional measurement techniques) by indefinitely prolonging the test.

It is true that if the test is prolonged to the point where error saturation occurs (as discussed in Sections IV and V), the upper portions of the curves of Fig. 47 will bend over to the left and force a narrowing of the confidence intervals in the saturation region. But this will not increase the reliability of the measurement in the major portion of the curves.

In conclusion, one could say that measurement of error performance essentially breaks down into two parts:

1. What is the quantity to be measured or, at least, eventually reported?

2. How long should the test last?

At the present time firm numerical answers to these questions cannot be given. It can only be said, from the discussions which have been given, first that the average error rate is not too bad if the conditions are specified; and, second, that a short test is quite likely to give a misleading sample of the performance of a data
system, so that tests should be as long as practical. The most significant picture of system performance is given by a distribution plot something like Fig. 19.

MERGING OF INDIVIDUAL MEASUREMENTS

In the case of conventional measurements, it generally makes little difference in the trend of a distribution whether individual data are plotted or whether successive groups of points are averaged together. In the latter case the curves come out smoother, but their center of gravity is not changed, and the major consideration is that enough points for a good curve are left after the merger.

However, it will be found that where the data follow a simple hyperbolic distribution this is no longer true, and the center of gravity of the plotted points varies according to the extent to which data are merged—that is, according to whether the points are taken minute by minute, hour by hour, week by week, etc. In all of this it is assumed, of course, that the numerical error density is specified per some fixed unit of time, regardless of the average measuring interval. That is, that the data for successive points, regardless of whether they come every second, minute, etc., are specified as errors per bit, or per $10^5$ bits, or per some other fixed unit for all the cases.

To understand this we can repeat Eq. (158), namely

$$P(a, c) = h/(c + h)$$ (162)
where
\[ a = hA(k) \]  \hspace{1cm} (163)

\[ a = h\left(\log_e k - 1 + \frac{1}{k}\right) \]  \hspace{1cm} (164)

This can be applied to the total test, for example, of Fig. 47. There are 30 daily test readings so that
\[ k_1 = 30, \ A(k_1) = 2.435 \]  \hspace{1cm} (165)

It is possible to merge the test readings for every five days, so that
\[ c_5 = \left(\frac{1}{5}\right) \sum_{j=1}^{5} c_{1j} \]  \hspace{1cm} (166)

This gives six of the 5-day test readings, and
\[ k_5 = 6, \ A(k_5) = 0.958 \]  \hspace{1cm} (167)

The total number of errors in both descriptions of the same test is the same, so that the rates, say per $10^5$ bits, are the same and
\[ a = \frac{2.435}{h_1} = 0.958 \frac{h_5}{h_3} \]  \hspace{1cm} (168)

or
\[ h_5 = (2.435/0.958) \frac{h_1}{h_1} = 2.54 \frac{h_1}{h_1} \]  \hspace{1cm} (169)

Thus
\[ P_1(a, c) = \frac{h_1}{(c_1 + h_1)}, \quad P_5(a, c) = \frac{h_5}{(c_5 + h_5)} \]  \hspace{1cm} (170)

For a given $P$, say $P = \frac{1}{2}$
\[ \frac{1}{2} = \frac{h}{(c + h)} \]  \hspace{1cm} (171)
\[ c = h \]  \hspace{1cm} (172)

Thus, for \( P = \frac{1}{3} \)

\[ c_5 = 2.54 \, c_1 \]  \hspace{1cm} (173)

The two distributions have been plotted in Fig. 50, with \( c_1 \)

![Graph showing error distributions in long test with original and merged data](image)

Fig. 50 -- Error distributions in long test with original and merged data

illustratively equal to 1, so that \( c_5 = 2.54 \), both at \( P = \frac{1}{3} \). The distribution for the 5-day test readings comes distinctly higher than for the daily test readings.

An experimental check has been made for the same tests \( P \) and
of Figs. 48 and 49, by merging the daily test readings into 5-day test readings. The results are plotted in Figs. 51 and 52, respectively. These show the same general appearance as Fig. 50, and the separations between the two descriptions of the same tests are roughly the same.

The results indicate that when one compares two plots of error distributions which follow, or approximately follow the simple hyperbolic law, it is necessary to make sure that the individual

---

**Fig. 51—Experimental effect of merging data**

(F of Fig. 19)
data points cover the same measuring periods. This holds whether the overall test periods are the same or not.

TIME DOMAIN MEASUREMENTS

At the IEEE convention of 1964 in New York, two papers from the Bell Telephone Laboratories, (45, 46) discussed a problem which has some bearing on data circuit performance rating, though the question is not exactly the one we have been considering. Their problem is the prediction of probable performance of modems over specific telephone lines. The objective is to determine whether a given line is suitable for use, or whether it needs modification; and
the test must be very simple and quickly applied, so that many circuits may be tested by a relatively unskilled operator.

The principles of both test sets are essentially the same and consist in sending periodically a fairly sharp pulse over the circuit. At the receiver the signal is full-wave rectified and averaged in a d'Arsonval galvanometer. The peak is also detected and the ratio of peak to average readings (or the reciprocal) determined. Surprisingly enough, when certain precautions are observed, this gives a fair correlation with the potential circuit performance.

It is clear, of course, that such quick methods, especially when they purport to include the effect of noise, cannot possibly reflect the extraordinary variations in the intensity and time distributions in the noise and other disturbances, which cause the errors that have been discussed in the preceding sections. If repeated measurements are taken over a long period, it would be expected that these would fall along some sort of hyperbolic law.
VIII. CONCLUSIONS

The exploratory studies presented here on the subject of impulse noise and error performance on data transmission circuits lead to the following conclusions.

IMPULSE NOISE

With regard to the impulse noise:

1. It is principally characterized by two distributions, one in amplitude and the other in times of peak occurrences.

2. The cumulative amplitude distribution of the impulse noise is generally described by a hyperbolic law, of order usually between 3 and 5. As compared with Gaussian or Rayleigh distributions, this gives considerably more probability to the amplitudes which are very much larger than the rms value.

3. The times of peak occurrences are distributed in much the same manner as transmission errors, considered below.

4. Short-time circuit opens or "hits," or more generally "multiplicative noise," has hardly been studied, but the indications are that its share in causing errors can occasionally be quite significant as compared with impulsive noise. The occurrence statistics appear indirectly to be much the same as for impulsive noise peaks or errors.
ERROR INCIDENCE

With regard to error incidence:

5. The timing of error occurrences can be described either in terms of the number of errors in successive 1-minute (or other appropriate unit) intervals of time, or in terms of successive spacings between errors. Description in one form is essentially a reciprocal of that in the other.

6. It appears that some change is occurring in the distribution of errors over comparatively short periods of time (fractions of an hour to a few hours) and over much longer periods of time (days to weeks and months). How this varies with wide ranges in bit rate has not been studied.

7. The shorter-period error distribution has been described in terms of a variety of models which are more or less reasonably realistic. Perhaps the best-fitting model is a Pareto distribution of the inter-error spacings. A Pareto distribution is hyperbolic, with the order as a parameter varying with over-all density of error incidence. None of these models extrapolates successfully to the longer-period distribution.

8. The longer-period error distribution has been described with good success in terms of an error density distribution which follows a simple hyperbolic law with a bias. The bias appears also as a normalizing constant
and varies with the over-all error density. The hyperbola is essentially of the first order. However, some occasional deviations, to somewhat higher than the first order, have been noted.

9. The distributions in both paragraphs 7 and 8 lead to error occurrences which appear clustered or "bunched" as compared with wholly random (or Poisson) occurrences. That is, they show a higher probability of high error density in occasional short intervals of time. Correspondingly they also show a higher probability of very long inter-error spacings.

10. As a consequence of the distributions in both paragraphs 7 and 8, "error chains" can occur, lasting over substantial periods of time (10 min or so). These occur, in long tests over long-haul circuits, with a very much higher probability than if the errors followed a purely random or Poisson distribution. These error chains have been observed experimentally. The probabilities increase appreciably with duration of test, so as to cast doubt on long synthetic tests made with noise or error recordings taken over moderate periods, and played over and over for the long duration.

**ERROR PERFORMANCE MEASUREMENTS**

The conclusions also touch upon the problem of error performance measurements of a data transmission circuit, namely:
11. The distribution laws in paragraphs 7 and 8, for error density, do not integrate to finite averages over an infinite duration test. Consequently, a technique has had to be devised to relate average error rate over the finite test to the parameters of the distribution. This involves, for example, a ratio between the average and median rates, which varies with the test duration. The median rate in such a case is equal to the bias.

12. As a consequence of paragraph 11 the error rates do not obey the laws of large numbers. Data taken over successive time intervals fluctuate widely, and a reasonably reproducible average error rate is not obtainable from measurements over a comparatively short test. Even over a long test the results are not as reliable as would be expected for a parameter that follows conventional statistics.

13. The most descriptive indication of error performance appears to be a long-time plot of the cumulative probability of error densities in short-time intervals (such as in Fig. 19). When comparing the performance of systems, unusual precautions must be taken if these are to be significant.

**AREAS FOR FURTHER STUDY**

It is clear from the discussions given in this study that there are areas which call for further exploration. Some of these are:

14. The study has been exploratory and empirical. Correlation and internal consistency among the various areas has not
been completely plumbed because the different portions were studied at different times, and time itself was limited. The general description secured from the experimental picture is undoubtedly correct, but continual checking over is required to obtain a complete detailed consistency among empirical equations.

15. The philosophy of measurement needs to be examined further. It is doubtful that it will ever be possible, by means of a brief measurement, say of some few minutes duration, to characterize reliably the error performance of a data transmission system.

16. Similarly, it will be necessary to develop a simple method of determining the relative performance of an arrangement A compared with an arrangement B, when the difference between these is not large. From paragraphs 11 and 12 it appears that this is not too easy. The problem is again an aspect of the philosophy of measurement.

17. It would be very desirable to present a physical model of error occurrence. One has been presented by Gilbert, but it is not at all realistic over long tests; and another by Mandelbrot, but it also breaks down for long tests, and it does not correspond exactly with his formula.

18. It would be very desirable to present error performance as a simple unified theory. The need for change of laws between short-time and long-time tests is awkward and unelegant. Perhaps a study of the simple hyperbolic law
of an order slightly greater than one (which deviation has been encountered occasionally in practice) is bound up with this.

19. In time it will become important to find the reasonably complete sources of the impulsive noise, and the reasons for the characteristic distribution of errors, in order most effectively to prepare systematic remedial measures. To have such sources would also please many engineers and mathematicians who do not fancy empirical laws and like to see a cause for the awkward mathematical processes needed.
Appendix

SPECIAL COORDINATE PAPERS

In the course of these studies on noise pulse and error distributions, it has been found desirable to introduce special coordinate scales on graph paper for certain specialized plots.

The reasons are primarily to adapt the scale to the range of the variable that is plotted on it. Namely, this is to include zero in a plot that is desired to be otherwise logarithmic; or radically to extend a portion of the scale to spread out points or curves so that they can easily be distinguished, without extending the remainder of the scale; or sharply to condense a section of scale that shows no particular curve features, so as not to need an inordinate size of paper; and so on.

Another important reason for some cases is to plot a significant law as a straight line.

BIASED LOGARITHMIC SCALES

Conventional logarithmic scales permit showing in at least some detail on one plot the course of a variable over an extreme range of values. There are many physical processes, however, where a parameter may have a very wide range of values; but because of a loss of sensitivity, or for other reasons, at the lower end of this range it goes to zero on a number of occasions. This is the case, with the error rate in a number of situations in the present report. The zero, however, does not naturally appear in the logarithmic scale, except at minus infinity, and therefore it cannot be plotted. Such a situation makes
for awkward graphs.

Or, in another application, it may be desirable to present the scale as an approximately linear one in the range from zero up to some particular value, and as approximately logarithmic from there on up. The zero would, of course, naturally appear on such a scale.

The problem can be readily solved by adding a bias to the argument of the logarithm. Thus if the normal argument is \( x \), and the new argument \( y \),

\[
y = x + a
\]

\[
\log y = \log (x + a)
\]

Then, if it is desired to plot the point \( x = 0 \), this appears at \( y = a \). The usual logarithmic plot of \( x \) would appear at minus infinity, but in the modified plot it appears at \( \log a \), which is quite finite and can be included in the sheet of paper.

This is illustrated in Fig. 53. At A is shown a normal scale for which the bias \( a \) is taken as zero. At B the bias \( a \) is taken equal to 1. Thus the scale of \( y \)'s starts at 1 and appears as a conventional scale. It is possible, however, to plot values of \( x \) down to zero. The scale of \( x \)'s, however, as it appears on conventional paper, is an inconvenient one.

The principal lines in the plot can be changed, however, to give a convenient scale, as shown at C. Here the accented lines are placed at \( y = 2; 11; 101; 1,001; \) and \( 10,001 \); to give the more convenient scale for \( x \).
Fig. 53 — Biased logarithmic scale to include zero in plot
A — Conventional logarithmic scale
B — Biasing of argument to include zero
C — Revision of grid in B

Fig. 54 — Biased logarithmic scale to expand grid just above 1

Fig. 55 — Probability scale based on log scale
(See Table 10)
It is seen that there is some visible distortion of the usual logarithmic scale between \( x = 1 \) and \( x = 10 \). However, at \( x = 100 \) and above, the distortion from a true logarithmic scale is hardly visible.

It is important to note that the scale for \( x \) is completely analytic from \( x = 0 \) (and even for a short interval below, i.e., to almost \(-a\)), on up to \( x \) equals infinity. Thus, there are no breaks in the biased scale, or the plots on it, at any point.

A biased scale of this type has been used in this report in Figs. 20, 21, 34, 37, 38, 49, and 52.

The case of Fig. 26 represents another problem. Here the incomplete Zeta function runs from 1 to very large values. As the parameter \( s \) changes over fairly large values, the function changes only very slightly, just above 1. On conventional logarithmic paper, the function for \( s \) from 3 to 10 would have been crowded and, in some cases, superposed and indistinguishable on a normal size sheet.

The solution adopted consists in subtracting 1 from the value of the function and then plotting the logarithm of this biased argument. Since the function is always greater than 1, the biased argument is always positive, and no trouble is experienced with its logarithm. The logarithmic scale then greatly stretches out the values near to 1 and permits their being distinguished. The resulting scale gives a good over-all indication of the course of the function, but it is still impossible on a normal size sheet to permit scaling off its values from the plot, with any reasonable degree of accuracy.

An illustration of the scale is given in Fig. 54, showing the parameter before and after biasing.
HYPERBOLIC SCALES

Hyperbolic scales of various orders are of use in a number of plots of parameters that can be encountered. They are usually designed to plot specific equations as straight lines, but they also have the property of extreme condensation of large values. This is because

\[ \frac{1}{x^n} = 0 \]  (176)

for any positive finite order \( n \). Thus, one can show values all the way up to infinity on the same sheet. Towards zero, on the other hand, the scale is correspondingly spread out.

The particular case in the present study was for Fig. 28, where \( N \) very slowly approaches an asymptote as \( k \) becomes large. This approach, at the larger values, becomes so slow that a considerable condensation of the scale (much more than given by the logarithm) is desirable. The scale could have gone all the way up to infinity on the plot, but it was not necessary.

As was noted in Section V, a reciprocal square root plot (i.e., \( n = 1/2 \)) is provided by Keuffel and Esser paper No. D-180, and this was used for the figure. To give the best spacing over the range of \( k \) depicted, a factor of 1/50 was imposed over the printed ordinates. Thus, these were translated as:

<table>
<thead>
<tr>
<th>Printed</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y )</td>
<td>0.1</td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
<td>0.8</td>
<td>1.0</td>
<td>2</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>
The results went a long way towards straightening out the plotted curves—though not enough to go all the way in the process.

**PROBABILITY SCALES**

The two objectives of the conventional normal probability paper are:

1. To plot a straight line for a normal distribution
2. To open out drastically the scale near 0 and near 1 and permit showing more detail in these portions of the plotted curves, where needed.

One occasional disadvantage is that commercially printed paper usually goes out only to 0.01 per cent at each end of the scale. If a plot needs to go further than this, the extrapolation of the grid is quite onerous.

A modification of the normal, or Gaussian, probability paper is the Rayleigh paper. On this a Rayleigh distribution plots as a straight line. It has also been used by Gumbel* to plot some of his special probability distributions. This paper is characterized by being unsymmetrical between the low and high ends. It stretches out the region near 1 and compresses that near 0, as compared with Gaussian paper. It has not been used in this study.

If one has a plot that goes down very close to 0, but not too close to 1, it is possible to use standard logarithmic paper as is.

The plot is then very simply

\[ y = \log p \]  

*(See Ref. 14, p. 176-177.*
This was used, for example, in Figs. 1, 8, and 32. A modified form of it, to be discussed below, was used in Fig. 18.

If one has a plot that goes up very close to 1, but down not too close to 0, one can use the standard logarithmic paper, but the log scale must be inverted. That is, one plots

\[ y = \log (1 - p) \]  \hspace{1cm} (178)

An illustration of this appears in Fig. 22.

The great advantage of both of these plots is that they permit going as far as necessary in one direction or the other, merely by using an appropriate number of log cycles. With these, however, one cannot go far out in both directions on the same plot (as can, for example, be done in principle on the conventional probability paper).

It is possible to do this, however, by combining both of the logarithmic treatments. Then one plots

\[ y = \log \left( \frac{p}{1 - p} \right) \]  \hspace{1cm} (179)

This makes it possible to go out as far as one likes in both directions, merely by using the number of log cycles needed. Points in the central part of the range, however, will require a computation using Eq. (179) before plotting on a conventional log grid. It will also be necessary, in this region, to re-draw some grid lines, with renumbering, to make the plot intelligible to the reader.

An illustration of both the logarithmic and probability scales appears in Fig. 55. Table 10 lists the principal grid lines.
Table 10

PRINCIPAL GRID LINES--HYPERBOLIC PAPER

<table>
<thead>
<tr>
<th>( p )</th>
<th>( 1 - p )</th>
<th>( p/(1-p) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>.99</td>
<td>.01</td>
<td>99.</td>
</tr>
<tr>
<td>.98</td>
<td>.02</td>
<td>49.</td>
</tr>
<tr>
<td>.97</td>
<td>.03</td>
<td>32.3</td>
</tr>
<tr>
<td>.96</td>
<td>.04</td>
<td>24.</td>
</tr>
<tr>
<td>.95</td>
<td>.05</td>
<td>19.</td>
</tr>
<tr>
<td>.94</td>
<td>.06</td>
<td>15.7</td>
</tr>
<tr>
<td>.93</td>
<td>.07</td>
<td>13.3</td>
</tr>
<tr>
<td>.92</td>
<td>.08</td>
<td>11.5</td>
</tr>
<tr>
<td>.91</td>
<td>.09</td>
<td>10.1</td>
</tr>
<tr>
<td>.9</td>
<td>.1</td>
<td>9.</td>
</tr>
<tr>
<td>.8</td>
<td>.2</td>
<td>4.</td>
</tr>
<tr>
<td>.7</td>
<td>.3</td>
<td>2.33</td>
</tr>
<tr>
<td>.6</td>
<td>.4</td>
<td>1.5</td>
</tr>
<tr>
<td>.5</td>
<td>.5</td>
<td>1.</td>
</tr>
<tr>
<td>.4</td>
<td>.6</td>
<td>.667</td>
</tr>
<tr>
<td>.3</td>
<td>.7</td>
<td>.429</td>
</tr>
<tr>
<td>.2</td>
<td>.8</td>
<td>.25</td>
</tr>
<tr>
<td>.1</td>
<td>.9</td>
<td>.111</td>
</tr>
<tr>
<td>.09</td>
<td>.91</td>
<td>.099</td>
</tr>
<tr>
<td>.08</td>
<td>.92</td>
<td>.087</td>
</tr>
<tr>
<td>.07</td>
<td>.93</td>
<td>.0753</td>
</tr>
<tr>
<td>.06</td>
<td>.94</td>
<td>.0639</td>
</tr>
<tr>
<td>.05</td>
<td>.95</td>
<td>.0526</td>
</tr>
<tr>
<td>.04</td>
<td>.96</td>
<td>.0417</td>
</tr>
<tr>
<td>.03</td>
<td>.97</td>
<td>.0311</td>
</tr>
<tr>
<td>.02</td>
<td>.98</td>
<td>.0204</td>
</tr>
<tr>
<td>.01</td>
<td>.99</td>
<td>.0101</td>
</tr>
</tbody>
</table>
An important advantage of this probability paper is that hyperbolic laws of the first order plot as straight lines. Thus

$$p = \frac{h}{x + h}$$  \hspace{1cm} (180)

is converted to

$$y = \log \left( \frac{h}{x + h} \right) - \log \left( 1 - \frac{h}{x + h} \right)$$  \hspace{1cm} (181)

$$= \log \left( \frac{h}{x + h} \right) - \log \left( \frac{x}{x + h} \right)$$  \hspace{1cm} (182)

$$y = \log h - \log x$$  \hspace{1cm} (183)

This is a straight line on conventional log paper.

Similarly if

$$p = \frac{x}{x + h}$$  \hspace{1cm} (184)

$$y = \log \left( \frac{x}{x + h} \right) - \log \left( \frac{h}{x + h} \right)$$  \hspace{1cm} (185)

$$y = \log x - \log h$$  \hspace{1cm} (186)

This is a straight line with the opposite inclination to that for Eq. (183).

"NEARLY WORST" INTERVALS

It has been indicated (Section VII) that a useful description of how a data transmission system is behaving is to divide its operation into one-minute (or other appropriate unit) intervals and arrange these in order of the number of errors occurring in each. One can start with the interval, in the entire test, which has the most errors, and let it be followed by the one having the next most errors,
etc. This is obviously the method followed in plotting Fig. 19.

One can then separate out, say, the worst decile from this. The very worst cases, may, however, be highly accidental, and there may be some advantage in taking the decile between the worst 5 per cent and the worst 15 per cent. This might be called the "nearly worst decile."

One could then continue and study the "nearly worst centile," which would be the interval between the worst 0.5 per cent and the worst 1.5 per cent. One can go on to similar smaller intervals, some of which have not been named, as indicated below:

<table>
<thead>
<tr>
<th>Fractional Interval</th>
<th>Center Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decile</td>
<td>0.05 to 0.15</td>
</tr>
<tr>
<td>Centile</td>
<td>0.005 to 0.015</td>
</tr>
<tr>
<td>Milile</td>
<td>0.0005 to 0.0015</td>
</tr>
<tr>
<td></td>
<td>0.00005 to 0.00015</td>
</tr>
</tbody>
</table>

These "nearly worst intervals" were indicated in the abscissa grid of Fig. 18. The geometric mean of the number of errors in a given such interval is a fairly good index of the circuit performance, provided the test has been going on long enough.
REFERENCES


Addendum

ERROR PERFORMANCE DATA REPORTED BY BELL SYSTEM

After this Memorandum had been written, and while it was being processed for reproduction, a paper appeared* by Townsend and Watts which presented a number of summary curves on the error performance of the system under test, previous to error control. It is of interest to compare some of these with the general experience and the formulas which have been described in this Memorandum.

NATURE OF TESTS

A number of calls were made over switched telephone circuits, and one-half hour's data were transmitted back to one or the other of two receivers, at Murray Hill and Holmdel. The remote transmitting stations ranged from some at local exchanges and nearby cities (such as Rahway, Ridgewood, New York, etc.) to some at more distant cities such as Boston, Washington, Atlanta, Los Angeles, San Francisco, etc.

The data set and its characteristics, and the major results germane to the comparison here are:

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA-PHONE Set</td>
<td>No. 201 A</td>
</tr>
<tr>
<td>Baud rate</td>
<td>1000</td>
</tr>
<tr>
<td>Bit rate (2 bits per signal element)</td>
<td>2000 per second</td>
</tr>
<tr>
<td>Word length</td>
<td>31 bits</td>
</tr>
<tr>
<td>Repetitive test pattern length</td>
<td>511 words</td>
</tr>
<tr>
<td>Number of transmitter locations</td>
<td>28</td>
</tr>
</tbody>
</table>

### Comparison—Long-Time Distribution

In the Bell System report, Fig. 2 shows the distributions of bit error rates per call for a variety of types of calls, and for all the calls grouped together. If all calls represented in a curve had been over the same circuit, the curve would show the long-time distribution of the error rate. If all the calls for a curve were over the same type of circuit, and an ergodic principle were applicable, it would still represent a long-time distribution of the error rate. As it was, each curve represents a long-time sample of a user experience with the circuits categorized. While this is not the same as the long-time error rate on one circuit, it can be plotted in the same way and compared with it.

The curves have been re-plotted herewith in Fig. A. The figure differs from the authors' in that it has been translated to correspond with Fig. 19 of the present Memorandum. (The curves appear as series of broken segments which correspond with points taken from the figure in the paper, and which have not been re-smoothed.) The plot shows much the same characteristics as Fig. 19. Namely, the curves follow approximately the simple hyperbolic law of Eq. (80), although over most of their length they are slightly steeper. The straight line $S$ in Fig. A has the $-1$ slope of

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of hours of transmission</td>
<td>273</td>
</tr>
<tr>
<td>Number of calls</td>
<td>548</td>
</tr>
<tr>
<td>Total bits transmitted</td>
<td>$1.97 \times 10^9$</td>
</tr>
<tr>
<td>Total number of bits in error</td>
<td>$6.30 \times 10^4$</td>
</tr>
<tr>
<td>Average bit error rate</td>
<td>$3.19 \times 10^{-5}$</td>
</tr>
</tbody>
</table>
Fig. A—Bit error rate distributions for: All calls, exchange, long-haul, and short-haul.
After Fig. 2 of Townsend and Watts

Eq. (80), for comparison.

The curves show a slight concavity downward. This does not appear in Fig. 19, but it has been experienced before and is commented on in Section IV.

COMPARISON—EFFECT OF MEASURING INTERVAL DURATION

The Bell System authors show in their Fig. 16 the effect of averaging errors over 5-minute intervals as against 1-minute intervals. This sort of thing has already been touched on in Section VII, and Figs. 50, 51, and 52. The authors' Fig. 16 has been re-plotted herewith as Fig. B, with the curves smooth, and with the bit error rates plotted per $10^3$ bits instead of per
Fig. B—Distribution of error bit rates per one- and five-minute intervals. After Fig. A6 of Townsend and Watts measuring interval. This enables a more critical comparison between the two measuring intervals.

It will be noted that for the lower error rates (or to the bottom and right of the curves) the error rate is higher for the longer measuring interval. This accords with the observations in Figs. 51 and 52, in which both measuring intervals were very much longer than in Fig. B.

However, for the higher error rates, and shorter portions of the total test time, the curves cross; to indicate a lower error rate for the longer measuring interval. This is interesting in that it suggests what happens as the error distribution no longer follows the simple hyperbolic law. It will be commented on again below.
COMPARISON--SHORT-TIME DISTRIBUTION

The short-time error distribution in the Townsend and Watts report would be the distribution within a call duration. The first point to check would be to see how the error spacings compare with the Pareto distribution assumed in the Mandelbrot model, namely Eq. (113).

In their Fig. 7, the authors have plotted such a distribution. From the discussion in Part 5, it is clear that it would have been desirable to separate the calls into groups that each had approximately the same error performance. However, all the calls were merged together for the curve, and this will have to be used.

A re-plot of the authors' Fig. 7 is given in Fig. C herewith. The principal changes are, first, that error-free intervals are translated to spacings between successive errors (i.e., in which cases of errors in immediately successive bits are counted as one bit apart, instead of as zero-bit intervals); and, second, that the

![Graph showing error spacings distribution](image)

**Fig.C** — Distribution of error spacings. After Fig.7 of Townsend and Watts. Comparison with Mandelbrot model.
direction of probability cumulation is inverted.

The experimental curve is compared in Fig. C with the assumed Pareto distribution of Mandelbrot, in which the slope has been computed from the average error rate and the other parameters of the individual call in the test, in the following manner.

The parameters of the individual call are, using the notation of Section V:

\[ T = 3.6 \times 10^6 \text{ bits} \quad A-1 \]
\[ N = 63,000/548 = 114.8 \text{ errors} \quad A-2 \]

The formula relating these to the slope \(-k\) is Eq. (127), namely

\[ T = N^{1/k} S(N, 1/k) \quad A-3 \]

A quick solution for \( k \) may be obtained directly by examination of Fig. 27, which gives

\[ k \approx 0.32 \quad A-4 \]

By the method of successive approximations, and using interpolations from Table 5, one gets

\[ k = 0.318 \quad A-5 \]

This is the value used for plotting the dotted straight line of Fig. C. It seems to be a fair fit with the data if one insists on a Pareto line crossing the 1,1 point.

A comparison of Fig. C with Fig. 32 shows, however, that the Pareto law does not fit the Townsend and Watts data quite as well as it fits the Hoffman data of the IBM German Laboratories quoted
by Mandelbrot. Indeed, the Bell System data seem possibly to be better fit by the Gilbert curve \( x \) of Fig. 32. However, for the moment, the Mandelbrot approximation will continue to be considered.

Townsend and Watts also show in their Fig. 14 the probability that exactly \( m \) bits will be in error in a block of \( M \) bits. This is really a short-time distribution, first because as its largest value taken \( M \) bits covers only a very small fraction of a second, and second because the smallest incidence of error considered is high enough for the event to occur only a very small proportion of the time (even on a cumulative basis, from high error incidence down).

The curves have been re-plotted in Fig. D herewith. The probability density of the authors has been integrated into a cumulative probability for easy comparison with other figures, and the error occurrence has been changed to bit errors per bit, also for easy comparison (it is so large that the errors are taken per bit rather than per \( 10^5 \) bits).

The first point to be noted is that the error occurrence is lower for the larger \( M \)'s (or longer measuring intervals), entirely in conformity with what was noted in Fig. B for the larger and less frequent error occurrences. Of course, since the data are the same, this would be expected, but the real point of the observation is that the cross-over in Fig. B is distinctly systematic and not accidentally caused by any possible lack of accuracy in reading and translating the curves from the comparatively small illustrations in the printed Journal.
Fig. D — Error rate distributions measured in blocks of $M$ bits. After Fig. 14 of Townsend and Watts. Comparison with computation from extension of Mandelbrot model.

The second point which can be noted is that the curves are remarkably similar to the curves in Fig. 30, which were deduced by an extension of the Mandelbrot hypotheses. The testing parameters are different—in particular, the probabilities involved run toward the lowest extreme of Fig. 30 and extend two decimal logarithmic cycles lower.

To permit a more meaningful comparison a new curve has been computed, exactly similar to those in Fig. 30, but using the parameters described by Eqs. (A-1) to (A-5) above. The computations involved
interpolations of values listed in Table 5. This has then been plotted as the dotted curve in Fig. D, labelled as "Mandelbrot-Mertz."

This last curve generally fits the trend of the experimental data well. Such differences as there are can presumably be caused by the deviation noticed in Fig. C between the experimental data and the Pareto distribution.

CONCLUSIONS

One may conclude from the above observations that:

1) The newly reported Bell System error data fit quite reasonably well with the general experience and formulas which have been outlined in this Memorandum. This holds even though the experimental conditions are not always exactly parallel to those hypothesized in the Memorandum.

2) The long-time distribution of errors, over a variety of circuits, follows much the same approximate hyperbolic law that has been found to fit a single circuit over a long-time test.

3) The variation of observed average error incidence with duration of the measuring interval (over which the average is taken, for each observation point) follows that described in the Memorandum for the moderate error incidence periods which make up the substantial part of a long-time distribution; namely, the longer the measuring interval, the higher the average comes out. For the rarer short periods of extremely high error incidence there seems to be a cross-over, and the longer the measuring interval, the lower the average comes out. This cross-over would appear to merit further
examination at some time, but it cannot be done within the compass of the present note. It should be grouped with the "Areas for Further Study" of Section VIII.

4) The short-time distribution of spacings between errors fits the Mandelbrot model only rather approximately. Perhaps this is due to the merging of too great a variety of data. The distribution does seem to be more aptly describable by the Gilbert model.

5) In spite of conclusion 4), the short-time distribution of error occurrences seems to follow fairly well a deduction reached from an extension of the Mandelbrot model in this Memorandum.