NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Memorandum 33-699

Error Control in the GCF: An Information-Theoretic Model for Error Analysis and Coding

Oduoye Adeyemi

(NASA-CR-140852) ERROR CONTROL IN THE GCF: AN INFORMATION-THEORETIC MODEL FOR ERROR ANALYSIS AND CODING (Jet Propulsion Lab.) 193 p HC $7.00 CSCL 17B
N75-12171
Unclas
G3/32 03571

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

October 15, 1974

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Technical Memorandum 33-699

Error Control in the GCF: An Information-Theoretic Model for Error Analysis and Coding

Oduoye Adeyemi

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

October 15, 1974

Prepared Under Contract No. NAS 7-100
National Aeronautics and Space Administration

Preface

The work described in this report was performed by the Telecommunications Division of the Jet Propulsion Laboratory.

ACKNOWLEDGEMENT

This report benefits from many discussions with Howard Rumsey.

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. 33-699	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle ERROR CONTROL IN THE GCF: AN INFORMATION-THEORETIC MODEL FOR ERROR ANALYSIS AND CODING		5. Report Date October 15, 1974
		6. Performing Organization Code
7. Author(s) Oduoye Adeyemi		8. Performing Organization Report No.
9. Performing Organization Name and Address JET PROPULSION LABORATORY California Institute of Technology 4800 Oak Grove Drive Pasadena, California 91103		10. Work Unit No.
		11. Contract or Grant No. NAS 7-100
		13. Type of Report and Period Covered Technical Memorandum
12. Sponsoring Agency Name and Address NATIONAL AERONAUTICS AND SPACE ADMINISTRATION Washington, D.C. 20546		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This memorandum covers one aspect of the total effort to understand the structure of data-transmission errors within the Ground Communications Facility (GCF) and provide error control (both forward error correction and feedback retransmission) for improved communication. The emphasis is on constructing a theoretical model of errors and obtaining from it all the relevant statistics for error control. Thus, no specific coding strategy is analyzed, but references to the significance of certain error pattern distributions, as predicted by the model, to error correction are made.

17. Key Words (Selected by Author(s))	18. Distribution Statement
Information Distribution and Display Information Theory Mathematical Sciences Telemetry and Command	Unclassified -- Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	191	

CONTENTS

List of Tables

List of Figures

List of Figures
(Continued)

Section I

INTRODUCTION AND SUMMARY OF RESULTS

(i) Introduction

This report covers one aspect of our total effort to understand the structure of the errors on the Ground Communications Facility (GCF) and provide error control (both forward error correction and feedback retransmission) on it for improved communication. Here we are concerned mainly with constructing a theoretical model of errors and obtaining from it all the relevant statistics for error control. Thus no specific coding strategy is analyzed in this report, although references are made in appropriate places to the significance to error correction of the distributions of certain error patterns as predicted by the model. The success of our continuing efforts in designing specific error correction schemes on the basis of this GCF model will be reported elsewhere.

Our model is based on the 4800 bps high-speed GCF dataline test run provided by J. P. McClure [1] in March 1973, although we show that the same basic model is good for the 50 kbps wide-band data we analyzed earlier in [2]. Indeed all the error statistics that are calculated for the high-speed dataline are also obtained for the wide-band dataline. As shown in Table 1, the high-speed data set consists of 31 test runs on all the NASA lines between JPL and each of the outpost stations at Goldstone, Florida, Madrid (Spain), South Africa and Australia. McClure [1] has a detailed account of how the data were collected. There are two of the 31 test runs in which no errors are recorded (Madrid-JPL, duration 102 minutes; Goldstone-JPL, duration 146 minutes), but this perfect transmission is due to the line condition at

Table 1. Source of 4.8 kbps data (adapted from McClure [1])

CTA 21	JPL	DSS 51	South Africa
DSS 14	Goldstone	DSS 61	Madrid
DSS 42	Australia	DSS 71	Florida

Origin	Destination	Starting Time Day	Hour	Duration Min:Sec	Bit Error Rate, x10^{-5}	P(1|1)	Group
42	21	298	02	214:31	1.30	0.340	Amber
21	42	298	02	214:30	4.50	0.351	Amber
71	21	298	23	178:51	1.25	0.401	Amber
61	21	299	04	102:06	0	0	Green
21	61	299	04	107:49	2.70	0.366	Amber
42	21	299	23	192:42	0.23	0.285	Green
21	42	299	23	192:14	0.36	0.300	Green
14	21	300	03	146:26	0	0	Green
21	14	300	03	146:50	0.18	0	Green
71	21	300	23	52:22	25.4	0.410	Red
71	21	301	00	88:05	51.51	0.315	Red
71	21	301	01	24:36	2.36	0.186	Amber
61	21	301	02	222:16	15.1	0.388	Red
21	61	301	02	222:29	1.16	0.416	Amber
42	21	326	21	118:11	0.66	0.399	Green
21	42	326	21	118:10	1.13	0.342	Amber
42	21	326	23	44:58	0.02	0	Green
21	42	326	23	44:57	1.00	0.015	Amber
51	21	333	17	191:02	3.10	0.325	Amber
61	21	335	16	75:43	0.98	0.420	Green
21	61	335	16	75:27	2.56	0.374	Amber
61	21	335	18	99:51	2.30	0.324	Amber
21	61	335	18	99:51	2.48	0.154	Amber
42	21	335	21	170:12	4.02	0.271	Amber
21	42	335	21	170:15	6.49	0.323	Amber
61	21	340	16	76:00	0.51	0.277	Green
21	61	340	16	75:	3.97	0.392	Amber
61	21	340	18	100:36	1.63	0.405	Amber
21	61	340	18	100:	1.79	0.353	Amber
51	21	340	20	168:00	5.30	0.385	Amber
21	51	340	20	151:58	6.51	0.368	Amber
				3986.0 min. = 66.4 hrs.			

that time of day rather than a permanent feature of the lines between those stations. For example, the same Madrid-JPL link on another day (day 301) has one of the worst error rates recorded in the 31 runs. We follow McClure in dividing the different bit error rates obtained in the tests into Green, Amber and Red groups: the Green group consists of those with bit rates of less than 1×10^{-5}, those in the Amber group have bit rates between 1×10^{-4} and 1×10^{-5}, and in the Red group are those with a bit error rate higher than 1×10^{-4}. Of the 31 test runs, only 3 are in the Red group, 7 in the Green and 19 in the Amber group. (We discount the two error-free runs in the analysis.)

A rate-one code built into the GCF modems causes a fixed pattern of errors after each random error on the channel. In the 4800 bps high-speed data these fixed errors occur at bit positions 18 and 23 away from each random error. The positions are 3 and 20 in the 50 kbps wide-band data. It is now being determined whether to remove this fixed error-causing code or process the received data to remove the errors after each transmission.

It is not the high bit error rate, however, that makes this type of channel difficult to model. Rather it is the fact that the errors, when they do occur, tend to cluster together. In other words, the channels display some memory. How long or short a memory one should build into the model depends on the particular channel and the ease of handling the analysis of a model with a realistically long memory. On the GCF, the chance that a bit error will be followed by another bit error, denoted by $P(1|1)$ in Table 1, ranges from a high of 42% to a low of less than 2%, depending on the data-line condition and bit error rate. For example, in the Green group, a long error-free transmission followed by a burst of errors lasting only one second may have a high probability of consecutive bit errors while another test run with burst of errors scattered through the whole duration may result in high bit error rate and low probability of consecutive errors.

Before we summarize the results of the report, let us fix our ideas of a burst. As definition we adopt an intuitive notion of determining a burst from a sequence of transmissions on the channel as a sequence of bits beginning and ending with an error, separated from the nearest preceding and following error by a gap of no less than a given length, say G, called the guardspace and containing within it no gap of length equal to or greater than G bits. From this definition, it is clear that the length (in bits) of a burst depends on the guardspace G; the longer the guardspace, the longer the burst length, some of the bursts at shorter guardspace being combined into a single burst when the guardspace becomes longer. For example, for G = 400, the first test run contains 322 bursts, the longest of which is 6133 bits containing 141 errors. The same run for G = 3600 has only 100 bursts; the length of the longest burst is now 217362 bits containing 3550 errors. This is typical of the GCF data line; in the error mode, there are still some good runs several hundred bits long but not long enough to allow more than a few 1200-bit blocks to pass through error-free.

The histogram for the thirty-one runs of the high-speed 4.8 kbps data is shown in Figure 1. The error-free gap lengths are represented on the X-axis and their frequencies in the 31 runs, that is the number of times a gap of length X appears, on the ordinate. For example, the number of consecutive errors (at X = 1) is 17, 149, while the number of times gaps of length $100 \leq X \leq 499$ appear is 652.

(ii) Summary of Results

There are two broad classes of theoretical models that have been proposed for burst noise channels: the Independent Gap Model (or the Pareto Model), which assumes that successive gaps are approximately independent and suggests the Pareto distribution for the gaps, and the Markov model, which combines Markovian property with

Independent Gap property. The Markov model assumes that given that an error has occurred on the channel, the length of the gap following the error is independent of the length of the gap prior to that error bit. In general, however, when errors may occur in more than one state of the channel, the Markov model does not assume the Independent Gap property underlying the Pareto model. By looking at the graphs of certain functions of the empirical gap distributions, we showed that the Pareto model cannot be employed to model the errors on the GCF. Indeed, the Pareto model performs well only on good quality telephone channels (only in the Green error mode). Since our concern here is with the Red and Amber error modes, we have restricted our choice of a model to the Markov class and succeeded in getting a five-state model, diagrammed in Figure 3, that gives an acceptably good fit.

The five-state model we used has only one error state B which connects to perfectly good states G_1, G_2, G_3, and G_4. Errors occur in state B with probability one each time the process enters this state, consecutive errors occurring with the indicated probability $0 < q < 1$. The further away a good state is from B, the longer the sojourn time of the process in that state. Long gaps indicate the process is in the best state G_1 and the short bursts of errors indicate transitions between the error state B and G_4.

A single state B in which errors occur with probability one is not really acceptable, as close scrutiny would reveal, for it is well-known that the error-causing mechanism on the channel does not reverse the bit each time an error occurs, a fact which we seem to ignore in our model. We hasten to point out, however, that a model using a state $\overline{B}$ in which errors occur with some probability $0 < h < 1$ instead of B can be made to be mathematically equivalent to our model by appropriately increasing the number of good states and adjusting the corresponding transition probabilities. Moreover, introducing such a state $\overline{B}$ would involve unnecessary complications in the analysis.

A general method of getting maximum likelihood estimates (MLE) of the model parameters from the raw estimates obtained from data is presented. This method is applicable to any finite-state Markov process and hence to any Markov model.

We consider the gap distribution a basic property of the channel because, in our model, the process renews itself each time it enters the error state. In other words, the occurrence of an error is the renewal event which wipes out the memory of the past gap. That is why we judge the performance of our model by how good a fit it gives to a function of the gap distribution as calculated from the data. Sample graphs of typical fits in each of the three error modes are shown in Figures 4, 5 and 6. Goodness-of-fit tests are performed for each of the error runs, each of the Green, Amber and Red error mode channels, and for a single channel obtained by combining all the error runs and treating each as an independent sample from some basic common distribution. Since it is more important, for purposes of error control, to have very accurate predictions of error clusters when the gaps are short (high bit error rate) than during long intervals of error-free transmission, we concern ourselves with just how good a fit we obtain for gaps of 4000 bits or less. The results are very good indeed for individual test runs. The Kolmogorov-Smirnov test predicts that in 99% of the time the error of our prediction (the absolute difference between the model and empirical values) should not be more than 3.6%. In the Red group, the maximum error of our prediction is 2.3% (see Tables 4a and b). But the better the channel (the lower the error rate), the less spectacular this agreement becomes. For example, two of the 19 test runs in the Amber group fail this test only slightly while the fits obtained in the two cases that fail the test in the Green group are less than satisfactory; in this case, the bit error rates are 2×10^{-7} and 9.8×10^{-6}, and the percentage prediction errors are 51.4 and 4.5, respectively. We then ask that the errors not be more than 2.5%. The statistical test in this case says that about 85% of our test runs should have less than this percentage error. In the Red group, the highest prediction error of 2.3% falls below the theoretical bound of allowable deviation, while about 58% of the Amber group pass the test. The important thing is that those test runs with high error rates all give acceptably good fit with prediction errors of less than 2.5%.

The grouped channels (Red and Amber) give less excellent agreement with individual data runs. Errors of up to 8% are recorded in the Red group and 14.5% in the Amber. This fact is in great part due to the wide range of error rates recorded in each mode: 15×10^{-5} - 51×10^{-5} in the Red and equally wide variation in the Amber. But the greatest revealing fact was obtained when we attempt to fit a single channel to every one of the 29 error runs. The error is about 10% in the Red, between 2% and 68% in the Amber, and up to 77% in the Green group. It is therefore clear that the errors on the GCF do not follow a single distribution. In other words, the channel performance is significantly different for varying line conditions. It is now understood that this is caused by the varying load on the GCF. When users come onto or drop off the channel, the characteristic of the channel changes. A realistic model should incorporate the times between these changes and the characteristics of the channel when the changes occur. A way of constructing such a model is detailed. This and all the results mentioned above are presented in Section II.

Section III is devoted to the autocorrelation of bit errors. This is the probability of having an error k bits away following a given initial error, $k \geq 0$. From it we not only gain knowledge of significant error patterns but we also deduce the memory of the channel in the different error modes. For example, in both the Red and Amber groups the memory is very much longer than 1200 bits. It is only in the Green group that the memory is just about 1200 bits (see Tables 6 and 7).

In Section IV, we deal with the capacity of the channel (the maximal rate for which reliable transmission over the channel is possible). Since the capacity of a burst-noise channel is always larger than that of a binary symmetric channel (BSC) with the same bit error rate, which, at the error rate on the GCF, is large enough (>0.996 for the high-speed circuit and >0.997 for the wide-band), it is clear that for purposes of error control the capacity does not present any problem. The irony is that forward error-correction is more difficult than for the corresponding BSC.

One group of statistics that turns out to be very important in estimating the performance of block codes is the block-bit statistics. A block is defined as a sequence of n bits for a fixed integer n. This group includes:

(a) the block error rate as a function of block sizes;

(b) distribution of the number of errors in a block;

(c) distribution of distances between extreme errors in a block;

and

(d) distribution of errors in a code interleaved to some depth t.

All these statistics are presented in Section V.

Discounting the outages (those times when the error rate was so high that transmission was stopped) which McClure speaks about in [1], the block error rate for a 1200-bit NASA-standard block length in the 4800 bps data ranges from a low of .021 of 1% during the Green error mode to as high as 1.8% in the Red group. An error block is defined as a block having one or more bit errors. Tables 10 & 11 show the predicted and empirical block error probability for the HF 4.8 kbps and wide-band 50 kbps circuits. In both cases the predicted values are very close indeed to the data values. There is also close agreement between predicted values of block thruput and the empirical values calculated by McClure in [1]. The block thruput refers to those blocks received error-free. For the Green group the empirical value is 99.96%, the model value is 99.963%. For the Amber the values are 99.63% and 99.79%, the Red group gives 97.71% and 99.34% respectively. The total average empirical block thruput of 99.55% is close to the predicted value of 99.78% (see Tables 14 and 15).

But it is the density of errors in the error blocks that is more important. For if an error block contains only one error it is an easy matter to locate and rectify that error with only a few changes in the present specifications on the GCF. Even if there are more errors but they are all confined within a given length in the block it is still possible to find a burst-trapping code that will correct all of them. This is why we have calculated not only the distribution of errors in a block and the proportion of the blocks with more than a given number of errors but also the distances between the first and last errors in an error block. The proportion of error blocks containing twenty-five or more errors is less than 25% in most of the runs at 48 bps. Runs in which this proportion is more than 50% are really badly hit for in them the proportion containing fifty errors or more is equally high, so that the 3-bit maximum error correction capability which can be achieved on the GCF even if all the 33 bits currently allowed for error detection and correction in each 1200-bit block were used for error correction alone would still fall short of correcting a large proportion of the error-blocks (Table 13a and b). Graphs of the distribution of errors in a block are provided in Figures 8, 9, and 10.

The effect of the fixed error pattern introduced by the rate one code built into the modem becomes apparent in the distribution of distances between extreme errors (or of length of burst of errors) in a block mentioned earlier. If we consider only those blocks with two or more errors, the proportion having their burst confined to within exactly 23 bits is as much as 24% in the Red group; in the Amber group there is a run with as high as 83% while a percentage of 98 is recorded in the Green group (Table 5). This explains why the empirical and model values in this case are not as close as one has expected (Table 16) and further impels us to remove this error-causing modem code so as to be able realistically to assess the performance of the different error correcting schemes that are now being considered for the GCF.

A rather effective way to correct burst noise is to interleave the coded blocks to some depth t, say. Here the bits of each coded block are not transmitted consecutively but are interspersed in such a way that they are transmitted exactly t bit positions apart. For sufficiently large t, at the receiver, the blocks appear to have been corrupted by random noise thus spreading out the error clusters over many blocks. The trick is to thin out the errors in each block to a sufficiently low number that a known error correcting code (e.g., a BCH code) of high enough capability can be used to correct the resultant errors. Taking t = 6 and interleaving each block so that the length of each block transmitted separately is 200 bits we found that the proportion of blocks containing few errors has increased thus decreasing the proportion with a large number of errors. For example, in the Red group the proportion containing exactly one error increased to 0.06 of 1% with only 0.0045 having four or more errors; these proportions are 0.05 of 1% and 0.0006 respectively for the Amber group and proportionately higher numbers for the Green group. The encouraging fact is that in each case the proportion of blocks with three errors or less is at least twice as high as it is without interleaving. One therefore can expect better results (higher proportion having fewer errors) when the depth of interleaving t is increased thereby enabling us to correct a sufficiently high proportion of the errors without significantly changing the present GCF standard specifications.

As opposed to the distribution of bit errors in a block found above, in Section VI we concern ourselves with Block (symbol) Error distribution. In this case a block is considered as made up of symbols, each symbol being a fixed number of bits. Because of the burst noise it may be more efficient to employ an algorithm designed to correct up to a given number of symbol errors in a block rather than one that can correct only bit errors. This is especially so in cases where, although the number of bit errors in the block is higher than the error-correcting capability of the code employed, the errors are all confined to within only a few symbols.

Let us mention particularly the distribution of error symbols in n-symbol word and the autocorrelation of error symbols. For the standard 1200-bit block, if a symbol length of 6 bits is used, then the proportion in the Red group of the 200-symbol blocks that have 3 symbol errors or more is only 0.0061. Thus a code having two symbol error-correcting capability will fail to correct in only 0.61 of 1% of the time. To achieve this efficiency an error-correcting code must be able to correct up to 5 bit errors in the 1200-bit block. The proportions for symbol lengths s = 8 and 10 and for all the different error groups are shown in Table 21. Table 22 contains the autocorrelation of symbol errors for symbol lengths s = 6 and 10 bits. For s = 6 bits, the highest correlation between an initial error-symbol and another 200 symbols away is 0.08, the least is 0.0028 in all cases in the 4800 bps circuit where the symbol error memory is longer than 200. This means that unless combined with forward error-correction, a feedback retransmission scheme may be impractical on the GCF for error correction because occurrence of an error block (symbol) may cause a high number of others to occur in quick succession thus causing a problem of buffer over-flow.

As an immediate application of the block-bit and symbol error distributions we find the sync acquisition and maintenance probabilities. The two strategies we compare are both based on using a prefix sequence of 24 bits in each of the 1200-bit blocks. These strategies are:

1. to accept sync if there are not more than 3-bit errors in the prefix sequence
2. to accept sync if there is at most (only) one error symbol in the 24-bit prefix considered as four 6-bit symbols.

Our criterion of comparison is the efficiency of each of the algorithms in reacquiring a lost synchronization within a frame of 1200 bits after it is lost and of maintaining it once it is reacquired. It is found that the first algorithm will lock onto the wrong synchronization in over 16% of the time although it will hardly fail to identify the true sync sequence. On the other hand, the second scheme will lock onto the wrong sync in less than 2% of the time and it is equally as efficient as the first in not failing to identify the true sync sequence (Tables 23a and b, 24a and b).

This conclusion is not surprising since the second algorithm takes advantage of the burst noise by allowing up to 5 errors provided they all occur within a single 6-bit symbol.

To answer the question as to when the first algorithm is efficient, we increase the prefix sequence to 30 bits and find that allowing up to 3 bit-errors will falsely detect a sync in less than 0.5% of the time. In this case the second algorithm provides ample protection against both types of errors.

Lastly, in Section VII we take up the important question of burst distribution. The fact that no good error correcting device can be constructed without the knowledge of this distribution attests to its importance.

To understand the nature of the bursts we find out how long they are, how dense the errors within them are and particularly how many 1200-bit blocks are affected each time the channel enters into a burst mode. Specifically, we calculate the

(a) distribution and mean of burst lengths;

(b) the distribution of the density of errors in a burst of given length and the mean number of errors in such a burst;

and

(c) the block-burst distribution.

The last distribution is intended to give us an idea of the number of blocks that are likely to be affected each time a burst of error occurs.

But before we answer the above questions we review two criteria of choosing an optimum guardspace G since all the distributions depend on G. We agree to call a G optimal for a code C (with desirably high rate R) if a high proportion of the bursts (with respect to G) is less than the burst correcting capability of C. The burst correcting capability of a code relative to G is the largest integer, b, for which every noise sequence containing only bursts of length b or less is correctly decoded. It is shown that a guardspace of 400 bits hitherto being used in [1] is not adequate. However a guardspace of 3600 bits seems to work for both the 4800 bps high-speed and the 50 kbps wideband data. See Tables 27a and b for the bursts using different values of G.

In the 4800 bps circuit the mean burst length varies from 41 bits in the Green error mode to 340 bits in the Red with an average of 135 bits overall. The high standard deviation of burst lengths (Table 28) is explained by the wide variation in the bit error rate ($0 - 10^{-3}$).

The error density in a burst is obtained for guardspace G = 400 and 3600 bits (Table 29). This density can be as high as 6% in the highest error mode; in the 50 kbps circuit it can be up to 8% when G = 3600. As expected the mean number of errors (and the ratio of bad/good bits) in a burst decreases with increasing guardspace.

Using the standard 1200-bit block and a block guardspace of 10 blocks there is as high as 5% probability of getting a block-burst extending to 10 blocks or more (when the channel is in the Red error mode).

Our opinion of this work is contained in Section VIII which also lists a few problems indicating the line future investigations should follow.

Section II

CHANNEL MODEL AND PARAMETER ESTIMATES

(i) Criterion for Choosing a Model

The histogram for the thirty-one runs of the 4.8 kbps high-speed data is shown in Figure 1 below. The error-free (gap) lengths are represented on the X-axis and their frequencies in the 31 runs, that is the number of times a gap of length X appears, on the ordinate. For example, the number of consecutive errors (at $X = 1$) is 17,149 while the number of times gaps of lengths $100 \le X \le 499$ appears is 652. Actually a gap length as shown on the histogram includes the position of the error bit that ends the gap. Thus to get the number of gaps of length 500, say, we should read the ordinate at the point $X = 501$.

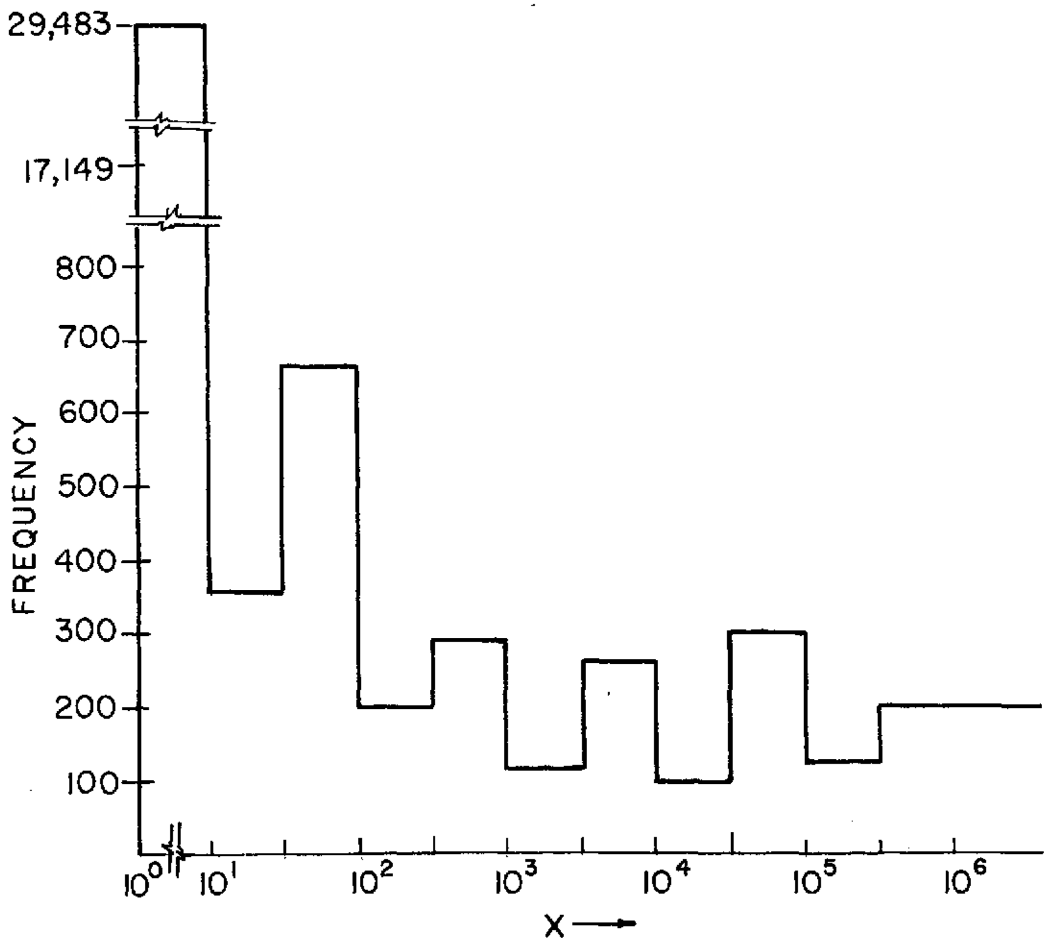

Figure 1. Histogram for the 4.8 kbps high-speed data.

The common feature of the 4800 bps high-speed and the 50 kbps wideband data is the way the errors tend to cluster together. Long gaps of error-free transmissions are followed by up to four seconds (or more) of sputtering errors. Within these "bursts" of errors there are intervals of good data. The problem is to construct a model for such a channel from which to derive statistics for error correction (both forward error correction and feedback retransmission).

Two broad classes of models have been proposed for these channels depending on what have been considered their main features. These are the Independent Gap or Pareto model and the Markov model. The Pareto model, so called because it assumes that successive gaps are approximately independent, was espoused by Berger and Mandelbrot in [3]. Pareto distribution was suggested for fitting the gaps. Earlier on, a Finite State Markov Chain had been suggested by Gilbert [4] for fitting the error sequence on such channels. The reason was that the gaps in the data seem to combine the Markovian property with the independent gap property. This class of models is called the Markov model. In the general case when the error clusters are different for different phases of the channel and hence more than one error state is used in the Markov model no assumption of independent gap distribution is made (see the generalization of Gilbert's model by Berkovits, Cohen, and Zierler in [5]). Up to now the choice of which model to use has not been based on an explicit criterion.

We shall briefly review both models and give a criterion for deciding which applies to a given set of data.

a. Independent Gap (Pareto) Models

Let $\{z_n\}$ be the error sequence, i.e., $z_n = 1$ if the $n^{\underline{th}}$ bit is in error and 0 otherwise. Let

$$V(k) = P(0^k 1|1); \qquad k \geq 0 \tag{1}$$

be the gap distribution where k is the gap length. Let

$$U(t) = \sum_{k \geq t} V(k) \ .$$

Thus

$$U(t) = P(\text{gap of length } k \geq t) \ .$$

Berger and Mandelbrot in [3] provided evidence, from the data they used, that successive gaps are approximately independent and suggested using Pareto distribution $t^{-\alpha}$ for the gap distribution. That is, they put

$$U(t) = t^{-\alpha} ; \qquad 0 < \alpha < 1 \tag{2}$$

or

$$P(\text{gap of length } k < t) = 1 - t^{-\alpha}$$

with probability density function $\alpha t^{-1-\alpha}$. For this range of α the $n^{\underline{th}}$ moment $\alpha \sum t^{n-1-\alpha}$ does not exist for any finite n. So t is restricted to some interval $0 < \delta \leq t \leq L < \infty$ and (2) is reduced to a three parameter model

$$U(k) = \begin{cases} 1 & k < \delta \\ (k/\delta)^{-\alpha} & \delta \leq k \leq L \\ 0 & L < k \end{cases} \ . \tag{3}$$

It is convenient to use

$$\log U(k) = \begin{cases} 0 & k < \delta \\ \alpha \log \delta - \alpha \log k & \delta \le k \le L \\ -\infty & L < k \end{cases} \quad . \tag{4}$$

From (4) it is not difficult to calculate the average bit error rate, P_1, the block error rate, the ratio of error probabilities for two different block lengths, etc. For example:

$$P_1 = \frac{(1-\alpha)\delta^{-\alpha}}{L^{1-\alpha}-\alpha} \quad . \tag{5}$$

It is also a straightforward matter to estimate the parameters of this model. The best fitting straight line for k between δ and L has slope α. The intercept at probability one occurs for a gap length $\delta \le 1$ and L can be estimated from relation (5), P_1 having been obtained from data. See Sussman [6] for further details.

But it is the shape of the graph of (4) between δ and L that we shall dwell on here. Within this interval $\log U(k)$ is always a straight line, so that any channel whose empirical $\hat{U}(k)$ cannot be fitted with a straight line on the log-log plot cannot be modelled by the Pareto distribution. For the GCF, Figures 4, 5 and 6 show that $\log \hat{U}(k)$ is convex for $k \le 2000$ and then becomes approximately a straight line. The interval of convexity of $\log \hat{U}(k)$ depends on the error mode; for the Green (low bit error rate) group $k \le 150$ and it increases as we enter the Red error group. Thus we reject the Pareto model for the GCF.

b. Markov Model

Gilbert's original model shown in Figure 2 consists of two states G and B. The channel alternates between the 'good' state G and the 'bad' state B according to a set of transition probabilities as shown. Errors may occur only in state B with some probability $0 < h < 1$. Transitions between states G and B plus the possibility of sojourn in either state (with probabilities Q and q, respectively) generate the bursts. Occurrence of an error implies the channel has returned to the state B and the process begins anew. Therefore successive gaps are independent and the model has both the Markov and independent gap property.

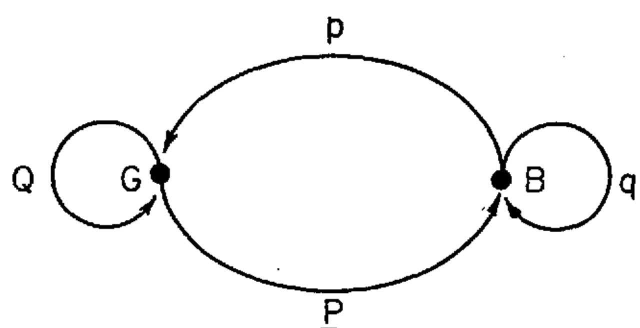

Figure 2. Gilbert Model

Gilbert showed, among other things, that

$$U(k) = ML^k + NJ^k; \quad k \geq 0 \tag{6}$$

where M, N, J, L all depend only on the model parameters and $0 < L \ll J < 1$.

Hence log U(k) is NOT a straight line for small to moderately large values of k. For sufficiently large k, log U(k) behaves like $c_1 + k \log J$, a straight line, where c_1 is a constant. This more closely resembles the shape of log U(k) shown in Figures 4, 5 and 6.

We shall therefore construct a Markov model for the GCF. The mant modes of the histogram (Figure 1) convince us that neither the Gilbert model nor its two-state generalization Zierler, et al [5] can provide adequate fit for our data. See also our earlier attempt in [2].

Let us now detail the procedure taken to construct a Markov model from the histogram.

(ii) The Model and Its Variations

We want to choose a "natural" model suggested by the histogram in the following way. Let us make the assumption that whenever an error occurs, the behavior of the channel at the time is independent of how good the channel was prior to that time. In other words the behavior of the channel each time it enters the bursty state is statistically the same irrespective of how long the time has been since it last showed this burst phenomenon. (In this report as in Reference 2 we use the same definition of burst, i.e. as a sequence beginning and ending with an error, separated from the nearest predeeding and following error by a gap of no less than a given length, say G - the guardspace and containing within it no gap of length equal to or greater than this guardspace.) Each time the channel enters a burst, that is each time we observe an error after a long gap, the length of the burst, the number of errors within it and the distribution of these errors are therefore all independent of what had gone on prior to the occurrence of this phenomenon. We can therefore represent distinct groups of gap lengths by distinct states of the channel and indicate the beginning of a burst by a return to a single error state from states representing long enough gaps. The short gaps and consecutive errors within a burst are then represented by transitions between this error state and those states representing appropriately short gaps. We shall make this concept more precise as we go along.

With respect to this "natural" way of constructing a model for the channel it would be necessary to represent each mode of the histogram by at least a state each of which connects to a single error state. We would therefore represent gaps of the following range of lengths corresponding to the modes of the histogram by distinct states:

$$X = 1; \quad 2 \le X \le 49, \ 100 \le X \le 499; \quad 1000 \le X \le 499;$$
$$10{,}000 \le X \le 49999; \quad 100{,}000 \le X \le 499999 \quad \text{and} \quad X \ge 10^6,$$

a total of seven states for the channel. But there are a number of objections to having so many states for the channel. These include the fact that a model with so many states may be unwieldy to analyze and even if we succeed in doing the analysis, such a model would be of very little practical use. A model should not be more complicated to understand than the phenomenon it is designed to explain!

The five state model we found to give acceptably good fit is shown in Figure 3 below. State B is the error state which connects to the perfectly good states G_1, G_2, G_3 and G_4. The good states represent gap lengths

$$X \ge 10^5, \ 1100 \le X \le 99999, \ 50 \le X \le 1099; \quad 2 \le X \le 49 \qquad (7)$$

respectively. These interval boundaries were determined from the histogram. All errors occur in state B, consecutive errors occurring with indicated probability $0 < q < 1$. Short bursts represent transitions between states B and G_4. Varying gap lengths are represented by transitions between state B and G_1, G_2, G_3; the very long gaps indicating the process is in state G_1.

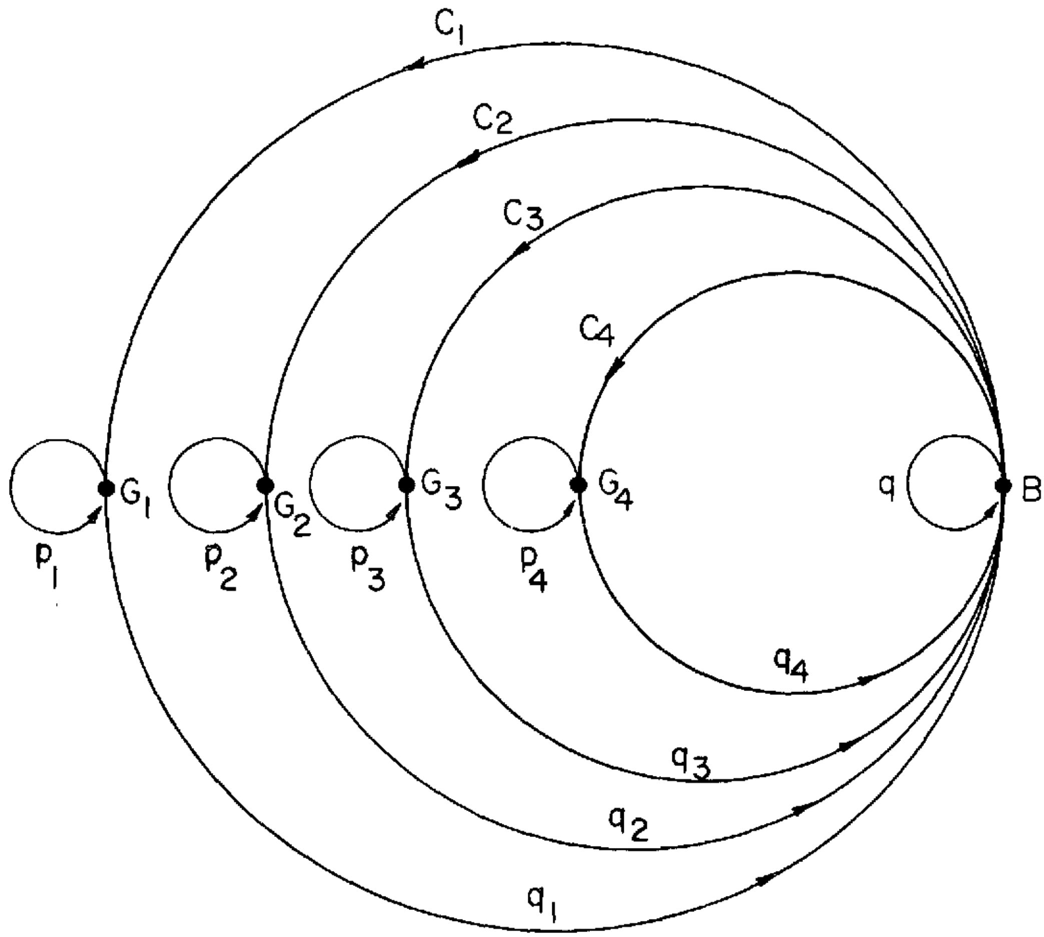

Figure 3. The five state model for the GCF

The model is not unlike four workmen with varying degrees of efficiency $p_1 > p_2 > p_3 > p_4$ employed to maintain a system. We agree to call a workman and his efficiency rate by the same name. Each time the system breaks down (in state B) any one of the four workmen is called upon to do the repairs, workman p_j being called with probability c_j; $j = 1, 4$. The probability is q that the maintenance supervisor will not call on any one of the workmen immediately the system breaks down. If he calls however, the length of time after the repairs are done for which the system remains in working condition is proportional to the workman's efficiency. In other words if workman p_j is called upon the chances are q_j that the system will not stay in working condition the next unit of time. Thus the lower the workman's efficiency the higher his q_j, $j = 1, 4$. (This analogy was suggested by E. C. Posner.)

In terms of transitions, p_j is the probability that the process stays in state G_j; c_j is the probability that the system moves from state B to state G_j in one step while q_j is the probability of returning from state G_j to state B in a single step, j = 1, 4; q is the probability of remaining in state B. We denote the one-step transition probability of going from state i to a state k by $P(k|i)$. Thus

$$\begin{aligned} P(G_j|G_j) &= p_j \\ P(G_j|B) &= c_j \qquad\qquad j = 1,4. \\ P(B|G_j) &= q_j = 1 - p_j \\ P(B|B) &= q = 1 - \sum c_j \end{aligned}$$

The physical explanation given above is not really acceptable as a close scrutiny would reveal. For it is fairly well-known that the error-causing mechanism on the channel does not reverse the bit each time an error occurs, a fact which we seem to ignore in our model in which we allow errors to occur with probability one each time the process reaches the state B. We hasten to point out however that a model using a state $\overline{B}$ in which errors occur with some probability $0 < h < 1$ instead of B can be made to be mathematically equivalent to our model by appropriately increasing the number of good states and adjusting the corresponding transition probabilities. Moreover, introducing such a state $\overline{B}$ would involve unnecessary complications in the analysis.

Even then the five-state model rather over-simplifies the actual channel. For instance we have assumed that it is possible to fit a channel with error rate varying between 0 and 10^{-3} and exhibiting three distinct error modes by a single stationary model. If the model performs well at high error rates it cannot be expected to depict the channel in the Green error mode. For when other users of the GCF come onto or drop off the line, the characteristics of the channel change significantly*. A realistic model must incorporate such changes.

One such model can be obtained as a generalization of our five state model. Instead of the p_j, c_j, $j = 1, 4$, being a fixed set of parameters we use

$$\{p_j^{(m)},\ c_j^{(m)},\ j = 1, 4,\ \ m = (\text{Red, Amber, Green})\} \ . \tag{8}$$

That is, we use a separate set of parameters for each of the error modes, much as we have done in this study, but further incorporate the varying line conditions caused by users coming onto and dropping off the line. The number of users on the channel at any given time can be modelled by the Poisson distribution, $P(\lambda)$, with some parameter λ, and the times between changes in the line condition then follow the exponential distribution. This means the line condition changes according to the Poisson distribution, the times between these changes following the exponential distribution. When a change does occur it can be to only one of the error modes Red, Amber or Green. The parameter λ of the Poisson distribution can be identified with the mean rate of user arrivals or the mean number of users per unit time. The estimation procedure for parameters in (8) is the same as is used in this study.

Because of its simplicity we have decided to use the stationary five-state model however. As a fitting model for the highest to moderate error mode (Red and Amber) our results bear us out.

Before we talk about how good a fit the model gives to the data let us take a look at the estimation procedure employed.

---

*We are grateful to L. R. Welch for discussions leading to this understanding.

(iii) Estimation of Parameters

In this section the procedure for estimating the model parameters p_j and c_j, $j = 1, 4$ from the data will be reviewed and the method for getting the Maximum Likelihood Estimates (MLE) indicated. For detailed analysis of the MLE method the reader is referred to Appendix I.

Let ℓ_j = number of times the process enters state j, $j = 1, 4$.

k_{ji} = the length of gap i, $i = 1, \ell_j$ in state j, $j = 1, 4$.

$\bar{k}_j$ = the threshold to state j or the minimal gap length determining state j, $j = 1, 4$.

N_e = the number of errors in the run.

N_{11} = cardinality of $[x = 1]$ or the number of occurrences of gaps of length zero in the run.

Then it can be shown (see Appendix I) using the method of maximum likelihood, that

$$\hat{p}_j = \frac{\sum_{i=1}^{\ell_j} k_{ji} - \ell_j \bar{k}_j}{\sum_{i=1}^{\ell_j} k_{ji} - \ell_j \bar{k}_j + \ell_j}$$

$$\hat{q}_j = 1 - \hat{p}_j \qquad j = 1, 4 . \tag{9}$$

$$\hat{c}_j = \frac{\ell_j}{N_e}$$

$$\hat{q} = \frac{N_{11}}{N_e}$$

The easiest way to understand the above expressions is to consider $\hat{p}_j$ as the proportion of time spent in state j as a fraction of the sum total of time spent in j and the number of times the process enters j; $\hat{c}_j$ as the number of times it enters state j as a fraction of the total number of times it crosses state B.

The above estimates we call raw estimates because they are obtained directly from the data using the gap length intervals stated in (7). These estimates are shown in Table 2(a). Table 2(b) contains the raw estimates for the 50 kbps wideband data for the same gap intervals used above for the 4.8 kbps high frequency data. The reader is asked to refer to Reference [2] for a description of and histogram for the 50 kbps data.

Let us now indicate how the optimum set of model parameters are obtained from these raw estimates.

Denote the probability of getting k error-free bits between a given error and the one immediately following it (i.e., a gap of length k) by $V(k)$:

$$V(k) = P(0^k 1|1)$$

where $\{0^k 1|1\}$ is the event that a given initial error is followed by a gap of length k. Then

$$\begin{aligned} V(k) &= P(0^k|1) - P(0^{k+1}|1) \\ &= U(k) - U(k+1) \end{aligned}$$

where we have denoted $P(0^k|1)$ by $U(k)$ and $\{0^k|1\}$ is the event that a given error is followed by at least k error-free bits. Also let us represent the sequence of noise digits by $z = \{z_n\}$ in which $z_n = 1$ if the n^{th} digit is in error and $z_n = 0$ otherwise.

Table 2 (a). Raw estimates of $\underset{\sim}{P}$ and $\underset{\sim}{C}$ for the 4.8 kbps high frequency data

1	$\hat{p}$	.9999987	.9999130	.7890661	.9956507
	$\hat{c}$	.2048373E-02	.9611595E-02	.6137241	.5924525E-01
2	$\hat{p}$	.9999994	.9999866	.7843750	.0000000
	$\hat{c}$	.1071429	.8928571E-02	.6160714	.0000000
3	$\hat{p}$	.9999998	.9999731	.6190476	.9979798
	$\hat{c}$	.1265823E-01	.6329115E-02	.5738397	.4219409E-02
4	$\hat{p}$	.9999986	.9999738	.7050070	.9942991
	$\hat{c}$	.2223089E-01	.1404056E-01	.5721529	.6630264E-02
5	$\hat{p}$	.9999985	.9999675	.7701746	.9915816
	$\hat{c}$	.404459E-01	.2872216E-01	.5709261	.3575616E-01
6	$\hat{p}$	.9999998	.9999557	.7037037	.9978564
	$\hat{c}$	.1866252E-01	.1555210E-01	.5598756	.6220840E-02
7	$\hat{p}$	.9999993	.9999381	.7216154	.9977547
	$\hat{c}$	.1331893E-01	.1979842E-01	.6054716	.1043916E-01
8	$\hat{p}$	.9999998	.9996789	.7849265	.9982025
	$\hat{c}$	.6000000E-01	.4500000E-01	.5850000	.1500000E-01
9	$\hat{p}$	.9999998	.0000000	.9047619	.0000000
	$\hat{c}$	.6666667	.0000000	.6666667	.0000000
10	$\hat{p}$	.9999998	.9978070	.7634730	.9950000
	$\hat{c}$	.7692307E-01	.7692307E-02	.6076923	.3076923E-01
11	$\hat{p}$	.9999998	.9999652	.7120360	.9934446
	$\hat{c}$	.5980860E-02	.3588517E-02	.6124402	.1315789E-01
12	$\hat{p}$	.9999999	.9992025	.5198413	.0000000
	$\hat{c}$	.9433962E-02	.4716981E-02	.5707547	.0000000
13	$\hat{p}$	.9999999	.9878049	.5378705	.9967664
	$\hat{c}$	.9459458E-02	.1351351E-02	.5689189	.5405404E-02
14	$\hat{p}$	.9999996	.9999225	.7606132	.9961320
	$\hat{c}$	.1815431E-01	.1966717E-01	.6142209	.2571861E-01

Table 2(a) Cont'd.

15	$\hat{p}$	.9999981	.9999480	.7990319	.9978045
	$\hat{c}$	.3760163E-01	.6402439E-01	.6117886	.1626016E-01
16	$\hat{p}$	.9999994	.9999822	.8977956	.0000000
	$\hat{c}$	.3200000	.1333333E-01	.6800000	.0000000
17	$\hat{p}$	.9999996	.9998726	.7000611	.9971001
	$\hat{c}$	.2729528E-01	.1736973E-01	.6091812	.7444169E-02
18	$\hat{p}$	.9999998	.0000000	.6358382	.9915493
	$\hat{c}$	.2690583E-01	.0000000	.5650224	.1345291E-01
19	$\hat{p}$	.9999995	.9999472	.7892157	.9981061
	$\hat{c}$	.4699739E-01	.1566580E-01	.5613577	.3655352E-01
20	$\hat{p}$	.9999985	.9999343	.8945498	.9974043
	$\hat{c}$	.1307693	.1461539	.6846154	.3076923E-01
21	$\hat{p}$	.9999991	.9998994	.6274128	.9959416
	$\hat{c}$	.3273322E-02	.4091654E-02	.5739225	.8728858E-02
22	$\hat{p}$	.9999973	.9998945	.8599168	.9978881
	$\hat{c}$	.8982038E-01	.6586826E-01	.6047904	.5988024E-01
23	$\hat{p}$	.9999995	.9999318	.6606772	.9960402
	$\hat{c}$	.3120125E-02	.5512219E-02	.5951118	.8840352E-02
24	$\hat{p}$	.9999992	.9999160	.8674912	.9953895
	$\hat{c}$	.2945302E-01	.2805049E-01	.7363254	.5329593E-01
25	$\hat{p}$	.9999992	.9999613	.7558528	.0000000
	$\hat{c}$	.3018109E-01	.1006036E-01	.5875251	.0000000
26	$\hat{p}$	.9999982	.9999579	.7334171	.9961382
	$\hat{c}$	.2268431E-01	.4190296E-01	.6014493	.1102709E-01
27	$\hat{p}$	.9999985	.9999592	.7450024	.9939882
	$\hat{c}$	.1987930E-01	.1668442E-01	.5750799	.2094427E-01
28	$\hat{p}$	.9999992	.9999778	.5950413	.9983525
	$\hat{c}$	.1706037E-01	.1181103E-01	.5787402	.1312336E-02
29	$\hat{p}$	.9999997	.9999112	.6872549	.9956168
	$\hat{c}$	.1717557E-01	.9541985E-02	.6087787	.1335878E-01

Table 2(b). Raw estimates of $\underset{\sim}{P}$ and $\underset{\sim}{C}$ for the 50 kbps wide-band data.

1	$\hat{p}$	.9999995E+00	.9999342E+00	.6041746E+00	.9922753E+00
	$\hat{c}$	.8959681E-02	.8461922E-02	.5191638E+00	.5475361E-02
2	$\hat{p}$	.9999997E+00	.9999660E+00	.6134831E+00	.9913206E+00
	$\hat{c}$	.6868917E-02	.1717230E-02	.4922725E+00	.8013736E-02
3	$\hat{p}$	.9999998E+00	.9999250E+00	.6000000E+00	.9951378E+00
	$\hat{c}$	.1130435E-02	.2608696E-02	.4434783E+00	.2608696E-02
4	$\hat{p}$	.9999999E+00	.0000000E+00	.8130435E+00	.9677419E+00
	$\hat{c}$	.4597701E-01	.0000000E+00	.4942529E+00	.1149425E-01
5	$\hat{p}$	.9999999E+00	.0000000E+00	.6063218E+00	.0000000E+00
	$\hat{c}$	.1923077E-01	.0000000E+00	.5269231E+00	.0000000E+00
6	$\hat{p}$	.9999999E+00	.9999100E+00	.4768519E+00	.9813084E+00
	$\hat{c}$	.1310044E-01	.4366811E-02	.4934498E+00	.8733623E-02
7	$\hat{p}$	.9999999E+00	.0000000E+00	.5869565E+00	.9740260E+00
	$\hat{c}$	.2259887E-01	.0000000E+00	.4293785E+00	.1129944E-01
8	$\hat{p}$	.9999999E+00	.0000000E+00	.6193354E+00	.0000000E+00
	$\hat{c}$	.1716738E-01	.0000000E+00	.5407726E+00	.0000000E+00
9	$\hat{p}$	.9999999E+00	.0000000E+00	.5322581E+00	.9929078E+00
	$\hat{c}$	.2016129E-01	.0000000E+00	.4677419E+00	.4032258E-02
10	$\hat{p}$	.9999998E+00	.9999595E+00	.5989176E+00	.9962997E+00
	$\hat{c}$	.9732362E-02	.8110300E-02	.5409570E+00	.3244120E-02
11	$\hat{p}$	.9999998E+00	.9999375E+00	.5638418E+00	.9960806E+00
	$\hat{c}$	.5791504E-02	.5148005E-02	.4967825E+00	.4504506E-02
12	$\hat{p}$	.9999998E+00	.9998188E+00	.8231986E+00	.9936668E+00
	$\hat{c}$	.6577650E-03	.3617707E-02	.6534895E+00	.3341446E-01

If we write the model transition matrix as

$$M = \begin{pmatrix} p_1 & 0 & 0 & 0 & 1-p_1 \\ 0 & p_2 & 0 & 0 & 1-p_2 \\ 0 & 0 & p_3 & 0 & 1-p_3 \\ 0 & 0 & 0 & p_4 & 1-p_4 \\ c_1 & c_2 & c_3 & c_4 & q \end{pmatrix}$$

then the probability of getting the particular pattern of errors observed on the channel $P(M, z)$ is given by:

$$P(M, z) = P_1 U(\ell) U(L) \prod_{j=1}^{N_e - 1} V(\ell_j) \tag{10}$$

where P_1 = theoretical bit error-rate.

ℓ = number of the error-free bits before the first error in the run.

L = number of the error-free bits after the last error in the run,

and here

ℓ_j = length of $j^{\underline{th}}$ gap, $j = 1, N_e - 1$. It is desired to maximize $P(M, z)$ subject to some restrictions. It is easy to show that:

$$P_1 = \frac{1}{\hat{c}}$$

where

$$\hat{c} = 1 + \sum_{i=1}^{4} \frac{c_i}{1-p_i} ,$$

$$U(k) = \sum_{i=1}^{4} c_i p_i^{k-1} \; ; \qquad k \geq 0 \tag{11}$$

and

$$V(k) = \sum_{i=1}^{4} c_i (1 - p_i) p_i^{k-1}$$

The set p_i', c_i', $i = 1, 4$, maximizing $P(M, z)$ also maximize

$$\log \frac{P(M,z)}{p_1} = \log \sum_i c_i p_i^{\ell-1} + \log \sum_i c_i p_i^{L-1} + \sum_{j=1}^{N_e-1} \log \sum_i c_i (1-p_i) p_i^{\ell_j -1} \tag{12}$$

p_i' is given by

$$p_i' = \frac{\ell \dfrac{\bar{c}_i p_i^{\ell}}{(1-p_i)\alpha_1} + \dfrac{\bar{c}_i p_i^{\ell+1}}{(1-p_i)^2 \alpha_1} + L \dfrac{\bar{c}_i p_i^{L}}{(1-p_i)\alpha_L} + \dfrac{\bar{c}_i p_i^{L+1}}{(1-p_i)^2 \alpha_L} + \sum_{j=1}^{N_e-1} \ell_j \dfrac{\bar{c}_i p_i^{\ell_j}}{\alpha_j}}{(\ell+1) \dfrac{\bar{c}_i p_i^{\ell}}{(1-p_i)\alpha_1} + \dfrac{\bar{c}_i p_i^{\ell+1}}{(1-p_i)^2 \alpha_1} + (L+1) \dfrac{\bar{c}_i p_i^{L}}{(1-p_i)\alpha_L} + \dfrac{\bar{c}_i p_i^{L+1}}{(1-p_i)^2 \alpha_L} + \sum_j (\ell_j+1) \dfrac{\bar{c}_i p_i^{\ell_j}}{\alpha_j}}$$

and if

$$\bar{c}_i' = \frac{\left\{ \dfrac{\bar{c}_i p_i^{\ell}}{(1-p_i)\alpha_1} + \dfrac{\bar{c}_i p_i^{L}}{(1-p_i)\alpha_L} + \sum_j \dfrac{\bar{c}_i p_i^{\ell_j}}{\alpha_j} \right\}^2}{(N_e+1) \left\{ (\ell+1) \dfrac{\bar{c}_i p_i^{\ell}}{(1-p_i)\alpha_1} + \dfrac{\bar{c}_i p_i^{\ell+1}}{(1-p_i)^2 \alpha_1} + (L+1) \dfrac{\bar{c}_i p_i^{L}}{(1-p_i)\alpha_L} + \dfrac{\bar{c}_i p_i^{L+1}}{(1-p_i)^2 \alpha_L} + \sum_{j=1}^{N_e-1} (\ell_j+1) \dfrac{\bar{c}_i p_i^{\ell_j}}{\alpha_j} \right\}}$$

then

$$c_i' = \frac{\bar{c}_i' p_i'}{1-p_i'}$$

where

$$\bar{c}_i = \frac{c_i(1-p_i)}{p_i}; \quad \alpha_1 = \sum_{i=1}^{4} \frac{\bar{c}_i p_i^{\ell}}{1-p_i}; \quad \alpha_L = \sum_{i=1}^{4} \frac{\bar{c}_i p_i^{L}}{1-p_i} \tag{13}$$

and

$$\alpha_j = \sum_{i=1}^{4} \bar{c}_i p_i^{\ell_j} \quad .$$

To obtain the estimates p_i' and c_i' the raw estimates $\hat{p}_i$ and $\hat{c}_i$ in (9) are used in (13) as first approximations for p_i and c_i. Using the p_i' and c_i' thus obtained as initial estimates for p_i and c_i, the above procedure is repeated on a digital computer to give a new set of maximizing parameters, c_i'' and p_i''. This iterative method is repeated until a degree of stability sufficient for curve fitting purposes is achieved. See Baum and Welch [7] for further details of this iterative method.

In general, note that it is possible for distinct transition matrices M to yield the same z-process and thus the same P(M, z). Let us call all such matrices equivalent. For example, as shown by Blackwell and Koopmans [8], any two matrices M_1, M_2 yielding the same V(k) are equivalent. It suffices for our purpose therefore to find any one member M in this equivalence class, i.e., any transitions matrix M yielding a critical point of P(M, z).

Starting with the raw estimates in (7), two-hundred iterations on the computer yield the maximizing parameters shown in Table 3(a) for the 4.8 kbps data and Table 3(b) for the 50 kbps data.

(iv) Curve Fitting and Goodness-of-Fit Test

A basic statistic in our model is the gap distribution V(k) because the process renews itself each time it reaches state B. In other words the occurrence of an error is the renewal event which wipes out the memory of the

Table 3(a) Maximum likelihood estimates (MLE) of $\underset{\sim}{P}$ and $\underset{\sim}{C}$ for the 4.8 kbps HF dataline

1	p'	.9999953950E+00	.9968072943E+00	.5717325689E+00	.9106591544E+00
	c'	.8981760809E-02	.6745001194E-01	.4203317556E+00	.1713520791E+00
2	p'	.9999995475E+00	.9999917204E+00	.7619005941E+00	.0000000000E+00
	c'	.1047399680E+00	.1030998001E-01	.6742438692E+00	.0000000000E+00
3	p'	.9999998554E+00	.9999535030E+00	.5459808331E+00	.8961057897E+00
	c'	.1299700189E-01	.1017617426E-01	.5091828334E+00	.3962966780E-01
4	p'	.9999986050E+00	.9999567087E+00	.5519996996E+00	.9227811211E+00
	c'	.2665215750E-01	.1285778017E-01	.4786566726E+00	.8614839406E-01
5	p'	.9999984313E+00	.9999198994E+00	.5999923080E+00	.9601877004E+00
	c'	.5125120746E-01	.2359833744E-01	.4627144092E+00	.1478188950E+00
6	p'	.9999997989E+00	.9999390274E+00	.5425012681E+00	.9106503818E+00
	c'	.1905607045E-01	.2137889279E-01	.4755474738E+00	.7556629062E-01
7	p'	.9999993455E+00	.9999282191E+00	.6780365331E+00	.9972714645E+00
	c'	.1504603869E-01	.1804396336E-01	.6435550889E+00	.1771507139E-01
8	p'	.9999998185E+00	.9997777749E+00	.7532956976E+00	.9992440385E+00
	c'	.5989127950E-01	.3638922970E-01	.6622567025E+00	.2454708225E-01
9	p'	.9999999229E+00	.0000000000E+00	.9130433718E+00	.0000000000E+00
	c'	.7500019813E+00	.0000000000E+00	.4565198227E+00	.0000000000E+00
10	p'	.9999998558E+00	.9979309774E+00	.5581959110E+00	.8535666459E+00
	c'	.7636851511E-01	.4016395427E-01	.3200710332E+00	.2645899237E+00
11	p'	.9999999010E+00	.9999315091E+00	.5970954356E+00	.9689291927E+00
	c'	.6033735685E-02	.7273912417E-02	.5537134181E+00	.5750297725E-01
12	p'	.9999999537E+00	.0000000000E+00	.5454545438E+00	.0000000000E+00
	c'	.1408451107E-01	.0000000000E+00	.5377720829E+00	.0000000000E+00
13	p'	.9999999219E+00	.9006192752E+00	.5363738531E+00	.9979544427E+00
	c'	.9448104969E-02	.9756278883E-02	.5216192007E+00	.7212404850E-02
14	p'	.9999995490E+00	.9992407306E+00	.6163191023E+00	.9081214215E+00
	c'	.2258593837E-01	.3780571254E-01	.4839834401E+00	.1401222301E+00

Table 3(a) Cont'd

15	p'	.9999981532E+00	.9999391884E+00	.7730621732E+00	.9987174063E+00
	c'	.4522522687E-01	.5680939256E-01	.6806276662E+00	.1750866214E-01
16	p'	.9999996082E+00	.9999983751E+00	.9072722727E+00	.0000000000E+00
	c'	.2151236379E+00	.1138326308E+00	.6088191268E+00	.0000000000E+00
17	p'	.9999996682E+00	.9999115533E+00	.6802853951E+00	.9991929634E+00
	c'	.2775160346E-01	.1144913912E-01	.6442882584E+00	.1370192609E-01
18	p'	.9999998824E+00	.0000000000E+00	.5245536494E+00	.9819256259E+00
	c'	.2679576549E-01	.0000000000E+00	.4839356004E+00	.4972253088E-01
19	p'	.9999994394E+00	.9990228370E+00	.5303564611E+00	.9310260595E+00
	c'	.5485257315E-01	.4249021311E-01	.4000401305E+00	.1380993112E+00
20	p'	.9999986382E+00	.9999380983E+00	.9033324592E+00	.9993548564E+00
	c'	.1476422873E+00	.9198748428E-01	.6266307466E+00	.6663347228E-01
21	p'	.9999990224E+00	.9996616399E+00	.5619055084E+00	.9640913984E+00
	c'	.4247087939E-02	.7602542515E-02	.5396837177E+00	.2670050566E-01
22	p'	.9999978277E+00	.9998756379E+00	.8463196387E+00	.9990676567E+00
	c'	.1028148355E+00	.2779428687E-01	.6637432406E+00	.8503767638E-01
23	p'	.9999992919E+00	.9992437394E+00	.5704438190E+00	.9266249133E+00
	c'	.4881179917E-02	.1105176014E-01	.5328776355E+00	.4625063032E-01
24	p'	.9999991211E+00	.9991934723E+00	.8367628756E+00	.9732659686E+00
	c'	.3813345289E-01	.4954637321E-01	.6831363057E+00	.9331328594E-01
25	p'	.9999991663E+00	.9974426078E+00	.6732643675E+00	.0000000000E+00
	c'	.3647266453E-01	.1898180839E-01	.6358960597E+00	.0000000000E+00
26	p'	.9999980558E+00	.9999385496E+00	.5763860101E+00	.9029800511E+00
	c'	.2903556659E-01	.4430774668E-01	.4497416415E+00	.1321740257E+00
27	p'	.9999982226E+00	.9996096551E+00	.5650661490E+00	.9504853272E+00
	c'	.2792114797E-01	.1840825642E-01	.4777731974E+00	.1027912530E+00
28	p'	.9999992817E+00	.9999792257E+00	.5938876271E+00	.9991027533E+00
	c'	.2008279598E-01	.8168327315E-02	.5756273053E+00	.2493552415E-02
29	p'	.9999997544E+00	.9999128595E+00	.6438181406E+00	.9932046896E+00
	c'	.1733256091E-01	.9945088923E-02	.6098647740E+00	.2528598892E-01

Table 3(b) MLE of $\underset{\sim}{P}$ and $\underset{\sim}{C}$ for the 50 kbps W-B dataline

1	p'	.9999995441E+00	.9999004627E+00	.5338684774E+00	.9676298396E+00
	c'	.9845264894E-02	.8769695517E-02	.5129089059E+00	.1997571724E-01
2	p'	.9999997415E+00	.9971774533E+00	.4894906586E+00	.9033394419E+00
	c'	.7454561974E-02	.9653846970E-02	.4614024583E+00	.3635649359E-01
3	p'	.9999998071E+00	.9990447277E+00	.4376002743E+00	.9044734152E+00
	c'	.1217685823E-01	.4290337444E-02	.4124613968E+00	.3706154357E-01
4	p'	.9999999611E+00	.0000000000E+00	.4879713708E+00	.9437002150E+00
	c'	.4545482678E-01	.0000000000E+00	.3854866564E+00	.1553020579E+00
5	p'	.9999999132E+00	.0000000000E+00	.5472163544E+00	.0000000000E+00
	c'	.2297855466E-01	.0000000000E+00	.5346421124E+00	.0000000000E+00
6	p'	.9999999758E+00	.9999166130E+00	.4725751340E+00	.9852605868E+00
	c'	.1305383315E-01	.4417244510E-02	.4576236982E+00	.1395796319E-01
7	p'	.9999999552E+00	.0000000000E+00	.4705709962E+00	.9770023759E+00
	c'	.2247210426E-01	.0000000000E+00	.4471357312E+00	.2670114001E-01
8	p'	.9999999121E+00	.0000000000E+00	.5342551169E+00	.0000000000E+00
	c'	.2518403344E-01	.0000000000E+00	.5208004170E+00	.0000000000E+00
9	p'	.9999999416E+00	.0000000000E+00	.4740011959E+00	.9852936294E+00
	c'	.2008078715E-01	.0000000000E+00	.4582302267E+00	.1299721968E-01
10	p'	.9999998113E+00	.9999481848E+00	.5207492063E+00	.8653899580E+00
	c'	.9890531754E-02	.1061418918E-01	.4790819180E+00	.5149835679E-01
11	p'	.9999999024E+00	.9999854610E+00	.5099146477E+00	.9965654103E+00
	c'	.3965602111E-02	.5268077950E-02	.4990851372E+00	.1196290223E-01
12	p'	.9999998159E+00	.9996824757E+00	.7499412224E+00	.9872709999E+00
	c'	.7559548997E-03	.6304445321E-02	.6837962356E+00	.8010502177E-01

past gap. Each gap is then an independent statistical sample from the distribution $V(k)$, $k \geq 0$. The occurrence of bursts is a direct consequence of the form of the gap distribution. We shall therefore assess the performance of our model by how good a fit $U(k)$, $k \geq 0$, a function of the gap distribution, gives to the empirical $\hat{U}(k)$ - both from the wideband and high frequency data. (Recall from (11) that

$$\sum_{i=1}^{4} c_i p_i^{k-1} = U(k)$$

$$= \sum_{j \geq k} V(j) \Big)$$

Figures 4, 5, and 6 are representative of fits obtained from the Green, Amber and Red groups respectively.

For purposes of error control it is more important to have very accurate prediction of error clusters when the gaps are short (high bit error rate) than during long intervals of error-free transmissions. This is why we have compared $U(k)$ and $\hat{U}(k)$ for $0 \leq k \leq 4000$.

Let $F(k)$ and $F_n(k)$ defined by (14) be the distribution functions associated respectively with $V(k)$ and $\hat{V}(k)$.

$$F(k) = \sum_{j=0}^{k-1} V(j)$$
$$F_n(k) = \sum_{j=0}^{k-1} \hat{V}(j) \qquad (14)$$

Then

$$|F(k) - F_n(k)| = \left|1 - \sum_{j=k}^{\infty} V(j) - \left(1 - \sum_{j=k}^{\infty} \hat{V}(j)\right)\right|$$

$$= |U(k) - \hat{U}(k)|$$

Hence

$$\rho = \max_{0 \leq k \leq 4000} |U(k) - \hat{U}(k)| \qquad (15)$$

is the maximum absolute difference between the model and measured (empirical) distribution functions in the range shown. For the Green group (Figure 4) $\rho = .035$; for Amber (Figure 5) $\rho = .015$; and for the Red (Figure 6) $\rho = .012$, which shows that the higher the bit error rate the smaller the ρ (the better the fit). This is seen from the respective graphs.

Now write

$$Y(k) = U(k) - \hat{U}(k); \quad 0 \leq k \leq 4000 .$$

For a very good fit one would expect $Y(k)$ to have very small mean, $\overline{Y}$, and mean-square-error, s_y. A zero mean would indicate that the $\hat{U}(k)$ is symmetrical about $U(k)$ and the small standard error as a measure of the deviation of $\hat{U}(k)$ from $U(k)$ indicates that the $\hat{U}(k)$ does not deviate too widely from $U(k)$. For the Green group $\overline{Y} = -.007$, $s_y = .003$; for the Amber $\overline{Y} = -.004$, $s_y = .002$ and for the Red $\overline{Y} = -.001$, $s_y = .0008$. ρ and $\overline{Y}$ for all the 29 error-runs are shown in Table 4(a) for the HF circuit and in Table 4(b) for the wide-band.

Because of the wide range of error rates observed in the sample runs $(0 - 10^{-3})$ it is clear that there is no way of constructing a single channel with

N	INDIVIDUAL RUNS			GROUPED RUNS			OVERALL CHANNEL		
	ρ	$\overline{Y}$	s_y	ρ	$\overline{Y}$	s_y	ρ	$\overline{Y}$	s_y
2464200	.022	-.0003	.002	.063	.003	.006	.107	-.015	.008
14430000	.012	-.001	.0008	.083	-.008	.007	.098	-.22	.007
63632400	.023	-.0005	.001	.076	-.003	.006	.091	-.022	.006
28970400	.049	-.005	.002	.120	-.036	.005	.101	-.02	.007
48342000	.029	-.003	.001	.055	-.017	.003	.040	.006	.007
5500800	.015	-.004	.003	.063	.015	.004	.079	.039	.005
51313200	.023	-.005	.002	.063	-.020	.004	.058	.004	.006
61706400	.015	-.0008	.002	.056	-.022	.002	.021	.002	.004
30902400	.025	-.004	.002	.081	-.044	.003	.065	-.021	.006
64060800	.018	-.001	.0006	.145	-.044	.005	.130	-.021	.008
28754400	.022	-.001	.003	.059	-.020	.005	.034	.004	.003
49009200	.009	-.0005	.004	.135	.045	.006	.151	.067	.004
61592400	.011	-.001	.002	.048	-.013	.002	.033	.011	.004
34022400	.027	-.002	.003	.068	.010	.009	.054	.034	.005
12943200	.017	-.005	.009	.670	.195	.030	.686	.219	.029
7084800	.027	-.005	.008	.336	.089	.025	.352	.113	.021
28753200	.029	-.002	.003	.035	-.001	.018	.120	.022	.016
19392000	.073	-.002	.004	.050	-.017	.002	.035	.007	.006
48927600	.031	-.006	.002	.033	.009	.002	.059	.033	.004
43287600	.024	-.002	.002	.035	-.018	.002	.023	.006	.004
19036800	.023	-.001	.001	.118	-.026	.005	.103	-.002	.008
28974000	.030	-.001	.002	.067	-.028	.003	.042	-.004	.005
21888000	.012	-.008	.004	.033	.016	.003	.119	.078	.008
55360800	.035	-.006	.007	.038	-.006	.016	.126	.065	.011
12952800	.514	-.416	.013	.721	.243	.021	.770	.305	.022
55497600	.025	-.007	.003	.037	-.017	.008	.109	.045	.004
21592800	.045	-.007	.003	.200	-.084	.006	.147	-.022	.008
42288000	.025	-.008	.003	.654	.229	.019	.704	.292	.020
340272000	.035	-.004	.002	.167	-.069	.004	.129	-.007	.007

Table 4(a). Curve fitting parameters for 4.8 kbps HF dataline

$Y = U(k) - \hat{U}(k)$

$\overline{Y}$ = Mean of Y, s_y = standard error of Y

$\rho = \max_{0 \le k \le 4000} |U(k) - \hat{U}(k)|$

N	INDIVIDUAL RUNS			OVERALL CHANNEL		
	ρ	$\bar{Y}$	S_y	ρ	$\bar{Y}$	S_y
39198003	0.0046	-0.0007	0.0004	0.118	0.0079	0.0072
42681997	0.0053	-0.0006	0.0008	0.133	-0.0009	0.0071
62300445	0.0079	-0.0008	0.0005	0.188	0.0038	0.0084
51451345	0.0289	-0.011	0.001	0.031	0.0265	0.0042
46043171	0.022	-0.008	0.0008	0.144	0.0068	0.0082
41468126	0.015	-0.004	0.0008	0.167	0.0046	0.0085
44651211	0.0247	-0.006	0.0008	0.162	0.0086	0.0078
44301174	0.046	-0.012	0.001	0.153	0.0043	0.0082
51389935	0.011	-0.004	0.0006	0.169	0.0077	0.0086
54343756	0.008	-0.002	0.0004	0.119	0.0085	0.0074
43441434	0.0195	-0.0007	0.001	0.155	0.00064	0.0078
52020411	0.079	-0.001	0.004	0.122	-0.0034	0.0051

Table 4(b). Curve fitting parameters for 50 kbps W-B dataline

$$Y = U(k) - \hat{U}(k)$$

$\bar{Y}$ = mean of Y, S_y = standard error of Y

$$\ell = \max_{0 \le k \le 4000} |U(k) - \hat{U}(k)|$$

transition matrix M which will give a good fit to each of the sample $\hat{U}(k)$, $k \geq 0$. One thing that can be done is to construct a channel for each of the Green, Amber and Red groups. By plotting averaged model $U(k)$ for each of these groups we shall get a better picture of what is happening at comparable error rates. As raw estimates for the optimum averaged parameter set for each group we use weighted averages of the parameters p_i' and c_i' which gives the critical points of $P(M, z)$ in each of the samples in that group, weighted according to the number of bits transmitted in the sample run. For example, if N_j represents the number of bits transmitted in sample j, j running over the number of samples in that group, then

$$\hat{\hat{p}}_i = \frac{\sum_j N_j p'_{ij}}{N} ; \quad N = \sum_j N_j ; \qquad i = 1, 4 \tag{16}$$

is the averaged raw estimate of p_i for the group and p'_{ij} is the estimate of p_i in sample j which maximizes $P(M, z)$.

Now let

$$\begin{aligned} \alpha_{1k} &= \sum_i \frac{\bar{c}_i p_i^{\ell_{1k}}}{1-p_i} \\ \alpha_{L_k} &= \sum_i \frac{\bar{c}_i p_i^{L_k}}{1-p_i} \\ \alpha_{jk} &= \sum_i \bar{c}_i p_i^{\ell_{jk}} \end{aligned} \tag{17}$$

where $k=1, \ldots$ runs over the samples in the group and

$j=1, \ldots, N_{ek}$ is the number of errors in sample k.

ℓ_{1k} = number of the error-free bits before the first error in sample k.

L_k = number of the error-free bits after last error in sample k.

ℓ_{jk} = $j^{\underline{th}}$ gap length in sample k.

Then if

$$D_1 = \sum_{k\geq 1}\left[\frac{\bar{c}_i p_i^{\ell_{1k}}}{(1-p_i)\alpha_{1k}} + \frac{\bar{c}_i p_i^{L_k}}{(1-p_i)\alpha_{L_k}} + \sum_{j=1}^{N_{ek}-1}\frac{\bar{c}_i p_i^{\ell_{jk}}}{\alpha_{jk}}\right]$$

$$D_2 = \sum_{k\geq 1}\left\{\frac{\ell_{1k}\bar{c}_i p_i^{\ell_{1k}}}{(1-p_i)\alpha_{1k}} + \frac{\bar{c}_i p_i^{\ell_{1k}+1}}{(1-p_i)^2\alpha_{1k}} + \frac{L_k\bar{c}_i p_i^{L_k}}{(1-p_i)\alpha_{L_k}} + \frac{\bar{c}_i p_i^{L_k+1}}{(1-p_i)^2\alpha_{L_k}} \quad \sum_{j=1}^{N_{ek}-1}\ell_{jk}\frac{\bar{c}_i p_i^{\ell_{jk}}}{\alpha_{jk}}\right\} \tag{18}$$

$$D = D_1 + D_2 ,$$

the optimum averaged parameter set for that group is given by:

$$p_i' = \frac{D_2}{D}$$

$$\bar{c}_i' = \frac{D_1^2}{\left(\sum_{k\geq 1} N_{\ell k}+n\right)D} \tag{19}$$

$$c_i' = \frac{\bar{c}_i' p_i'}{1-p_i'} .$$

Our notation here is as was used in (9), (10), and (11). n is the number of samples in the group. Two hundred iterations on the computer of expression (19)

Table 5a MLE of $\underset{\sim}{P}$ and $\underset{\sim}{C}$ for the Red, Amber and Green groups and overall channel; 4.8 kbps dataline

MLE (Optimum) Set for Group R (10^{-3}) (Red)

p	.9999982928E+00	.9975858132E+00	.9114607347E+00	.5645880886E+00
c	.6947437705E-02	.3811463333E-01	.1103252624E+00	.4707554813E+00

MLE for Group A (Amber 10^{-4})

p	.9999991048E+00	.9998733543E+00	.9278527620E+00	.5752459153E+00
c	.3038593488E-01	.2886455315E-01	.1214899946E+00	.4658393836E+00

MLE for Group G (Green 10^{-5})

p	.9999996962E+00	.9165101985E+00	.5635646054E+00	.5635646054E+00
c	.9148175377E-01	.1450558925E+00	.9796649509E-01	.3254327807E+00

Single Channel (Parameters) for 4.8 kbps Data

p	.9999989701E+00	.9987625506E+00	.9134863812E+00	.5668315129E+00
c	.2346445968E-01	.2937987288E-01	.1211068491E+00	.4617199786E+00

Table 5b MLE of $\underset{\sim}{P}$ and $\underset{\sim}{C}$ for 50 kbps Channel; M^k for $k = 6$

p	.9999997795E+00	.9027180023E+00	.9967973995E+00	.5138910937E+00
c	.6226522694E-02	.2027754955E+00	.2531177075E-01	.3822078635E+00

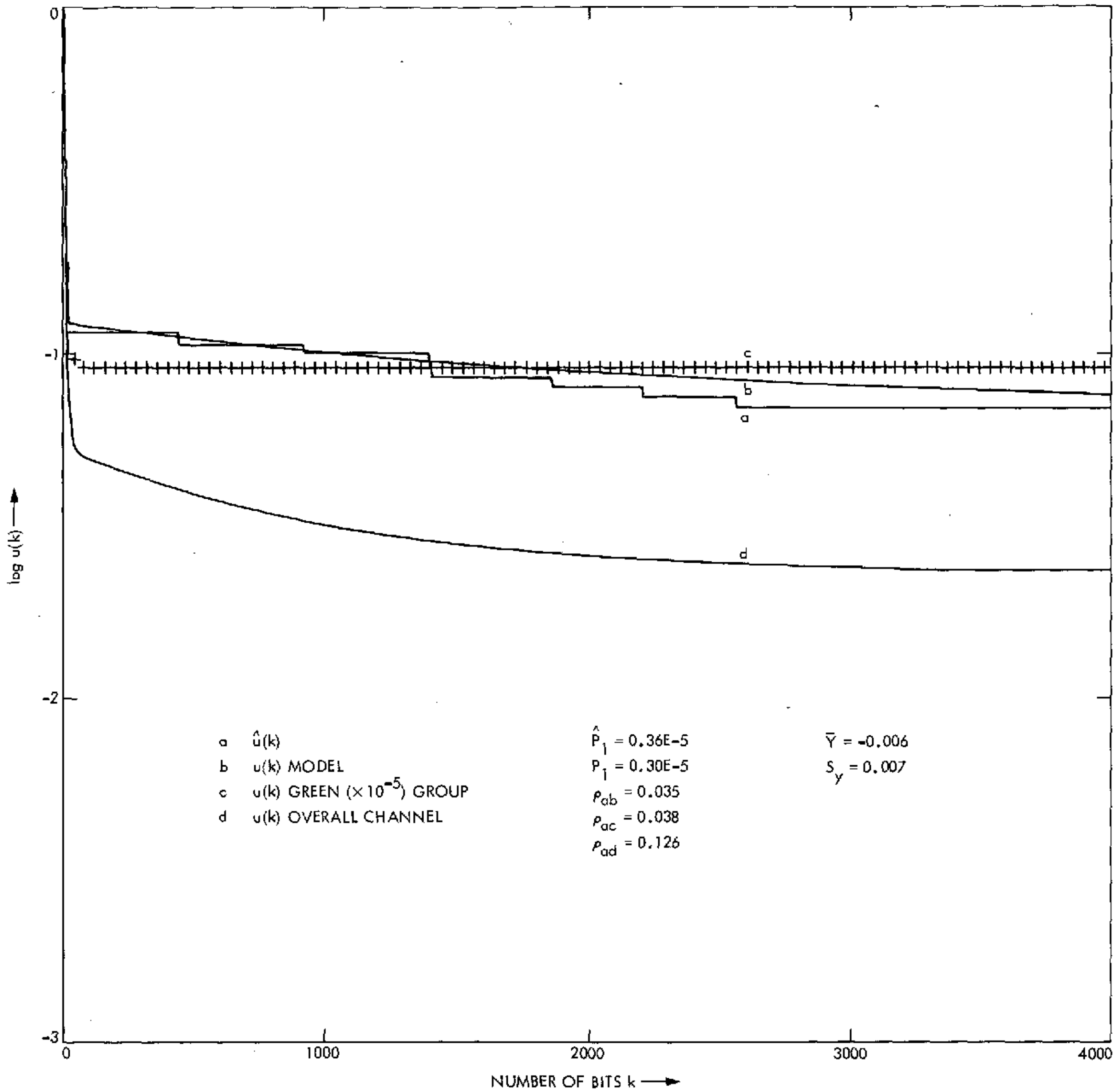

Fig. 4. Gap distribution for Green error group

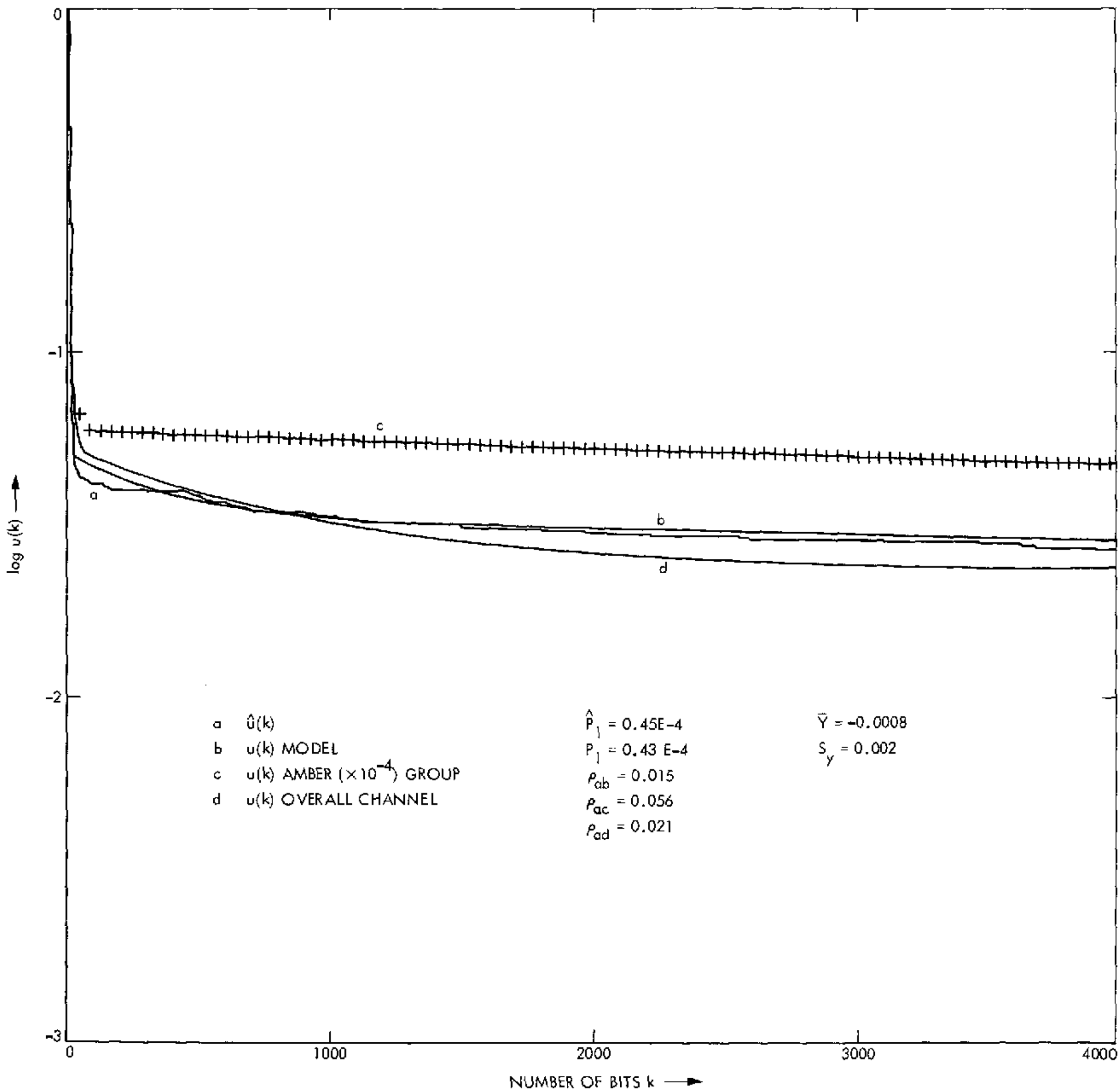

Fig. 5. Gap distribution for Amber error group

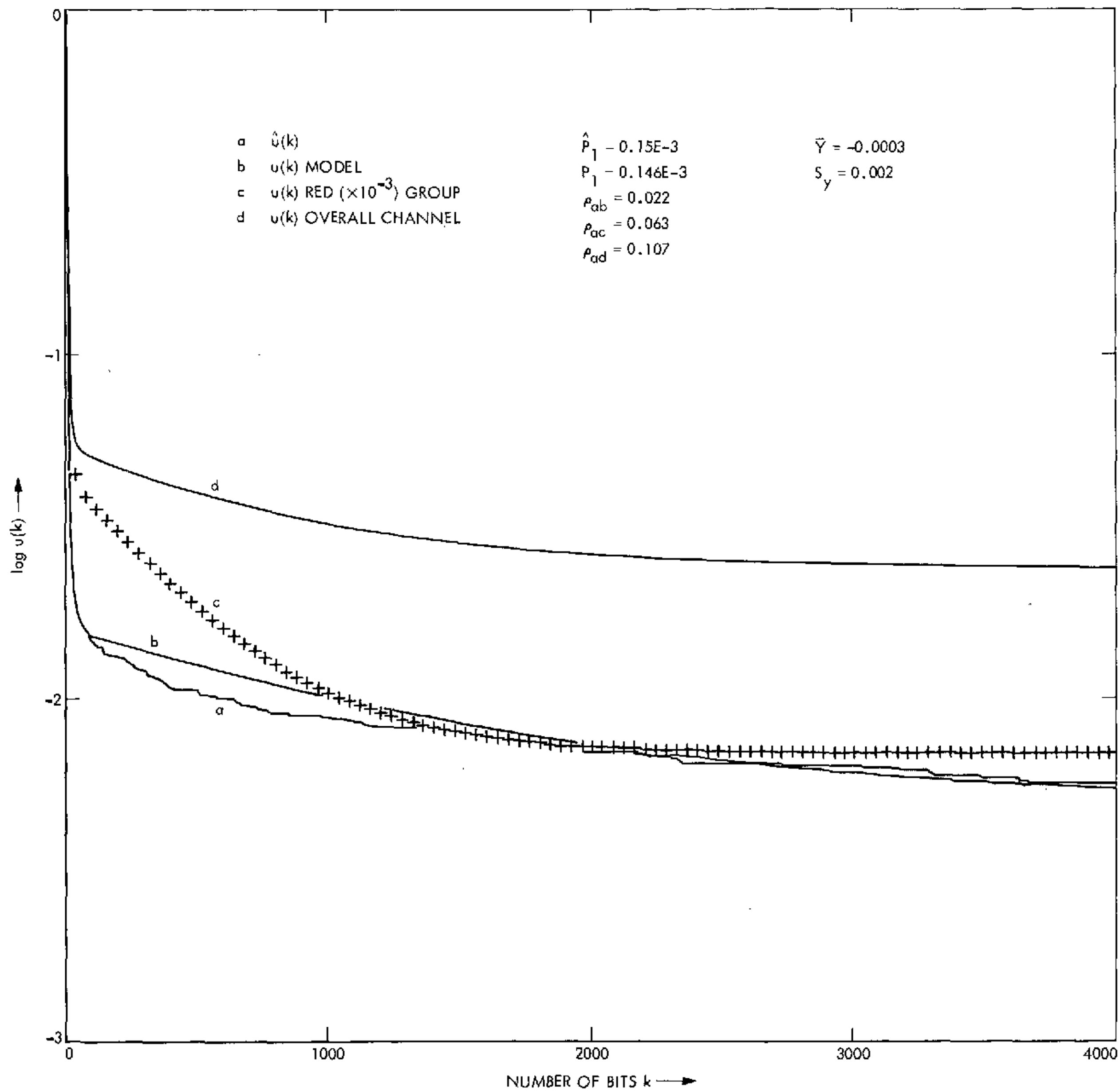

Fig. 6. Gap distribution for Red error group

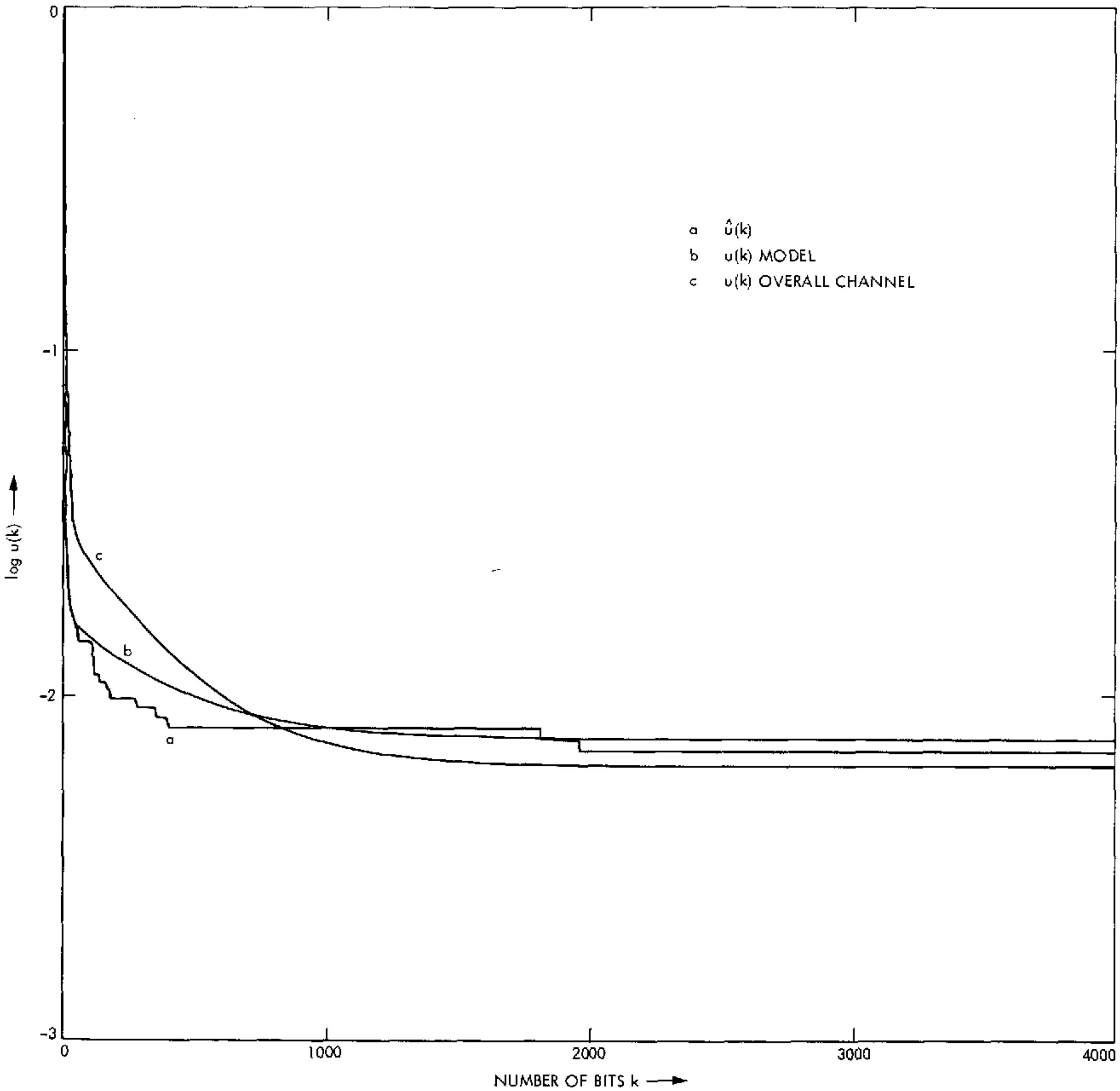

Fig. 7. Gap distribution (50 kbps circuit, BER = 0.51×10^{-4})

starting with raw estimates in (16) yield the stable parameter set for the 4.8 kbps and 50 kbps shown in Tables 5a and 5b respectively.

Figures 4 through 6 also show comparative plots of the function $U(k)$ for each group and the overall channel. As expected the overall channel does not predict any particular group closely enough. On the other hand, the less the spread among the values of the empirical bit error rate in a group, the better the fit between the averaged $U(k)$ and any given sample in that group.

Let us go back to the definition of ρ in (15) and use it to construct confidence intervals for the distribution function $F(k)$. What we want to do is to construct a statistical test of the hypothesis that the true parent distribution $G(x)$, $x \geq 0$ is $F(x)$ against the alternative that $G(x) \neq F(x)$ using the metric ρ at significance levels $\alpha = 0.01, 0.05, 0.10$ and 0.15. That is to say we shall test the hypothesis

$$H_o : G(x) = F(x) \tag{20}$$

at levels α given above. The empirical distribution function is $F_n(k)$. Since $F(x)$ is continuous we shall use the Kolmogorov-Smirnov test statistic.

$$K_n = \sqrt{n} \sup_{-\infty<x<\infty} |F(x) - F_n(x)| \quad . \tag{21}$$

Now we shall confine the supremum in (21) to the range of sample points k for $0 \leq k \leq 4000$. Indeed we shall use equally spaced 2000 points in the range. Thus our $n = 2000$ and K_n then becomes:

$$K_n = \sqrt{n} \max_{0\leq k\leq 2000} |F(k) - F_n(k)| \tag{22}$$

$$= \sqrt{n}\, \rho \quad .$$

It is clear that H_o is to be rejected if K_n is too large. The different rejection regions we shall use will be α. Since no tables exist for n as large as 2000 we shall use the limiting form of the distribution of K_n to get the cut-off points for the test.

Let

$$\phi_n(\lambda) = P(\sqrt{n}\ \rho \le \lambda) \ . \tag{23}$$

Kolmogorov showed, (see Darling [11]), that

$$\begin{aligned} \phi_n(\lambda) &\to \phi(\lambda) \\ &= \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2} \ . \end{aligned} \tag{24}$$

Thus for large n, approximate confidence bands for $F(x)$ are given by $F_n \pm \lambda_\alpha/\sqrt{n}$, where $1 - \phi(\lambda_\alpha) = \alpha$. Since $F_n(k)$ converges to the true parent distribution $G(k)$ with probability one (Cantelli-Gilivenko lemma), $\phi(\lambda_\alpha)$ is thus the probability that the maximum absolute difference between $G(x)$ and $F(x)$ is at most $\lambda_\alpha/\sqrt{n}$ when n is large enough. This statement is not true for small n (see the modification of $\phi(\lambda)$ by the error term $D_n(x) = |\phi_n(\lambda) - \phi(\lambda)| = 0(n^{-1/8})$ in K. Kunisawa, et al [9] for n as large as ours) but tables available for small n do not cover values of n as large as 2000. Kunisawa, et al, further point out that for $\alpha = 0.01$ or 0.05 and the corresponding values of λ_α, $D_n(\lambda)$ are very small for $n > 100$. In any case the effect of the correction factor $D_n(\lambda)$, when $\phi(\lambda)$ is used rather than $\phi_n(\lambda)$, is to increase the rejection region α thus forcing us to reject certain distributions $F(x)$ which would have been accepted as not essentially different from the true $G(x)$ had we used $\phi_n(\lambda)$. Hence the use of the limiting distribution $\phi(\lambda)$ makes the test of the hypothesis H_o very conservative.

Using tables of $\phi(\lambda)$ by Kunisawa, et al [10] the metric ρ was calculated for the different values of α and corresponding λ_α (see Table 5 (c)).

Table 5(c). Kolmogorov-Smirnov Test Statistic

α	$\phi(\lambda_\alpha)$	λ_α	ρ
0.01	0.99	1.628	0.0364
0.05	0.95	1.3581	0.0304
0.10	0.90	1.224	0.0274
0.15	0.85	1.138	0.0255

We now compare the values of ρ with those listed in Table 4a obtained from each of the test runs using equation (15) and notice the following:

1. At level $\alpha = 0.01$, almost all the test runs agree that the distribution function predicted for each of them by the model is the true distribution. In other words with probability 99%, the maximum absolute difference between $G(x)$ and $F(x)$, ρ, is no greater than 0.0364. (Equivalently put, the error of our estimate is at most 3.6%.) In the Red group each ρ is less than this value showing that at level $\alpha = 0.01$, $F(x)$ given by the model is accepted as the true parent distribution. Two of the 19 test runs in the Amber group fail our test only slightly while also two of the seven tests in the Green group reject the distributions our model assigns to them. The encouraging fact is that the higher the bit error rate the smaller the ρ and hence the more we are wont to accept the hypothesis H_o. Remember the emphasis in this study is to model for very short bursts (associated with high bit error rates) that cause decoding errors.

2. Even if we allow as much as 15% ($\alpha = 0.15$) probability against accepting H_0, every one of the test runs in the Red group still accepts the model F(x) as the true distribution. This fact gives us confidence to use our model for predicting significant error patterns in the worst mode of the channel. Notice also that we obtain good results in the Amber group.

3. The grouped runs and the overall channel, however, show poor agreement as we have noted earlier. This is as should be expected owing to the wide range of error rates obtained on the channel ($0 - 10^{-3}$). But the fact that only ten of the test runs show errors of less than 6% indicates that the error patterns on the GCF do not follow a single distribution. The overall channel predicts the Amber error mode better than any other. For example, no run in the Red group shows less than 9% error while the Green group contains runs with as much as 77% disagreement with the averaged channel. Any statistics calculated using a single parameter set will therefore not be valid for the different error modes and hence not reliable.

4. The conclusions for the 50 kbps data are even more striking. At level $\alpha = 0.01$ all the tests except one accept the model distribution as the true distribution (see Table 4(b)).

Section III

AUTOCORRELATION OF BIT ERRORS AND CHANNEL MEMORY

In the last section we presented the Maximum Likelihood Method for obtaining the estimates of the model parameters and demonstrated how closely our model predicts the pattern of errors on the GCF. While we consider the gap distribution or the distribution of return times to the bad state, $V(k)$, $k \geq 1$, to be of fundamental importance in our model, there are functions of this distribution which play a central role in our understanding of the error patterns on the channel. Certainly knowledge of significant error patterns is necessary for error correction. The way the bit errors and error blocks are correlated should be known. In this section bit error correlation will be discussed.

Denote by $r(k)$, $k \geq 1$, the auto-correlation of bit errors. That is

$$r(k) = P(z_k = 1 | z_o = 1); \qquad k \geq 1 . \tag{25}$$

By definition $r(k)$ is the probability of having an error at time k following a given initial error. Let

$$\begin{aligned} G_i(k) &= P(s_k = G_i | s_o = B); \qquad i = 1, 4 \\ B(k) &= P(s_k = B | s_o = B) . \end{aligned} \tag{26}$$

Then

$$r(k) = B(k) . \tag{27}$$

The following recursions are satisfied:

$$B(k + 1) = qB(k) + \sum_i (1 - p_i)G_i(k)$$
$$i = 1, 4 .$$
$$G_i(k + 1) = c_i B(k) + p_i G_i(k)$$

Equivalently,

$$(G_1(k+1), G_2(k+1), G_3(k+1), G_4(k+1), B(k+1)) = (G_1(k), G_2(k), G_3(k), G_4(k), B(k))M \tag{28}$$

where

$$M = \begin{pmatrix} P_1 & 0 & 0 & 0 & 1-P_1 \\ 0 & P_2 & 0 & 0 & 1-P_2 \\ 0 & 0 & P_3 & 0 & 1-P_3 \\ 0 & 0 & 0 & P_4 & 1-P_4 \\ C_1 & C_1 & C_2 & C_3 & q \end{pmatrix}$$

is the transition matrix. Using the initial conditions

$$G_i(0) = 0, \quad i = 1, 4; \quad B(0) = 1$$

we can write (28) as

$$(G_1(k+1), G_2(k+1), G_3(k+1), G_4(k+1), B(k+1)) = (0, 0, 0, 0, 1)M^{k+1} \tag{29}$$

Using the method outlined in Appendix II, for some selected values of k, the matrix M^{k+1} was found. For example, for $k = 5$, and for each of the error groups and the overall channels (both for the 4.8 and the 50 kbps data), M^6 is as shown in Table 6. Above method is not the only one available for finding the autocorrelation $r(k)$. That method is developed specifically for use in Section V for finding the autocorrelation of error-blocks and is presented here only for comparison with the following more directly programmable method.

By definition

$$r(k) = P\{0^j 1 \leftarrow k\text{-}j\text{-}2 \rightarrow 1 | 1\}; \qquad j = 0, k - 1 \tag{30}$$

The (k-j-1) bits indicated are any (k-j-1) binary digits. Hence we can write

$$r(k) = \sum_{j=0}^{k-1} P(0^j 1|1) P\{0^n 1 \leftarrow k\text{-}m\text{-}j\text{-}3 \rightarrow 1|1\}; \; m=0, k\text{-}j\text{-}2$$

that is

$$r(k) = \sum_{j \geq 0} V(j) r(k-j-1) \tag{31}$$

subject to $r(0) = 1$. Call the above methods, Methods I and II respectively. $r(k)$, for $k = 6$ and for both methods, is shown in Table 7 for comparison. $r(k)$, for $k = 1200$, range from .0278 for bit error-rate $P_1 \sim 10^{-3}$ to 0.0000033 for $P_1 \sim 10^{-5}$ with value of 0.007 for the averaged (overall) channel. For the 50 kbps data, $r(1200) = 0.0382$; showing in each case that the memory of the channel is longer than 1200 bits. It is only in the Green group ($P_1 \approx 0.3258E\text{-}05$) that the bit correlation, at 1200 bits apart, of 0.332889E-05 is closest to P_1 showing that in this group the memory is almost 1200 bits.

Table 6. M^k for $k = 6$

Error-rate 10^{-3}

.9999899	.55E-06	.140E-05	.358E-05	.461E-05
.000143	.9863834	.001973	.00537	.006463
.004575	.024998	.6359246	.1539447	.1805568
.01347	.073556	.17742	.40545	.33009
.01874	.10204	.22498	.35689	.2973375

Error-rate 10^{-4}

.9999948	.215E-06	.8068E-06	.182E-05	.230E-05
.3204E-04	.9992708	.0001141	.0002575	.0003256
.01626	.015445	.69563	.12573	.14693
.05634	.053509	.19305	.39236	.30473
.07814	.07420	.24742	.334208	.266023

Error-rate 10^{-5}

.9999984	.313E-06	.124E-06	.412E-06	.747E-06
.0543	.66724	.02838	.094289	.15578
.16635	.219696	.084864	.2818996	.27923
.16635	.219696	.08486	.2818996	.27923
.22482	.27038	.06261	.20799	.23419

Over-All Channel (4.8 kbps)

.9999940	.255E-06	.9178E-06	.2089E-05	.2714E-05
.0002447	.9929	.00110	.0025	.003247
.01494	.01866	.6476	.1459	.17287
.04465	.05576	.19163	.39096	.31699
.06182	.07710	.24198	.33788	.28122

Over-All Channel (50 kbps Data)

.9999987	.328E-06	.475E-07	.349E-06	.5866E-06
.00444	.66362	.0179497	.124576	.189414
.0001697	.004734	.981624	.00504	.00843
.01255	.3303	.050667	.293315	.31321
.01656	.394816	.066642	.246665	.275713

Table 7. Autocorrelation r(k) for k = 6, High-Speed Circuit.

	Method I	Method II
Red	.297337	.297337
Amber	.266023	.266023
Green	.234194	.234876
Overall Ch. (4.8 kbps)	.28122	.281232
Overall Ch. (50 kbps)	.275713	.2757125

Section IV

THE CHANNEL CAPACITY

For estimating the maximal rate for which reliable transmission over the channel is possible we shall now calculate the capacity of the channel.

Following Gilbert [4], the capacity, C, of the burst-noise channel is given by

$$C = 1 - H \tag{32}$$

where

$$H = \lim_{n\to\infty} \sum_{z_i = 0 \text{ or } 1} P(z_1,\cdots,z_{n+1}) \log P(z_{n+1}|z_1,\cdots,z_n) .$$

As shown in Appendix III, H can be written in terms of $U(k)$, $k \geq 0$ as

$$H = - p_1 \sum_{k=0}^{\infty} U(k)\left\{\frac{U(k+1)}{U(k)} \log \frac{U(k+1)}{U(k)} + \left(1 - \frac{U(k+1)}{U(k)}\right)\log\left(1 - \frac{U(k+1)}{U(k)}\right)\right\} . \tag{33}$$

H is the entropy of the noise sequence $z = \{z_n\}$. Since

$$U(k) = \sum_{i=1}^{4} c_i p_i^{k-1}$$

for large values of k, $U(k) \sim c_1 p_1^{k-1}$ where p_1 is the largest of the p's. Thus $\frac{U(k+1)}{U(k)} \sim p_1$ for sufficiently large $k = k_o$ say. So that for all $k \geq k_o$

we can approximate the summand in (33) by

$$h_o = p_1 \log p_1 + (1 - p_1)\log(1 - p_1) \tag{34}$$

and H can then be written as

$$H \approx - P_1 \sum_{k=0}^{k_o-1} U(k)\left\{\frac{U(k+1)}{U(k)} \log \frac{U(k+1)}{U(k)} + \left(1 - \frac{U(k+1)}{U(k)}\right) \log\left(1 - \frac{U(k+1)}{U(k)}\right)\right\}$$

$$- h_o P_1 \sum_{k_o}^{\infty} U(k) \quad .$$

Using the fact that

$$P_1 \sum_{k=k_o}^{\infty} U(k) = \sum_{k=k_o}^{\infty} P(10^k) = 1 - \sum_{k=0}^{k_o-1} P(10^k)$$

$$= 1 - P_1 \sum_{k=0}^{k_o-1} U(k) .$$

We can approximate H by the finite sum:

$$H' = - P_1 \sum_{k=0}^{k_o-1} U(k)\left\{\frac{U(k+1)}{U(k)} \log \frac{U(k+1)}{U(k)} + \left(1 - \frac{U(k+1)}{U(k)}\right)\log\left(1 - \frac{U(k+1)}{U(k)}\right)\right\}$$

$$- h_o\left[1 - P_1 \sum_{k=0}^{k_o-1} U(k)\right]$$

and thus

$$C = 1 - H \approx 1 - H'$$

C was evaluated on the computer for values of k_o from 40 to 1200 in steps of 20 and for both the wide-band and the high frequency model parameters. In each case C converged (to six decimal places) for $k_o = 1000$ while for many of the runs k_o is very much smaller.

Table 8 shows the model values of C for each of the 29 error runs of the 4.8 kbps high frequency data, the three groups Red, Amber and Green and for the overall channel. The lowest capacity is $C = 0.9986$ (the C for a binary symmetric channel with the same bit error rate, C(BSC), is 0.996) and the highest $C = 0.9999984$ (C(BSC) = 0.9999978). The capacity for the Red group is 0.9994, Amber group $C = 0.99992$, Green group $C = 0.999988$ and the overall channel capacity is 0.99988.

The capacity for the wide-band error runs is shown in Table 9. In this case C ranges from a minimum of 0.9993 to a maximum of 0.999997. The overall channel capacity is 0.99991 which is only 0.003 of 1% higher than the capacity of the average channel at high frequency.

At 4.8 kbps during the highest bit error phase (discounting the outages) we can still transmit reliably at rates close to 0.9986 while the rate is 0.9993 for the 50 kbps. During the Green phase (at 4.8 kbps) the worst we can achieve is 0.999983 while for the 50 kbps during the least bit error mode the worst is 0.999992.

We should remember however the recording problems, mentioned in [2], encountered in the gathering of the wide-band data which have the effect of increasing the error-free gap lengths at the end of each test run. These have the effect of lowering the bit error rate and thus increasing the capacity of the channel based on wideband (50 kbps) data.

The calculations based on the 4.8 kbps high frequency data are thus more reliable.

Table 8. **High Speed (4.8 kbps) Channel Capacity**

	Group	Bit-Rate P_1	C	C(GROUP)	C(BSC)	C(BSC)/C
1	Red	.506E-03	.998580		.995653	.997
2	.244E-03	.229E-03	.999576	.9994	.997853	.9983
3		.145E-03	.999716		.998576	.999
4		.111E-04	.999978		.99986	.9999
5		.515E-04	.99988		.99944	.99956
6		.303E-04	.9999		.99965	.99975
7		.105E-04	.999977		.99987	.9999
8		.43E-04	.9999		.99952	.99963
9	Amber	.164E-04	.999965		.9998	.99984
10		.827E-05	.999985	.99992	.999895	.99991
11	.293E-04	.99E-04	.99994		.99976	.99982
12		.393E-04	.99987		.99956	.99969
13		.119E-04	.999969		.99985	.99988
14		.102E-04	.99997		.9999	.9999
15		.909E-05	.99995		.99989	.99994
16		.21E-04	.9999		.99975	.99985
17		.23E-04	.99991		.99973	.9998
18		.229E-04	.99994		.99973	.9998
19		.639E-04	.9998		.9993	.9995
20		.635E-04	.9998		.9993	.9995
21		.353E-04	.99993		.9996	.9997
22		.141E-04	.99997		.99983	.9999
23		.453E-05	.999983		.999943	.99996
24	Green	.303E-05	.999989		.999958	.99997
25		.128E-06	.9999984	.999988	.9999978	.9999994
26	.332E-05	.189E-05	.999993		.99997	.99998
27		.329E-05	.999994		.999955	.99996
28		.162E-05	.999989		.999977	.999988
29		.439E-05	.9999898		.999941	.999952

For the overall channel: P_1 = .438E-04, C = 0.99988

$$C(BSC) = 0.9995, \quad \frac{C(BSC)}{C} = 0.9996$$

Table 9. Wideband (50 kbps) Channel Capacity

Overall Channel	Bit-Rate P_1	C	C(Overall)	C(BSC)	C(BSC)/C
	.461E-04	.99992		.99949	.9996
	.347E-04	.99994		.9996	.9997
	.158E-04	.99997		.99981	.9998
.354E-04	.856E-06	.999997	.99991	.999987	.99999
	.38E-05	.999992		.999950	.99996
	.185E-05	.999997		.99997	.999977
	.199E-05	.999996		.99997	.999976
	.349E-05	.999993		.999953	.99996
	.291E-05	.9999946		.99996	.99997
	.19E-04	.999965		.99988	.99981
	.244E-04	.99996		.9997	.99976
	.242E-03	.9993		.9977	.998

$\overline{C}$ = C(BSC) for overall parameter = 0.9996

$\overline{C}$/C(Overall) = 0.9997

Section V

BLOCK-BIT STATISTICS

A block of size n, (also referred to as a message) is defined as n consecutive bits of data. For the GCF, n = 1200. In data transmission a block is considered as correct, in which case it is accepted, or incorrect, in which case it is ignored or retransmitted. Thus the proportion of blocks to be ignored after possible repeated retransmissions is a measure of the performance of the forward error correcting code employed on such a channel. This is one reason that the block error statistics is a very important group of distributions to be evaluated on the GCF.

(i) Block Error Rate as a Function of Block Size

Let us start this section by calculating the block error rate. An error block is defined as a block having one or more error bits. An unbiased estimate of block error rate or of the probability of a number of errors in a 1200-bit block from the data is obtained by supposing every bit in the test run is a possible beginning of a block. In doing this, each run is divided into consecutive blocks 1200 bits long starting at the $i^{\underline{th}}$ bit of the run ($i = 1,2,\ldots,1200$) and the number $N_i(k)$ of blocks containing k bit-errors is noted. We thus obtain, for each $i = 1,\ldots,1200$, a probability $P_i(k,1200) = N_i(k)/N$ that a block in the subdivision contains k bit-errors or $\sum_{k\leq 1} P_i(k,1200)$ as the block error rate in the subdivision. (N is the total number of blocks in the subdivision). Now average over all the possible 1200 starting positions and take the probability P(k,1200) that k errors occur in a block of length 1200 to be $\frac{1}{1200}\sum_{i=1}^{1200} P_i(k,1200)$ and $\sum_{k\leq 1} P(k,1200)$ to be the block error rate.

From Appendix IV, the probability of no errors in a block of size n is given by:

$$P(0,n) = P_1 \sum_{j=1}^{4} \frac{c_j}{1-p_j}\, p^{n-1}. \qquad (35)$$

Hence the probability of getting an error block is given by

$$P(\text{error block}) = 1 - P(0, n) \; . \tag{36}$$

Remember that $0 < p_j < 1$ for $j = 1, 4$. Hence in (35), p_j^{n-1} goes down to zero as n becomes large. And hence $P(0, n)$ goes down to zero for large n. Therefore by (36) P(error block) goes up to 1 as n gets larger. This is as should be expected: if the bit error rate is not zero, that is to say if it is possible for error to occur on the channel at all, it will occur eventually. So that any block that is almost as long as the total test run is sure to include the error bit.

The empirical and predicted block error rates for $n = 1200$ are shown in Table 10 for the 4.8 kbps and in Table 11 for the 50 kbps data. For the 4.8 kbps data, the block error rate ranges from a low of 0.021 of 1% during the Green phase of transmission to as high as 1.8% during the high bit error mode. The predicted values are 0.0066 for Red, 0.0021 for Amber and 0.00037 for the Green groups with overall value of 0.00156 for the averaged channel. The block rates for the wideband data range between a low of 0.009 of 1% and a high of 0.12 of 1%. The predicted value for the overall channel in this case is 0.00073.

In both cases, as shown in the tables, the predicted and empirical values agree very closely.

(ii) Distribution of the Number of Errors in a Block - P(k, n)

If a block is in error how many of its bits have been received in error? What is the average probability of an undetected block error for block codes? To answer questions like these we need to know $P(k, n)$.

Appendix IV shows that

$$P(k, n) = P_1 \sum_{\ell=0}^{n-k} U(\ell)\overline{P}(k-1, n-\ell-1)$$

where

$$\overline{P}(j, t) = \sum_{m=0}^{t-j} V(m)\overline{P}(j-1, t-m-1) \tag{37}$$

$$\bar{P}(0, n) = U(n)$$

$$= \sum_{i=0}^{4} c_i p_i^{n-1}$$

and hence, for example

$$P(1, n) = P_1 \sum_{\ell=0}^{n-1} U(\ell)\bar{P}(0, n-\ell-1)$$

$$= P_1 \sum_{\ell=0}^{n-1} U(\ell)U(n-\ell-1) \quad .$$

$P(k,n)$, for $n = 1200$ was evaluated for each of the 29 error runs on the HF circuit. Figures 8, 9, and 10 are some of the graphs of the probability of k or more errors in a block, $P(\geq k,n)$, $k = 0,1,\ldots,n$, for block length of $n = 1200$ bits as given by the data and the model for each of the different error modes. To emphasize the effect of the wide range of the bit error rate ($0 - 10^{-3}$) on the block error distribution, the predicted $P(\geq k,n)$ by the overall channel is plotted on each of the graphs (to same scale). For example, in the Red group, of all blocks containing errors, 64.78% contain 25 or more errors and 36.92% have 50 errors or more while the percentages are 9.26 and 0.77 respectively for the Green group. The overall channel predicts 37.00% and 12.29% for ≥25 and ≥50 errors (see Table 13). Tables 14 and 15 show the proportion of all the blocks that were correctly received compared to the expected proportion for both the HF and WB circuits.

Some of the codes now being considered for use on the GCF can correct burst of errors in a block if all the errors in the burst are confined to within a given length apart. In other words, if the distance between extreme errors in a

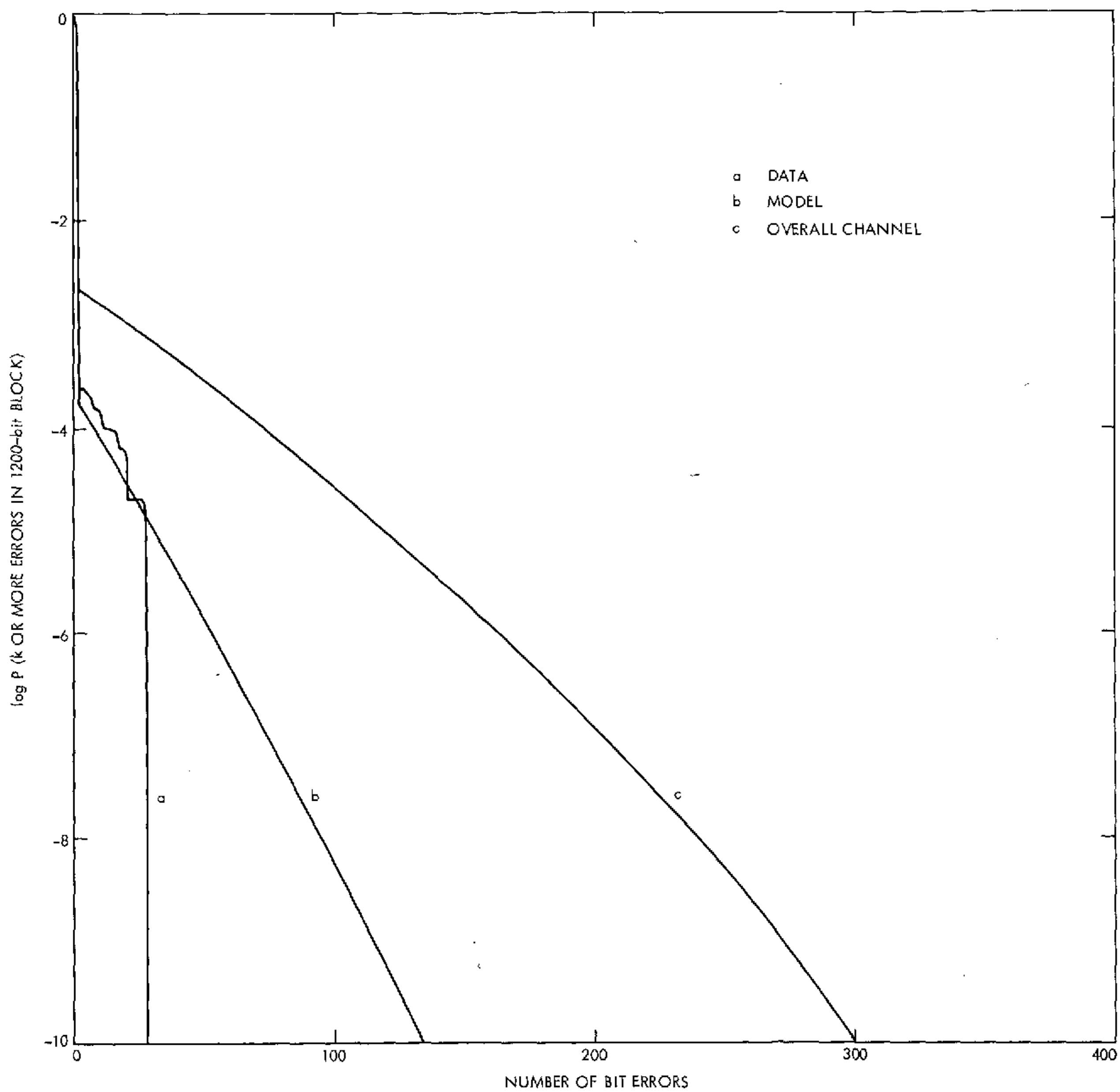

Fig. 8. Distribution of errors in a block (4.8 kbps; Green group)

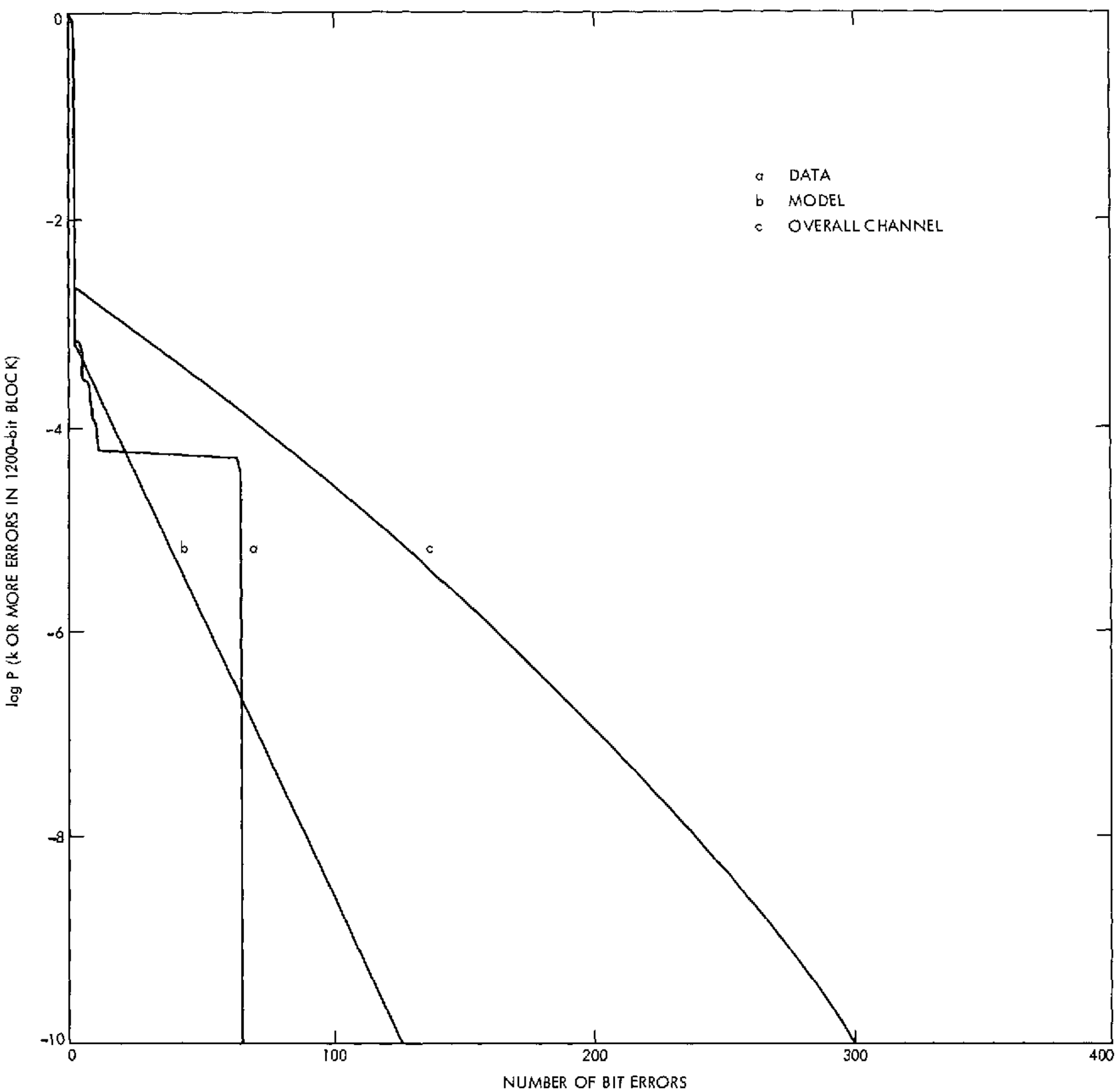

Fig. 9. Distribution of errors in a block (4.8 kbps; Amber group)

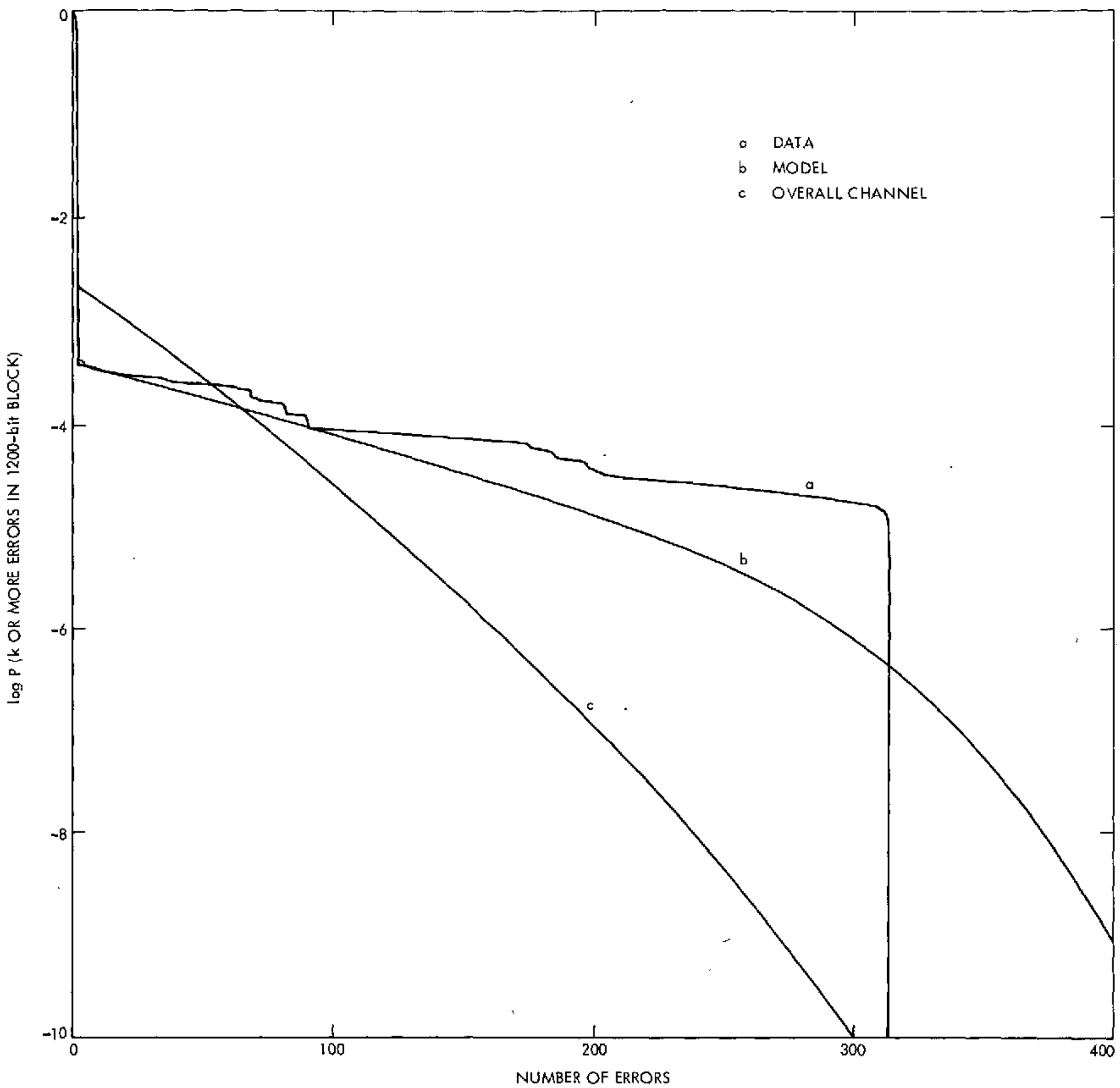

Fig. 10. Distribution of errors in a block (4.8 kbps; Red group)

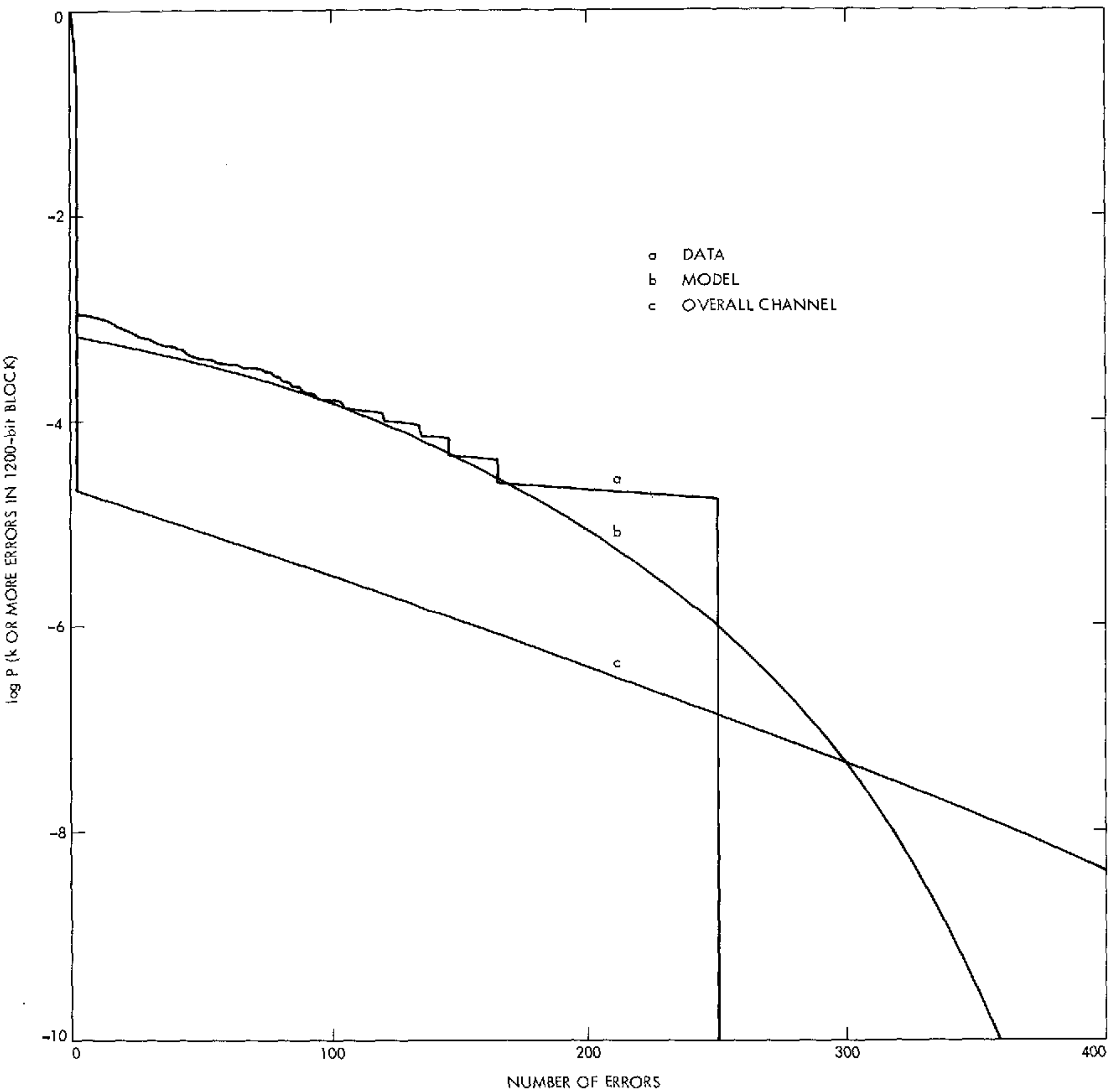

Fig. 11. Distribution of errors in a block (50 kbps, line; bit rate = 0.52×10^{-4})

Table 10 HF circuit (4.8 kbps)

Block Error Rate

	$\hat{P}$(blk Err)	P(blk Err)	P_1
1	.018	.018	.505E-3
2	.00068	.00064	.453E-5
3	.00041	.00027	.915E-5
4	.0025	.0025	.507E-4
5	.0028	.0028	.2997E-4
6	.00057	.00047	.935E-5
7	.0021	.0021	.431E-4
8	.00048	.00040	.298E-5
9	.00095	.000072	.795E-7
10	.00024	.00017	.156E-5
11	.00043	.00039	.197E-4
12	.00013	.000081	.423E-5
13	.00018	.00011	.63E-5
14	.0013	.0013	.211E-4
15	.0054	.0053	.29E-4
16	.00069	.00069	.17E-5
17	.00076	.00077	.129E-4
18	.00021	.00016	.445E-5
19	.0011	.00097	.978E-5
20	.0035	.0032	.915E-5
21	.0035	.0035	.22E-3
22	.0051	.0046	.21E-4
23	.0024	.0025	.146E-3
24	.0022	.0022	.23E-4
25	.0013	.0012	.229E-4
26	.0054	.0058	.638E-4
27	.0033	.0035	.634E-4
28	.0015	.0013	.353E-4
29	.00066	.00054	.14E-4

Table 11 W-B circuit (50 kbps)

Block Error Rate

	$\hat{P}$(blk Err)	P(blk Err)	P_1
1	.0012	.0012	
2	.00054	.00046	
3	.00034	.00029	
4	.000078	.000078	
5	.00012	.000077	
6	.0001	.00009	
7	.000093	.00008	
8	.000094	.000077	
9	.00011	.00008	
10	.00054	.00048	
11	.00056	.00049	
12	.0049	.004	

Table 12 Group Block Error Rate

Group	P(blk Error)
Red	.0066
Amber	.0021
Green	.00037
Overall HF Channel	.00156
Overall W-B Channel	.00073

Table 13(a). HF 4.8 kbps data. % of error blocks containing ≥k errors, $\hat{P}(\geq k,n)$ data $P(\geq k,n)$ model estimates, for block length n = 1200 bits.

$\hat{P}(\geq 25,n)$	$P(\geq 25,n)$	$\hat{P}(\geq 50,n)$	$P(\geq 50,n)$
21.64	56.46	14.14	23.96
8.34	4.9	7.53	0.21
35.78	54.73	23.91	29.11
18.47	36.00	12.02	12.35
21.95	14.05	2.62	1.77
21.54	35.77	17.24	12.18
22.70	38.17	14.02	13.51
9.07	5.87	4.27	0.28
8.53	9.25	8.44	0.69
68.92	68.26	56.53	45.55
52.45	68.37	48.53	45.93
55.38	72.80	47.94	51.87
29.37	29.35	13.05	7.60
3.10	5.7	1.87	0.27
11.44	29.49	8.84	8.08
60.04	47.68	33.80	21.84
13.40	12.44	6.23	1.28
4.23	0.02	1.97	0.003
65.05	74.59	51.38	54.57
6.24	0.67	3.49	0.38
59.45	72.58	39.48	51.21
12.54	14.58	5.01	1.56
24.96	33.73	6.11	10.51
9.16	15.60	3.76	2.22
25.52	32.56	14.49	9.86
21.00	46.37	15.50	20.73

Table 13 (b). W-B 50 kbps data. % of error blocks containing ≥k errors, $P(\geq k,n)$ data $P(\geq k,n)$ model estimates, for block length n = 1200 bits.

$\hat{P}(\geq 25,n)$	$P(\geq 25,n)$	$\hat{P}(\geq 50,n)$	$P(\geq 50,n)$
59.28	61.46	35.68	36.86
67.60	76.96	44.32	58.01
78.49	67.75	56.65	44.94
29.78	29.91	24.80	8.47
50.65	54.89	45.34	29.33
93.19	62.31	79.22	37.82
84.00	54.77	56.06	29.15
65.71	52.04	59.84	26.31
68.42	58.37	40.56	33.98
71.10	58.67	43.91	33.56
88.36	73.78	61.18	53.19
64.92	80.60	50.26	62.21

Table 13 (c). Group estimate of % of error blocks containing ≥k errors for block length n = 1200 bits

Group	$P(\geq 25,n)$	$P(\geq 50,n)$
Red	64.78	36.93
Amber	22.56	4.71
Green	9.26	0.77
Av. 4.8 kbps	37.00	12.29
Av. 50 kbps	74.94	52.57

Table 14 Empirical ($\hat{P}(0,n)$) and predicted ($P(0,n)$) thruput for the H-F 4.8 kbps circuit. $n = 1200$.

	Group	$\hat{P}(0,n)$	$P(0,n)$	Group ($P(0,n)$)
1		.982	.982	
2	Red	.99654	.9965	.9934
3	($x10^{-3}$)	.99764	.99745	
4		.99959	.99973	
5		.9975	.9975	
6		.9972	.9972	
7		.99943	.99953	
8		.9979	.9979	
9		.99957	.9996	
10		.99982	.999894	
11	Amber	.99874	.9987	
12	($x10^{-4}$)	.9946	.9947	.99789
13		.99924	.9992	
14		.9989	.999	
15		.99655	.9968	
16		.99491	.9954	
17		.99782	.9978	
18		.99874	.9988	
19		.9946	.994	
20		.9967	.9965	
21		.99855	.99867	
22		.99934	.99947	
23		.99932	.99936	
24		.99952	.9996	
25		.99991	.99993	
26	Green	.99976	.99982	.99963
27	($x10^{-5}$)	.99987	.99992	
28		.9993	.9993	
29		.99979	.99984	

Overall Channel $P(0,n) = 0.99785$

Table 15 Empirical ($\hat{P}(0,n)$) and predicted ($P(0,n)$) thruput for the W-B 50 kbps circuit. $n = 1200$.

$\hat{P}(0,n)$	$P(0,n)$
.9988	.9988
.99946	.99954
.99966	.99971
.99992	.99992
.99988	.99992
.9999	.99999
.99991	.99992
.99991	.99992
.99989	.99992
.99946	.99952
.99944	.99951
.9951	.996

Overall Channel $P(0,n) = 0.99927$

burst is not too long we shall be able to correct all of those errors even without the use of feedback. This is why it is important to know:

(iii) Distribution of Distances Between Extreme Errors in a Block.

Since any error correcting code will be able to correct at least one error especially if it is the only one in a block, we shall find this distribution for those error blocks with two or more errors.

Let p_k denote the probability of exactly k bits between extreme errors in a block of length n given that the block contains at least two errors. Then as shown in the Appendix,

$$p_k = \frac{P_1 r(k+1) \sum_{m=0}^{n-k-2} U(m)U(n-k-2-m)}{1-P(0,n)-P(1,n)} \tag{38}$$

where $r(k) = P(1_k|1_o)$ is the autocorrelation of bit errors and $P(0, n)$ and $P(1, n)$ are respectively, as found above, the proportions of blocks that are received correctly and those that contain exactly one error.

As much as 1.75% of all blocks transmitted may contain more than two errors (Table 16) (about 0.48% in the wideband circuit) while less than 0.09 of 1% (0.004% for WB circuit) contain exactly one error in an error block which can be so easily corrected. It would thus be essential to use a "burst trapping" code on the GCF if a high proportion of the error blocks have their bursts confined to a correctable length. The length of a burst of errors in a block is the number of bits between the extreme errors in the block whatever the density of errors therein.

Examination of the HF data shows that an average of only 16.4% of the error blocks in the Red group have their bursts confined to within a length of twenty-five bits (22% to within fifty bits) while in some runs in the Green group almost all the bursts are confined to within a maximal length of 25 bits (Table 16). However in every one of the test runs a high percentage (Table 18) of the error blocks have all their errors at exactly a distance of twenty-three bits from the first error in the block. This is the effect of the fixed error pattern caused by the code built into the circuit modem. The code causes bit errors at 18th and 23rd bit positions after a random bit error. For example, in the Red group, as much as 24.2% of all the bursts in the error blocks are due to this effect; in the Amber group there is a run with as high as 83% while a percentage of as high as 98 is recorded in the Green group. The smaller the percentage of this fixed error pattern the better the agreement between the data and predicted values of the error bursts.

A block length of 1200 bits is so long compared to the effect of the fixed error pattern that the two errors caused by the modem code fall, in most cases, within the block having the affected random error. Thus only few blocks should contain exactly one error which may occur either at the beginning of the block (within the first four bits) or at the end (within the last 18 bits). Otherwise an error block would have at least two errors. This fact explains why the empirical probability of exactly one error in an error block, $\hat{P}(1,n)$, is so low (see Table 16).

In the wideband circuit these fixed error effects are not as pronounced although there are some jumps as high as 22% in the block error bursts at a distance of exactly twenty-eight bits from the first error in that block

Table 16 HF 4.8 kbps circuit

Proportion of error blocks containing two or more errors and whose errors are confined to not more than (i) 25 bits, (ii) 50 bits

	Group	P(1,n)	P(k≥2,n)	$\hat{P}$(k≤25| ≥2)%	P(k≤25| ≥2)%	Group P(k≤25| ≥2)%	$\hat{P}$(k≤50| ≥2)%	P(k≤50| ≥2)%	Group P(k≤50| ≥2)%
1		.34E-03	.0175	25.4	9.4		28	14.8	
2	Red	.42E-4	.0034	9.4	11	10.3	18.4	19.6	17.2
3		.33E-4	.0025	14.4	11.1		20.0	20.1	
4		.68E-5	.00027	51.7	16.0		53	39.1	
5		.11E-3	.0024	58.8	29.7		78.2	48.4	
6		.22E-3	.0026	54.5	34.7		67.8	50.8	
7		.20E-4	.00045	43.1	31.5		64.5	51.4	
8		.8E-4	.002	42.8	25.3		61	42.1	
9		.61E-5	.00038	15.1	11.7		16.8	20.0	
10		.14E-5	.0001	33	13.3		41.6	24.1	
11		.62E-4	.0012	27.7	26.6		48.2	42.2	
12	Amber	.59E-3	.0047	60.1	45.6	34.7	90.4	68.3	54.0
13		.38E-4	.00073	51	31.9		77.4	52.5	
14		.79E-4	.0009	68.9	39.7		72.9	56.8	
15		.91E-3	.0023	90.3	47.9		90.4	70.4	
16		.83E-3	.0037	74.2	45.7		81.00	65.5	
17		.16E-3	.002	52.5	20.6		66.3	33.7	
18		.54E-4	.0012	27.9	28.4		40.5	46.4	
19		.43E-3	.0053	47.9	42.2		80.5	64.3	
20		.16E-3	.0034	43.4	27.4		62.6	42.9	
21		.42E-4	.0013	60.9	27.3		78.6	47.0	
22		.164E-4	.00051	42.2	21.8		59.9	35.9	

Table 16 Cont'd.

	Group	P(1,n)	P(k≥2,n)	$\hat{P}$(k≤25|≥2)%	P(k≤25|≥2)%	Group P(k≤25|≥2)%	$\hat{P}$(k≤50|≥2)%	P(k≤50|≥2)%	Group P(k≤50|≥2)%
23		.76E-4	.00057	67	51.0		91.4	76.0	
24		.44E-4	.00035	70	47.4		88.4	69.5	
25		.54E-4	.00005	99.9	74.3		100	93.4	70.2
26	Green	.16E-4	.00016	18.8	41.7	48.0	59	60.3	
27		.13E-5	.00008	3.53	16.4		47.0	30.1	
28		.23E-3	.00046	95.9	54.8		99.96	79.6	
29		.50E-5	.00016	37.3	23.8		55.4	36.6	

Average $\hat{P}(k\leq 25|\geq 2) = 46\%$; $P(k\leq 25|\geq 2) = 24.2\%$

$\hat{P}(k\leq 50|\geq 2) = 63.2\%$; $P(k\leq 50|\geq 2) = 38.9\%$

Table 17 W-B 50 kbps Circuit

Proportion of error blocks containing two or more errors and whose errors are confined to not more than (i) 25 bits, (ii) 50 bits

	P(1,n)	P(k≥2,n)	P̂(k≤25| ≥2)%	P(k≤25| ≥2)%	Overall P(k≤25| ≥2)%	P̂(k≤50| ≥2)%	P(k≤50| ≥2)%	Overall P(k≤50| ≥2)%
1	.24E-4	.0012	6.4	19		36.5	33	
2	.78E-5	.00046	19.2	10.9		22.8	19.6	
3	.46E-5	.00029	4.5	17.7		19.5	31.3	
4	.38E-5	.00007	4.6	28.2		10.9	44.0	
5	.19E-5	.00008	25.1	24.5	8.6	27.8	43.0	13.3
6	.8E-6	.0001	3.4	20.8		7.1	35.2	
7	.20E-5	.00008	4.6	24.3		10.0	39.5	
8	.21E-5	.00007	31.5	27		33.8	46.7	
9	.18E-5	.00008	24.5	23.6		30.4	39.6	
10	.10E-4	.00047	15.7	20.8		22.7	36.9	
11	.61E-5	.0048	3.1	13.3		10.8	23.6	
12	.34E-4	.004	7.4	4.0		10.2	6.8	

Table 18

HF (4.8 kbps) circuit.

Values of k at which impulsive increase in $\hat{P}(—|\geq 2)$ occurs

	k	$\hat{P}(\frac{k}{—}|\geq 2)\%$
1	23	24.2
2	23	4.95
3	23	9.1
4	20	10.1
	23	39.4
5	23	54
6	23	44.4
7	23	34.6
8	23	36
9	23	13.1
10	23	31.2
11	23	25.3
12	23	49.8
13	23	28.1
	24	10.2
14	23	55.8
15	23	83.0
16	23	65.7
17	23	46.7
18	23	25.0
19	23	35.8
20	23	35.9
21	23	38.8
22	23	27.4
23	23	56.5
24	23	59
25	23	98.1
26	23	12.3
27	46	40.4
28	23	89.8
29	23	17.9
	46	16.5

Table 19

W-B (50 kbps) circuit.

Values of k at which impulsive increase in $\hat{P}(—|\geq 2)$ occurs

	k	$\hat{P}(—|\geq 2)\%$
1	--	--
2	28	16.4
3	--	--
4	--	--
5	28	22.1
6	--	--
7	--	--
8	25	28.5
9	51	21.3
10	28	12.4
11	--	--
12	--	--

Table 20

Predicted distribution of errors in interleaved codes

Group	$P_t(0,n)$	$P_t(1,n)$	$P_t(\geq 4,n)$
Red	.993860	.00059	.00447
Amber	.99831	.00049	.00061
Green	.99973	.00011	.00005
Overall 4.8 kbps	.99813	.00038	.00092
Overall 50 kbps	.99932	.000047	.00056

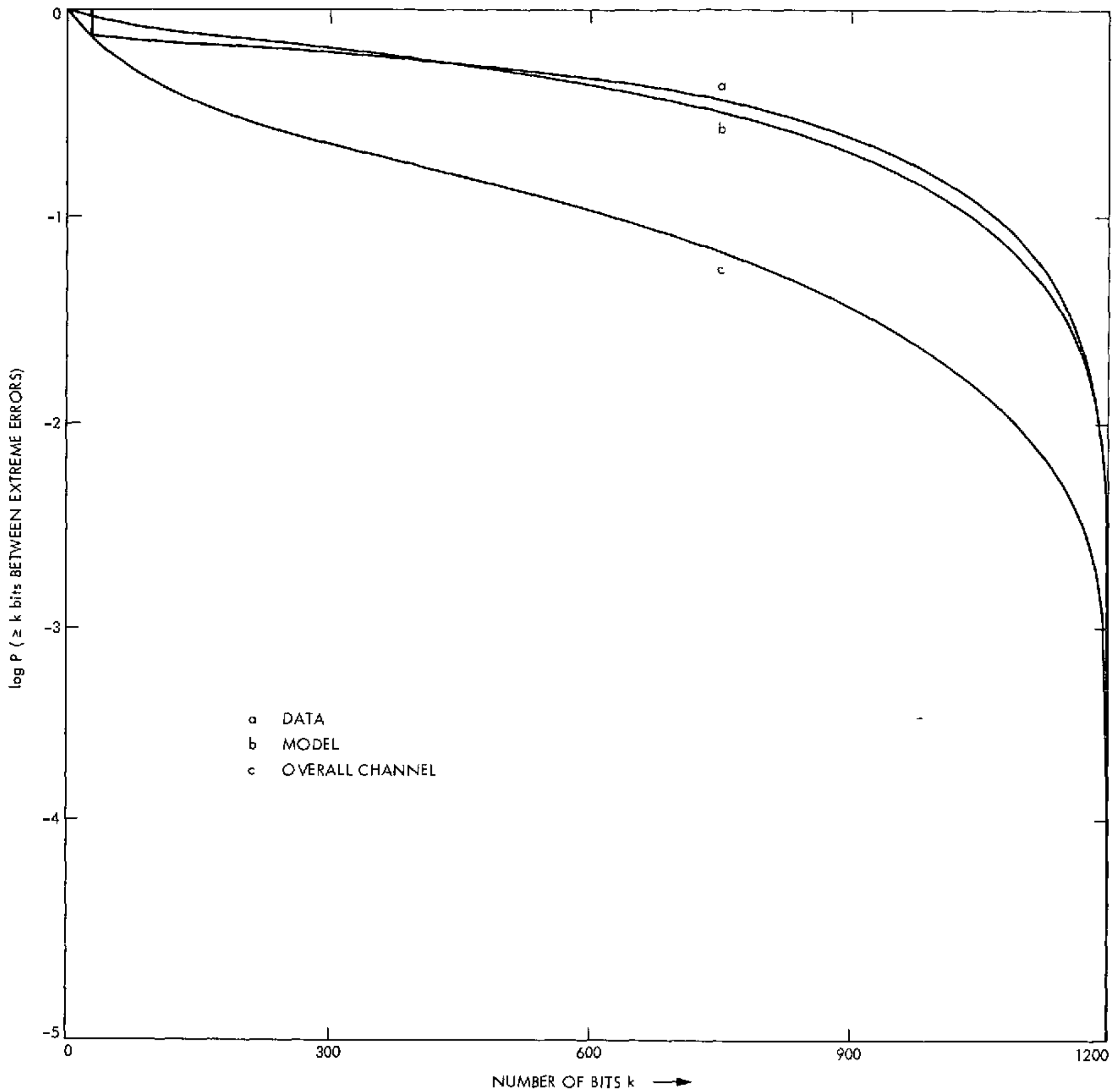

Fig. 12. Distribution of distances between extreme errors in a block (4.8 kbps; Red group

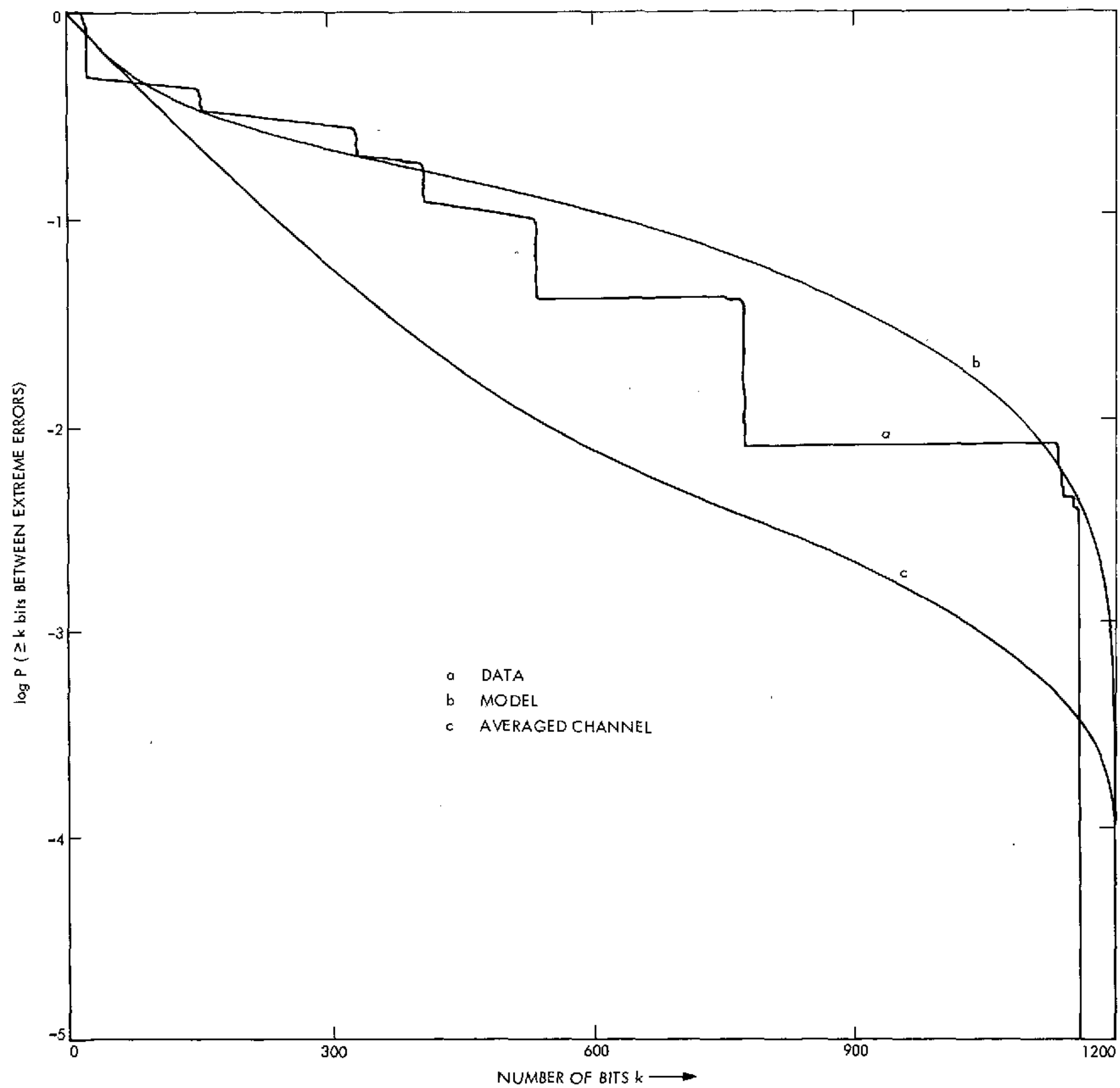

Fig. 13. Distribution of distances between extreme errors in a block (4.8 kbps; Amber group)

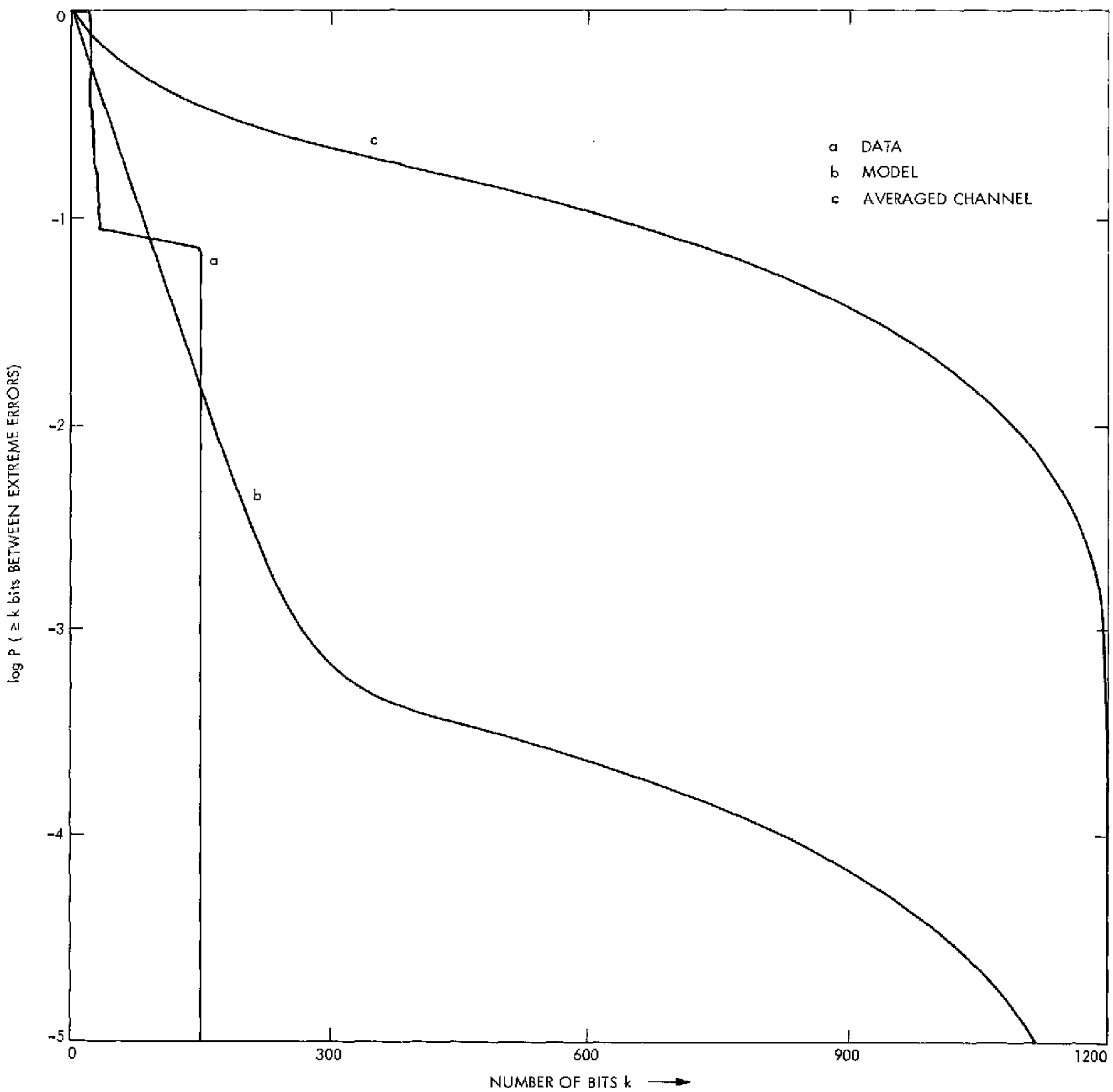

Fig. 14. Distribution of distances between extreme errors in a block (4.8 kbps; Green group)

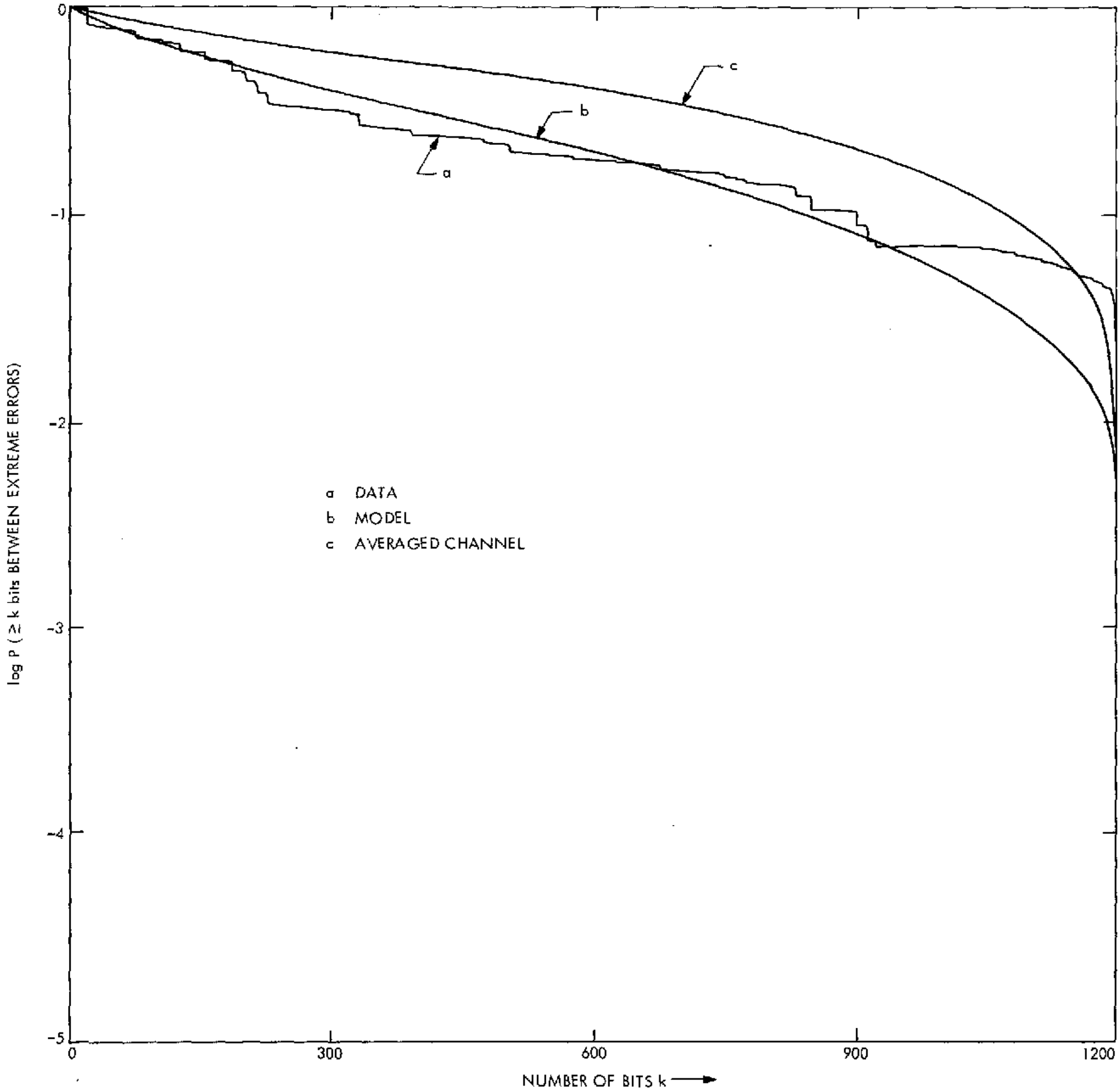

Fig. 15. Distribution of distances between extreme errors in a block (50 kbps; bit rate = 0.52×10^{-4})

(Table 19). Which explains why in this case the data and predicted values are much closer (Table 17).

In a seperate paper to deal with comparison of code performances on the GCF, a strategy will be developed to correct for these fixed error patterns before assessing the capability of any code on the channel.

A rather simple way to correct burst noise is to interleave the coded blocks to some depth t, say. Here the bits of each block are not transmitted consecutively but rather the bits of every t coded blocks are interspersed in such a way that the once consecutive bits are transmitted separated exactly t bit positions apart. When re-ordered at the receiver the blocks appear to have been corrupted by random errors, for sufficiently large depth t. It is interesting to observe the effect of interleaving on the distribution of block errors. The random error effect is to spread out the bit errors among many more blocks than would otherwise have been affected by the action of the bursty channel alone on the un-interleaved blocks with the result that the number, and hence the proportion of error-free blocks after the interleaved blocks are commuted together again at the receiver, is lower than without interleaving. But the error blocks now contain fewer errors. Thus if a code that can correct up to, say, r errors in a block is interleaved appropriately, many more blocks than otherwise would now be decadable. Hence a measure of performance of interleaving is the percentage increase in the proportion of decadable blocks. This is the main idea behind interleaving.

Suppose we interleave a block code of length N to depth t. That is each of the blocks is divided into t sub-blocks as shown in (39).

$$\begin{array}{l} u_1 u_{t+1} \cdots u_{ct+1} \\ u_2 u_{t+2} \cdots u_{ct+2} \\ \vdots \\ u_t u_{2t} \cdots u_{(c+1)t} \,. \end{array} \quad ; \qquad c = \frac{N}{t} - 1 \tag{39}$$

Each sub-block is seperately encoded and decoded before they are finally commuted together again. Let $X_1, X_2, \ldots, X_t$ be the number of errors in sub-block j, $j = 1, 2, \ldots, t$. By the time independence of our model (which in practical terms implies that the channel is as likely to introduce errors say in the first sub-block as it is in any other) it is not difficult to see that the X_j's are identically distributed. But they are not independent (because of the burst phenomenon). Now suppose the interleaved code can correct up to r errors in each of the t sub-blocks. Then we shall be able to correct up to a total of rt errors in the whole block of length N if no one of the component t sub-blocks contains more than r errors. Thus we should find the probability that $\{X_1 \le r, X_2 \le r, \ldots, X_t \le r\}$.

First we find the distribution of the X's denoted by $P_t(k,n)$.

(iv) Distribution of Errors in a Code Interleaved to Some Depth t

It is shown in the Appendix that

$$P_t(k,n) = P_t(1) \sum_{\ell_1=0}^{n-k} U_t(\ell_1)\overline{P}_t(k-1,\, n-\ell_1-1) \tag{40}$$

where

$$\bar{P}_t(k-1,\ n-\ell_1-1) = \sum_{\ell_2=0}^{n-\ell_1-k} V_t(\ell_2)\bar{P}_t(k-2,\ n-\ell_1-\ell_2-2)$$

$$U_t(k) = P_t(0^k|1)$$

$$V_t(k) = P_t(0^k 1|1)$$

$$P_t(1) = \text{bit rate}$$

all calculated using the t-step transition probabilities. Here $n = \frac{N}{t}$ is the length in bits of each of the t sub-blocks.

Let us interleave each block to depth $t = 6$ so that $n = 200$ bits. In this case for the Red group the thruput $P_t(0,n)$ is as high as 0.99386 with only 0.45 of 1% containing four or more errors. In the good mode (Green group) the thruput is 0.9997 with only 0.00005 containing four or more errors. See Table 20 for the complete numbers.

As we remarked above the important distribution for evaluating performance of interleaving is the probability of the joint events $\{X_1 \le r,\ X_2 \le r,\ \ldots,\ X_t \le r\}$ for the error correcting capacity r of the code employed. Details of this and some other strategies of burst correction will be given in a seperate paper.

Section VI

BLOCK (SYMBOL) ERROR DISTRIBUTION

There are some error correcting algorithms in which a block is considered as being made up of symbols, each symbol is a fixed number of bits. Such an algorithm is designed, for example, to correct up to a given number of symbol errors in a block. Because the errors on the GCF occur in clusters therefore, it may be more efficient to employ this type of algorithm rather than use one that is designed to correct only bit errors. Another reason for looking at symbols instead of the individual bits will be demonstrated presently when we consider the problem of acquisition and maintenance of synchronization on the GCF.

But even without consideration of forward error correction it is almost clear that dividing a block on the GCF into symbols of appropriate length is a more efficient way to take advantage of the burst noise for feedback and retransmission. For in doing so, we shall need only ask for and retransmit the symbols within each block that are received in error instead of having to retransmit the whole block. Retransmitting only the error symbols will particularly be preferred in cases when, although the number of bit errors in the block is higher than the error correcting capability of the code employed, the errors are all confined to within only a few symbols.

For a symbol of length s let us first find the statistics we shall employ to estimate the performances of the different algorithms designed for correcting symbol errors. Then we shall treat the problem of acquisition and maintenance of synchronization on the GCF. Specifically we shall look at the following statistics:

(i) $P^s(k,n)$: the distribution of error symbols in n-symbol word

$P^s(0^n)$ = probability of error-free n-symbol word

(ii) $P^s(0^k|1)$: the probability of k error-free symbols following a given error symbol,

$P^s(0^k 1|1)$ = the symbol gap distribution

(iii) $R^s(k)$ = P(symbol k in error|initial error symbol) or the correlation of symbol errors.

In the sequel a symbol is considered to be in error if at least one of its bits is in error. All the above expressions are derived in the Appendix V.

(i) Distribution of Error Symbols in n-Symbol Word

This is given by:

$$P^s(k,n) = P_1^s \sum_{\ell=0}^{n-k} P^s(0^\ell|1)\ \overline{P}^s(k-1,\ n-\ell-1) \qquad (41)$$

where

$$\overline{P}^s(k-1,n-\ell_1-1) = \sum_{\ell_2=0}^{n-k-\ell} P^s(0^{\ell_2}1|1)\ \overline{P}^s(k-2,\ n-\ell_1-\ell_2-2)$$

(ii) Symbol Gap Distribution

The probability of k error-free symbols following a given error symbol is:

$$P^s(0^\ell|1) = \frac{\sum c_i p_i^{s\ell} \dfrac{(1-p_i^s)}{p_i(1-p_i)}}{\sum \dfrac{c_i(1-p_i^s)}{p_i(1-p_i)}}$$

$$P_1^s = P_1 \sum \frac{c_i(1-p_i^s)}{p_i(1-p_i)}$$

$$P^s(0^\ell 1|1) = \frac{\sum c_i p_i^{s\ell} \frac{(1-p_i^s)^2}{p_i(1-p_i)}}{\sum \frac{c_i(1-p_i^s)}{p_i(1-p_i)}} \quad ;$$

P_1^s is the symbol error rate and P_1 is the bit rate. For $s = 6$, 8 or 10 in a block of length 1200 bits, $P^s(k, n)$ was found for the Red, Amber and Green groups and the overall H-F and W-B channels (Table 21). If we use a code that corrects up to two symbols say, the table shows the proportion of blocks that would contain more symbol errors than the capability of the code and may have to be retransmitted. For example, for symbol length $s = 6$ bits in the Red (10^{-3} bit rate) group about 0.6% of the blocks would be in this category. To achieve the same error rate we would have to be able to correct up to 6 bit errors in the 1200-bit block if we use a BCH code and we would need to use about 66 parity check bits to do it. At present only 33 bits are allowed for error detection and correction capability on the GCF. So that even if we use all the 33 bits for error correction alone we cannot correct more than 3 bit errors in the 1200-bit block. For the same symbol error correction capability the longer the symbol length, of course, the less the proportion of blocks left uncorrected and the longer the parity check bits of the BCH code that will give the same error probability (see Table 21).

As said in a number of places already our intention in this report is not to present detailed study of appropriate coding strategies (including feedback) for the GCF. This we intend to do in a separate paper. The above trade-off was mentioned only to emphasize the importance (and efficiency) of symbol error correction on the GCF.

Table 21. Symbol error rate (Group Statistics). The proportion having at least three symbol errors and the number of bit errors that give the same proportion in a 1200-bit block considered as made up of symbols of length $s = 6$, 8 or 10 bits.

	$P^s(0,n)$	s=6; n=200 P_1^s	$P^s(\geq 3,n)$	Number of bit errors k giving equivalent value of $P^s(\geq 3,n)$; & $(P(\geq k,1200))$	s=8; n=150 P_1^s	$P^s(\geq 3,n)$	Number of bit errors k giving equivalent value of $P^s(\geq 3,n)$; & $(P(\geq k,1200))$	s=10; n=120 P_1^s	$P^s(\geq 3,n)$	Number of bit errors k giving equivalent value of $P^s(\geq 3,n)$; & $P^s(\geq k,1200))$
Red	.9935	.00066	.0061	6 (.0061)	.00073	.00595	7 (.00596)	.000779	.00586	8 (.00586)
Amber	.99788	.000084	.00162	5 (.00165)	.932(-4)	.00153	6 (.00155)	.00010	.00145	7 (.00146)
Green	.999624	.966(-5)	.00024	5 (.00025)	.11(-4)	.00022	6 (.00023)	.12(-4)	.00021	7 (.00020)
Av. 4.8	.9978	.000122	.00183	5 (.00183)	.000135	.00176	6 (.00176)	.000145	.0014	7 (.0017)
Av. 50	.99932	.976(-4)	.000644	12 (.000642)	.00011	.000637	13 (.000635)	.00012	.000631	14 (.000628)

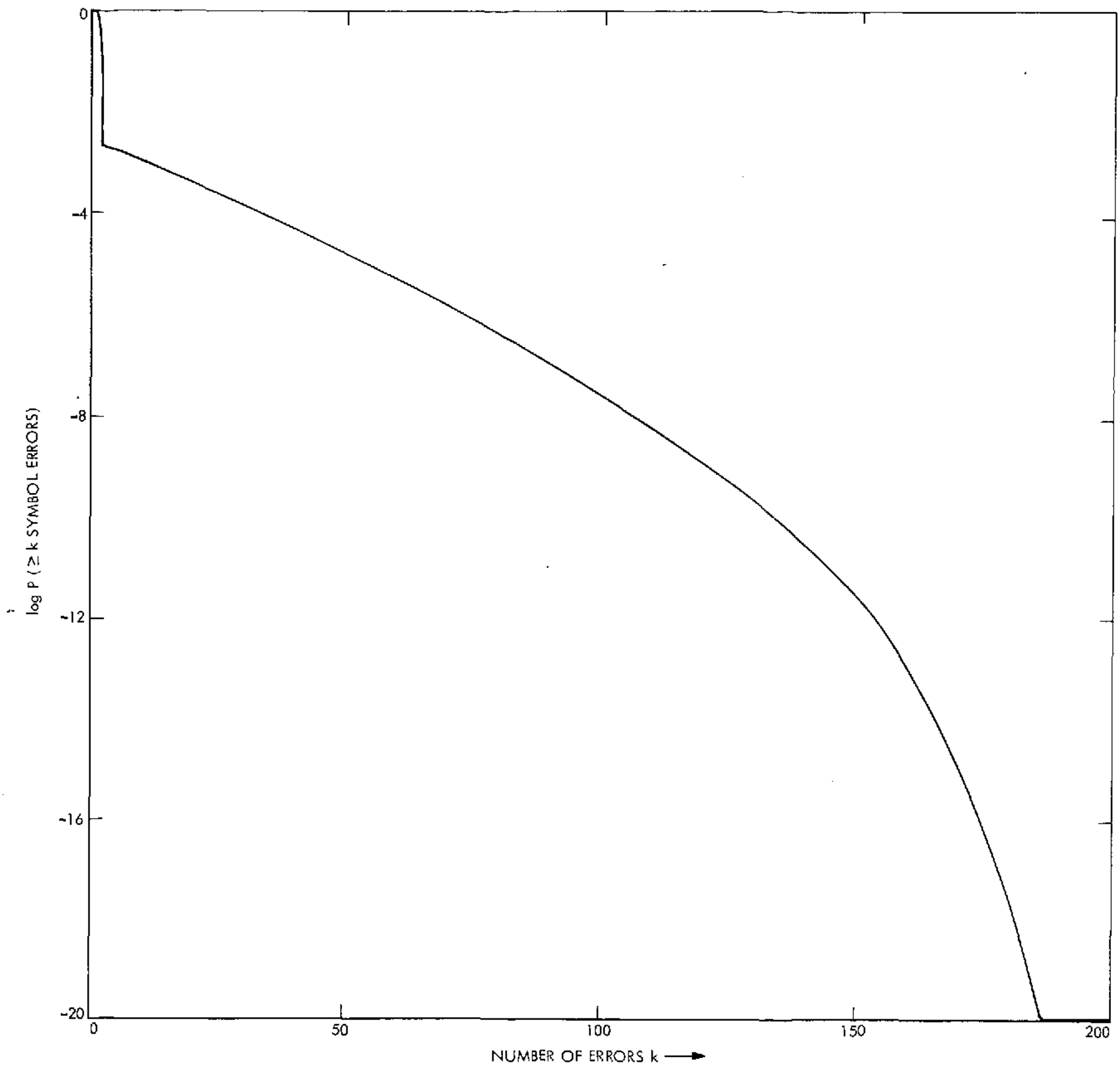

Fig. 16. Distribution of symbol errors (averaged 4.8 kbps channel; symbol length = 6 bits)

C - 2

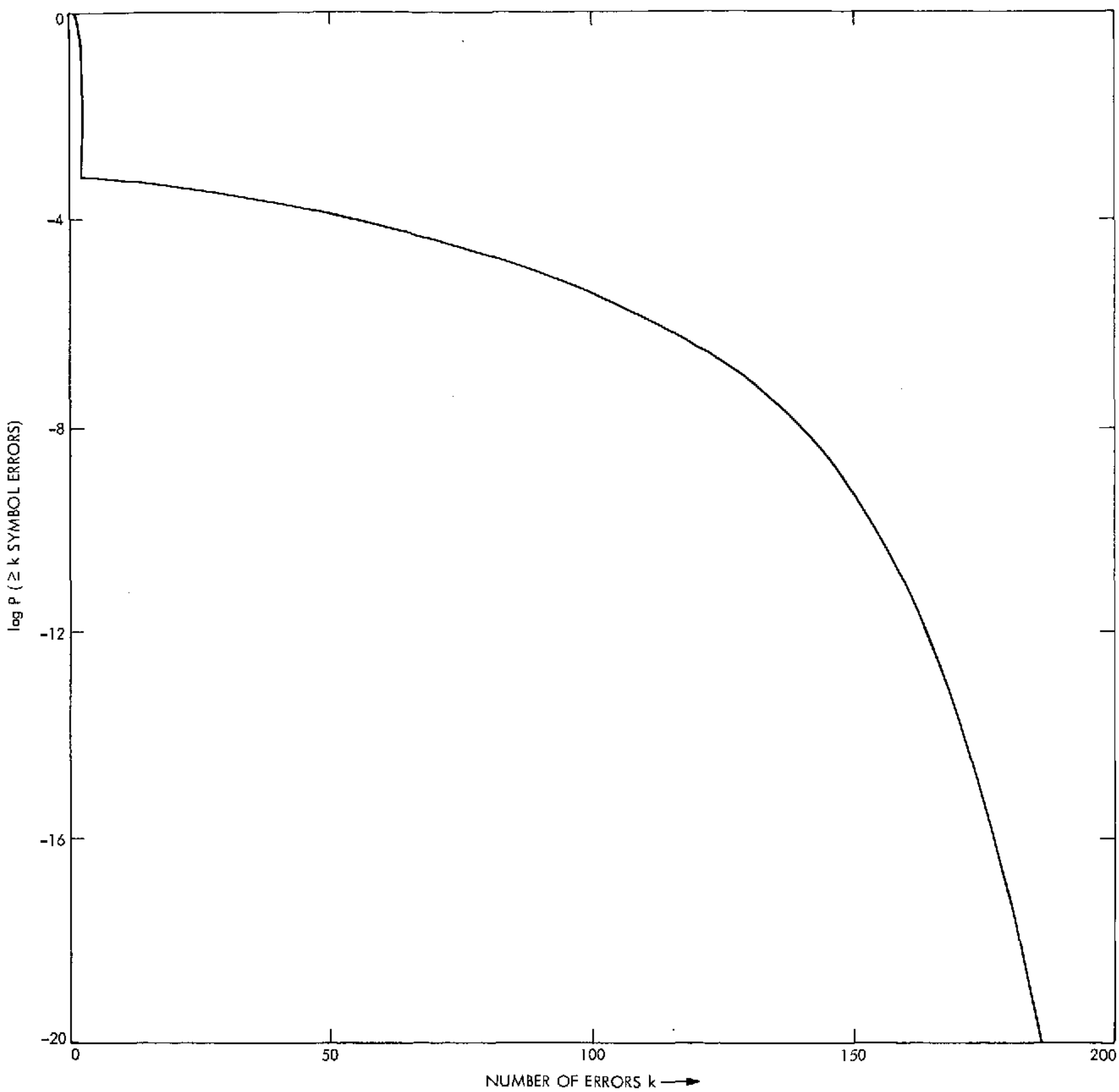

Fig. 17. Distribution of symbol errors (overall 50 kbps channel; symbol length = 6 bits)

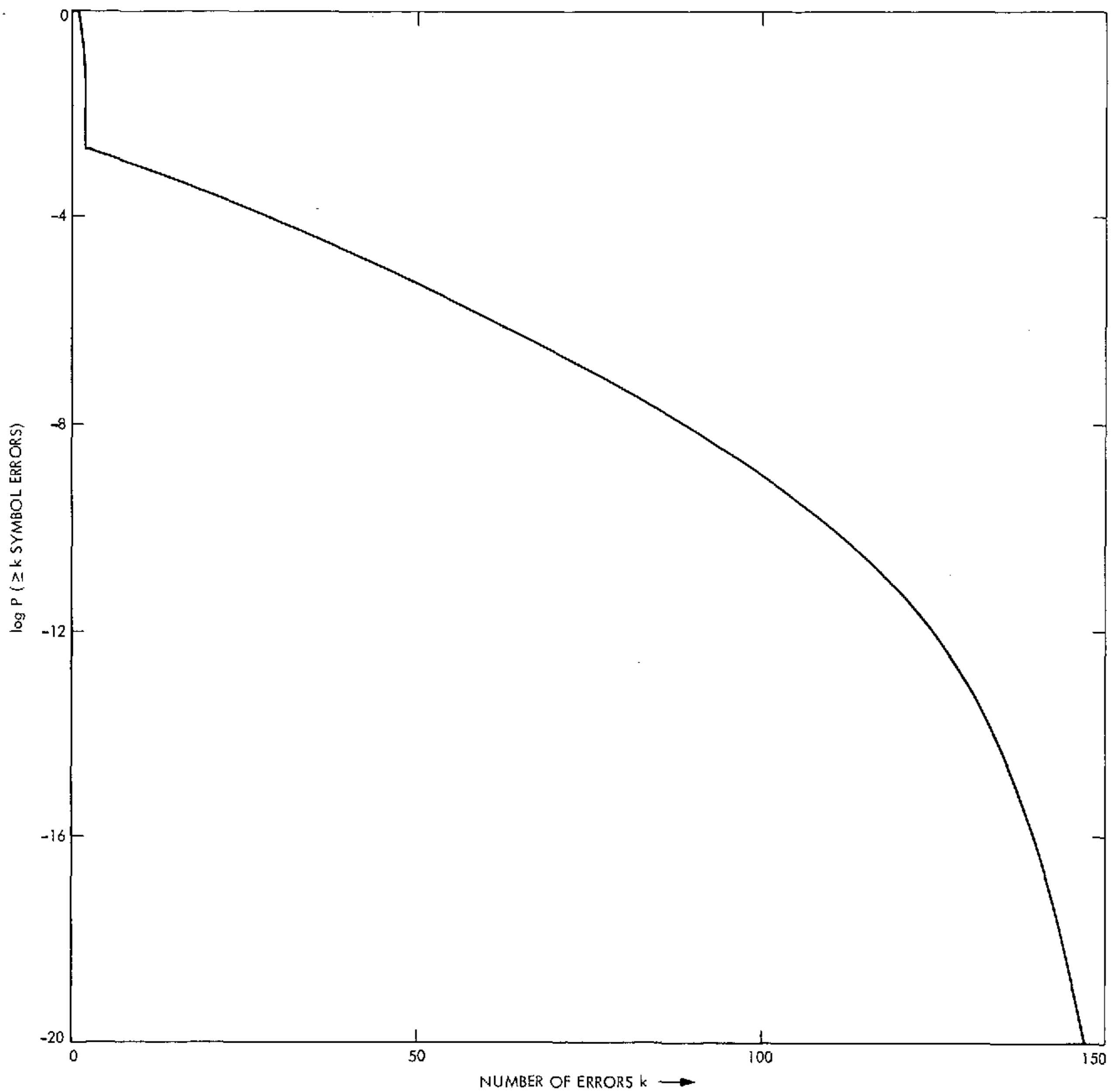

Fig. 18. Distribution of symbol errors (averaged 4.8 kbps; symbol length = 8 bits)

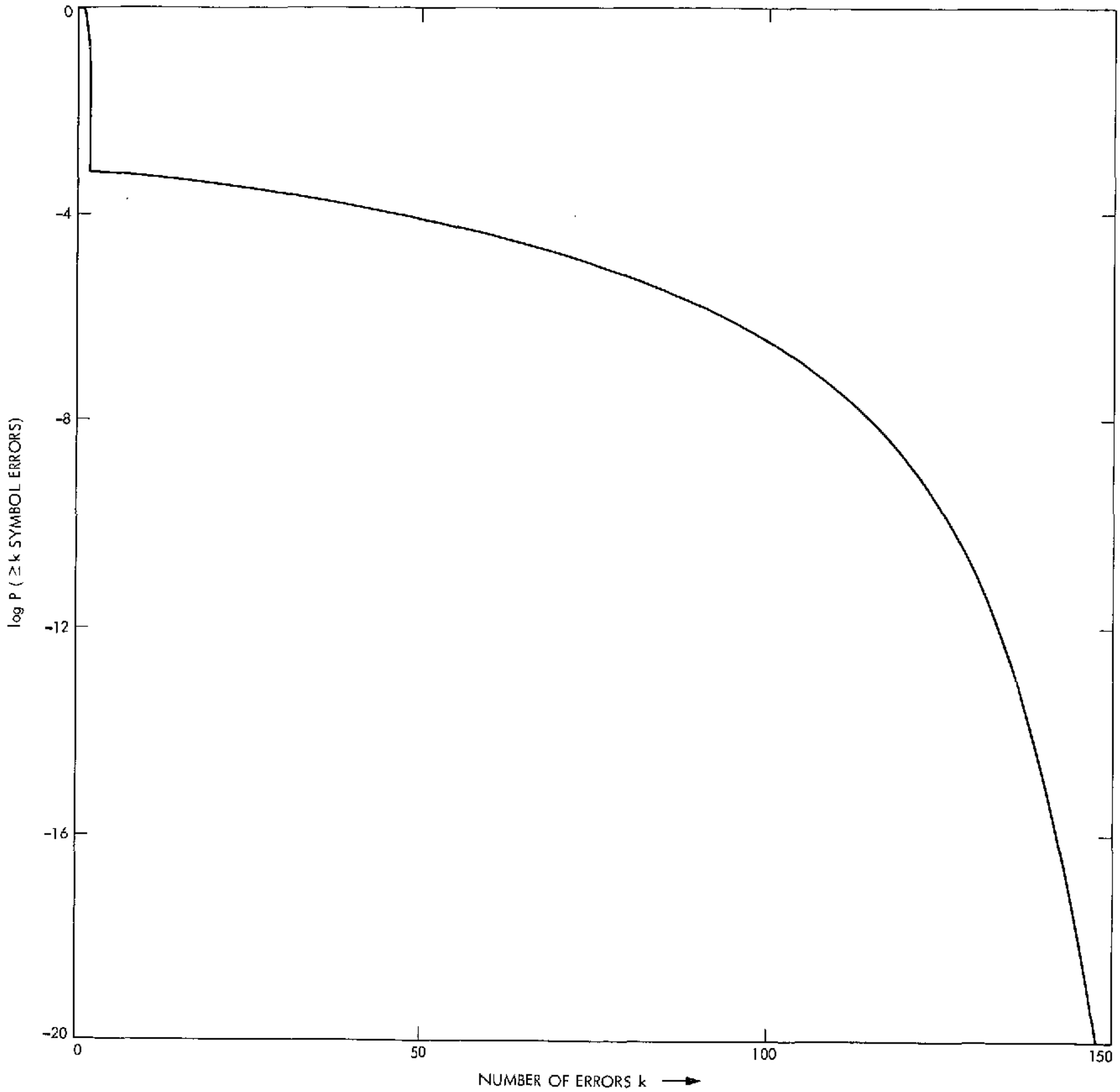

Fig. 19. Distribution of symbol errors (overall 50 kbps; symbol length = 8 bits)

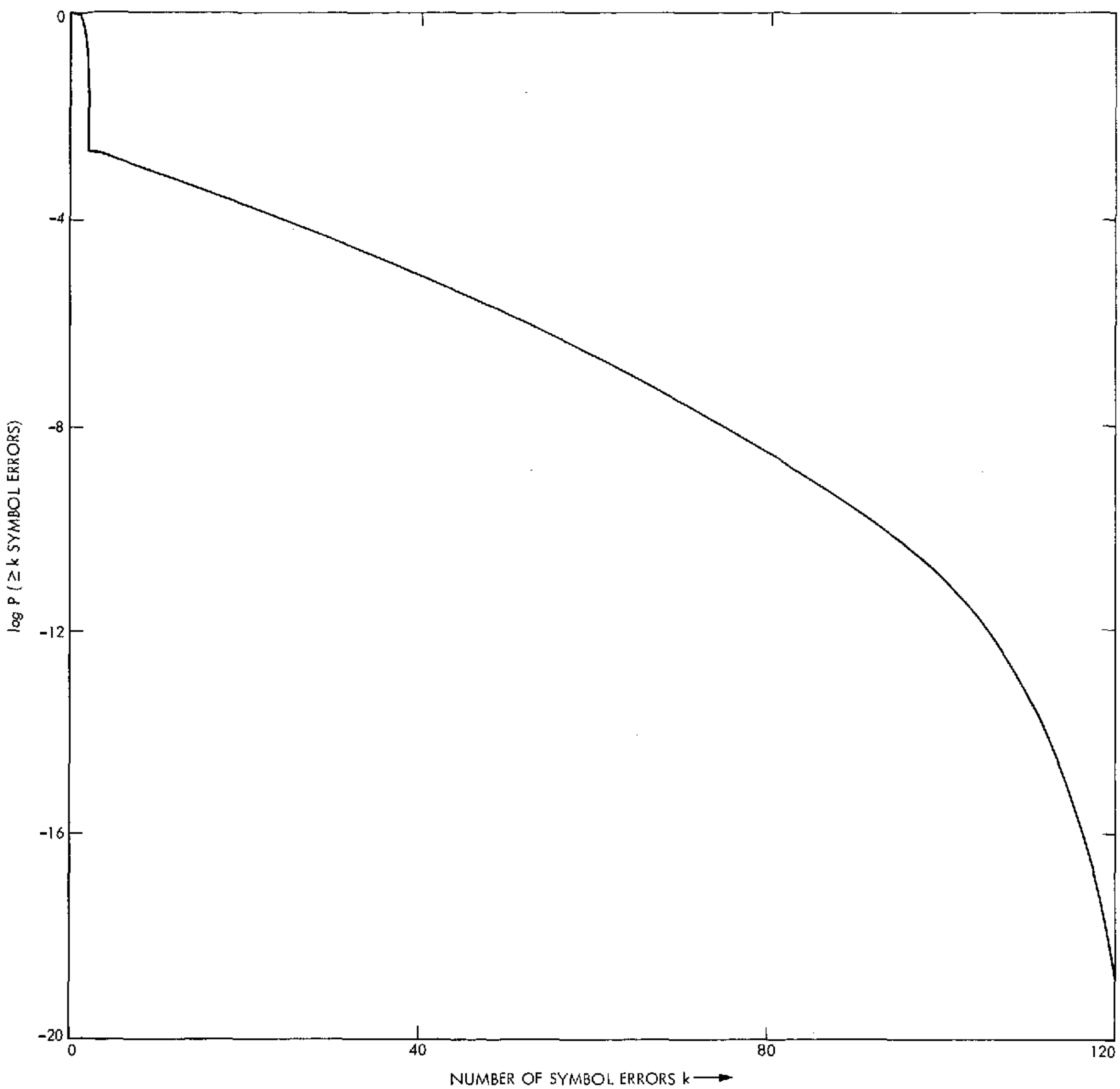

Fig. 20. Distribution of symbol errors (averaged 4.8 kbps; symbol length = 10 bits)

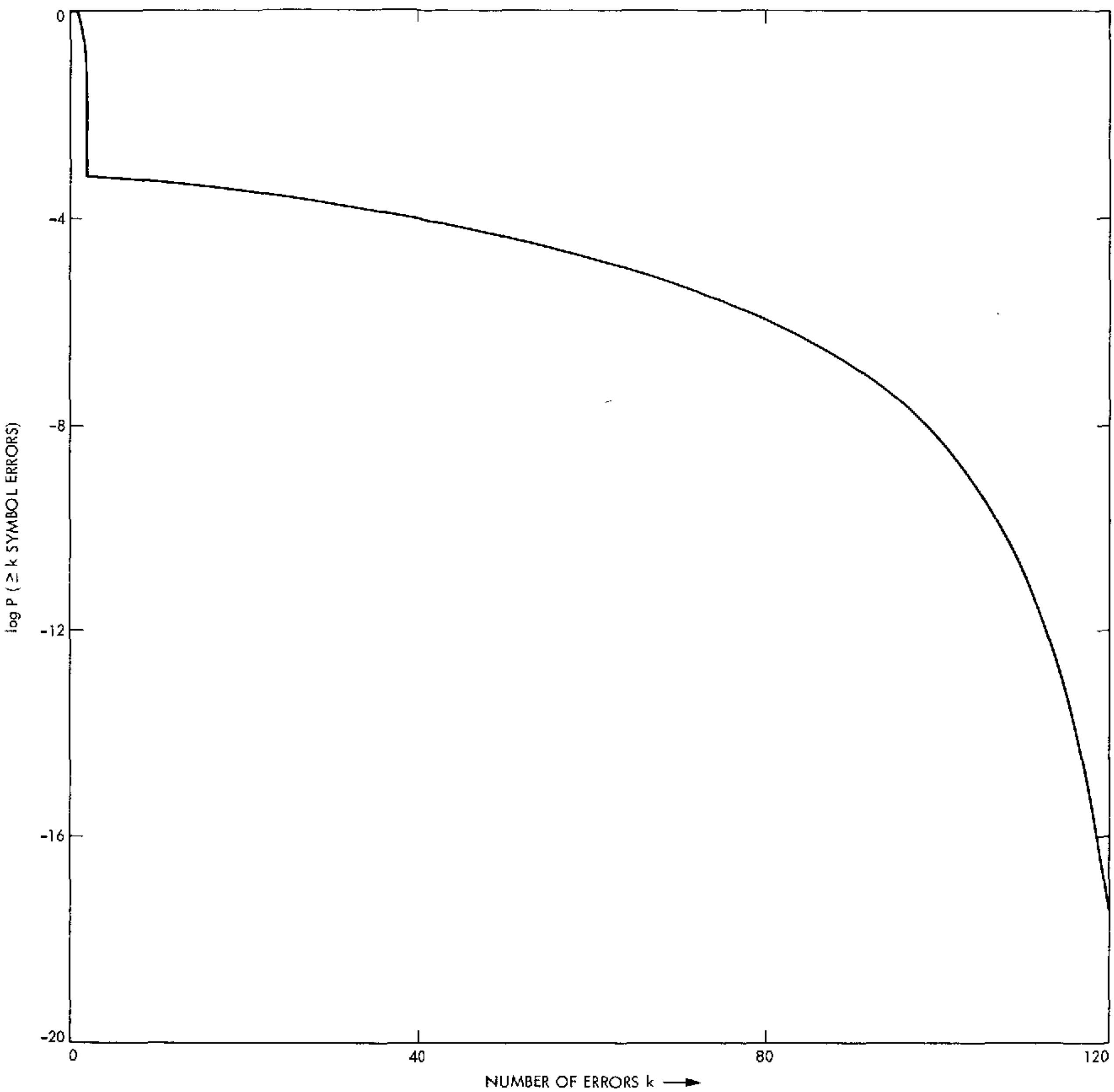

Fig. 21. Distribution of symbol errors (overall 50 kbps; symbol length = 10 bits)

(iii) Correlation of Symbol Errors

This section will not be complete without a word about the autocorrelation of symbol errors denoted by $r^s(k)$, $k = 0,1,2,\ldots$, i.e., $r^s(k)$ = P(symbol k in error|initial symbol error). For if the symbol errors are too highly correlated then we may not be able successfully to shorten the buffers at the transmitter and receiver since we may be forced to store for retransmission many blocks that are in error because occurrence of an error symbol may cause a high number of others to occur in quick succession. On the other hand if the capability or the code employed for forward error correction can handle most of the bursts when they occur it would be desirable to have high values of $r^s(k)$ for small values of k, i.e., just as in bit autocorrelation we would want to have not much longer bursts of symbol errors than the capability of the symbol correcting code so as not to have a problem of buffer over-flow for a buffer of "moderate" size. More detailed consideration of this problem will be studied in another paper.

In Appendix V it is shown that (iii) is given by:

$$r^s(k) = 1 - \frac{(CQ,1)M^{s(k-1)}SR^{s-1}U}{\sum \frac{c_i(1-p_i^s)}{p_i(1-p_i)}} \tag{42}$$

where

$c = (c_1, c_2, c_3, c_4)$

M_{5x5} : transition matrix (Fig. 2)

R_{4x4} : metrix of transitions between the good states only (obtained by deleting the last row and column of M)

$S = M$ - (last column of M)

U = column vector (4 x 1) of 1's

and

$$Q = \begin{pmatrix} \frac{1-p_1^{s-1}}{1-p} & & & 0 \\ & \frac{1-p_2^{s-1}}{1-p_2} & & \\ & & \frac{1-p_3^{s-1}}{1-p_3} & \\ 0 & & & \frac{1-p_4^{s-1}}{1-p_4} \end{pmatrix} .$$

For $s = 6$, Table 22 displays values of $r^s(k)$ for $k = 3, 200$. It is seen that the highest correlation exists in the Red (x10^{-3} error rate) group. In this case it is more probable (0.57) for a symbol error to be correlated with another three symbols away. This correlation reduces to less than 0.08 when $k = 200$ the symbol error probability is 0.00066. In designing a feed-back and retransmission strategy for the GCF note should be taken of the high symbol error correlation in each of the error groups for $k = 3$ (and indeed for all moderately small values of k) and of what we shall call a burst of block errors in the next section. It is only in the Green (x10^{-5} bit rate) group that $r^6(1200) = 0.00001$ is closest to the symbol error probability (0.0000097) showing as in the case of bit correlation that symbol memory is almost 1200. We may further observe that as symbol length becomes longer (for example $s = 10$ in Table 22) the correlation is weaker for small values of k. That is to say that in a great number of bursts of errors on the GCF the longer the symbol lengths the more likely it is that all the errors are contained within (i.e. affects) only a single symbol. Obvious fact! But no such general statement can be made for large values of k except in the case of the Red error group where an error symbol of length $s = 10$ bits is less likely to effect an error in another symbol at distance $k = 200$ away (than in the case when the symbol length is 6 bits).

Table 22. Group symbol-error correlation. The probability starting with initial symbol error (with probability P_1^s) of getting another symbol error in position $k = 3$ or 200 for symbol lengths $s = 6, 10$.

k	$r^s(k)$ for $s = 6$			$r^s(k)$ for $s = 10$		
	3	200	P_1^s	3	200	P_1^s
Red	.5722	.075	.00066	.554	.069	.000779
Amber	.4841	.0028	.000084	.456	.0032	.00010
Green	.4031	.00001	.966(-5)	.358	.000012	.12(-4)
Av. 4.8	.5304	.02	.000122	.51	.0155	.000145
Av. 50	.5643	.1054	.9 6(-4)	.61	.087	.00012

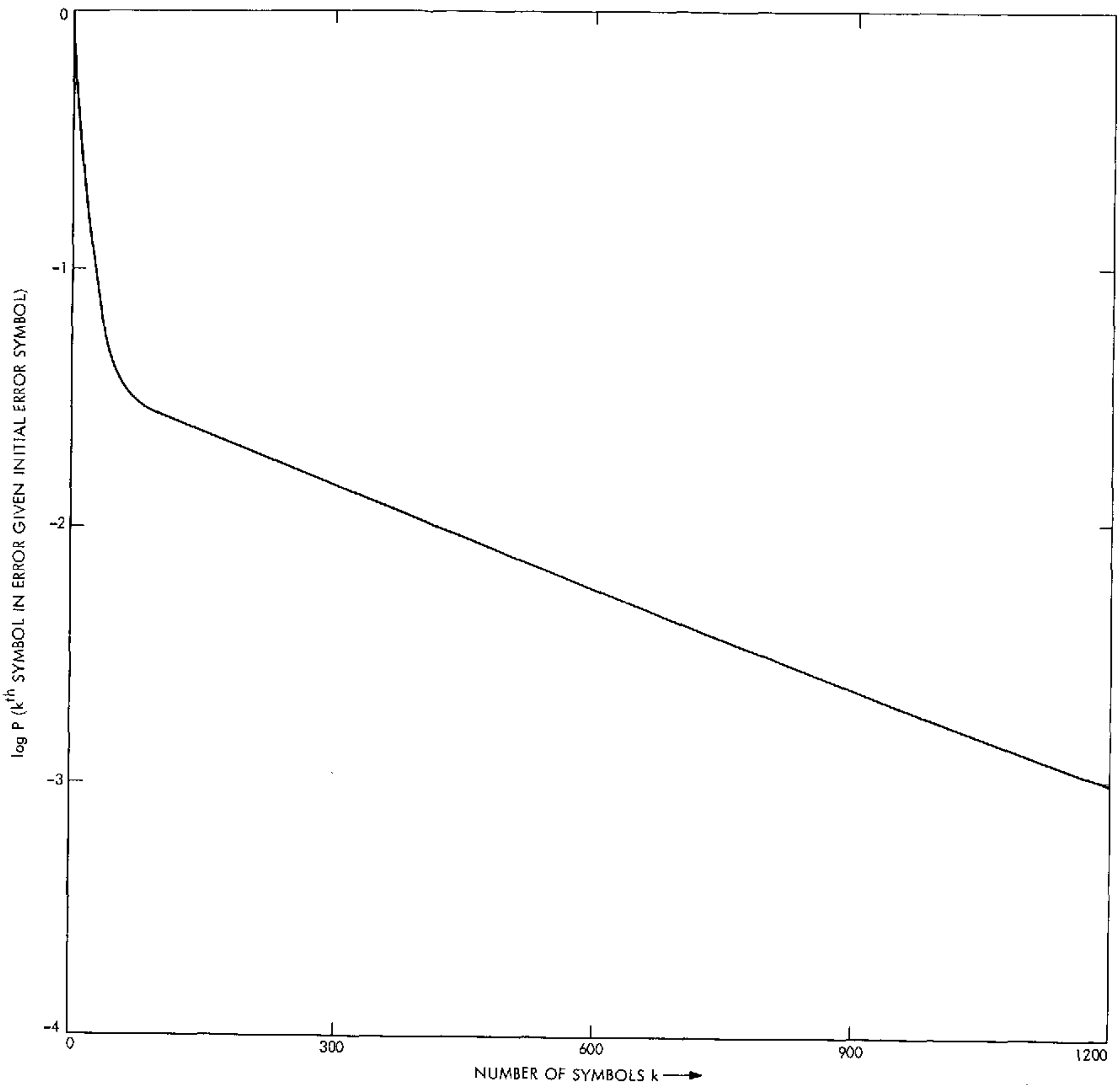

Fig. 22. Auto-Correlation of symbol errors (averaged 4.8 kbps; symbol length = 6 bits)

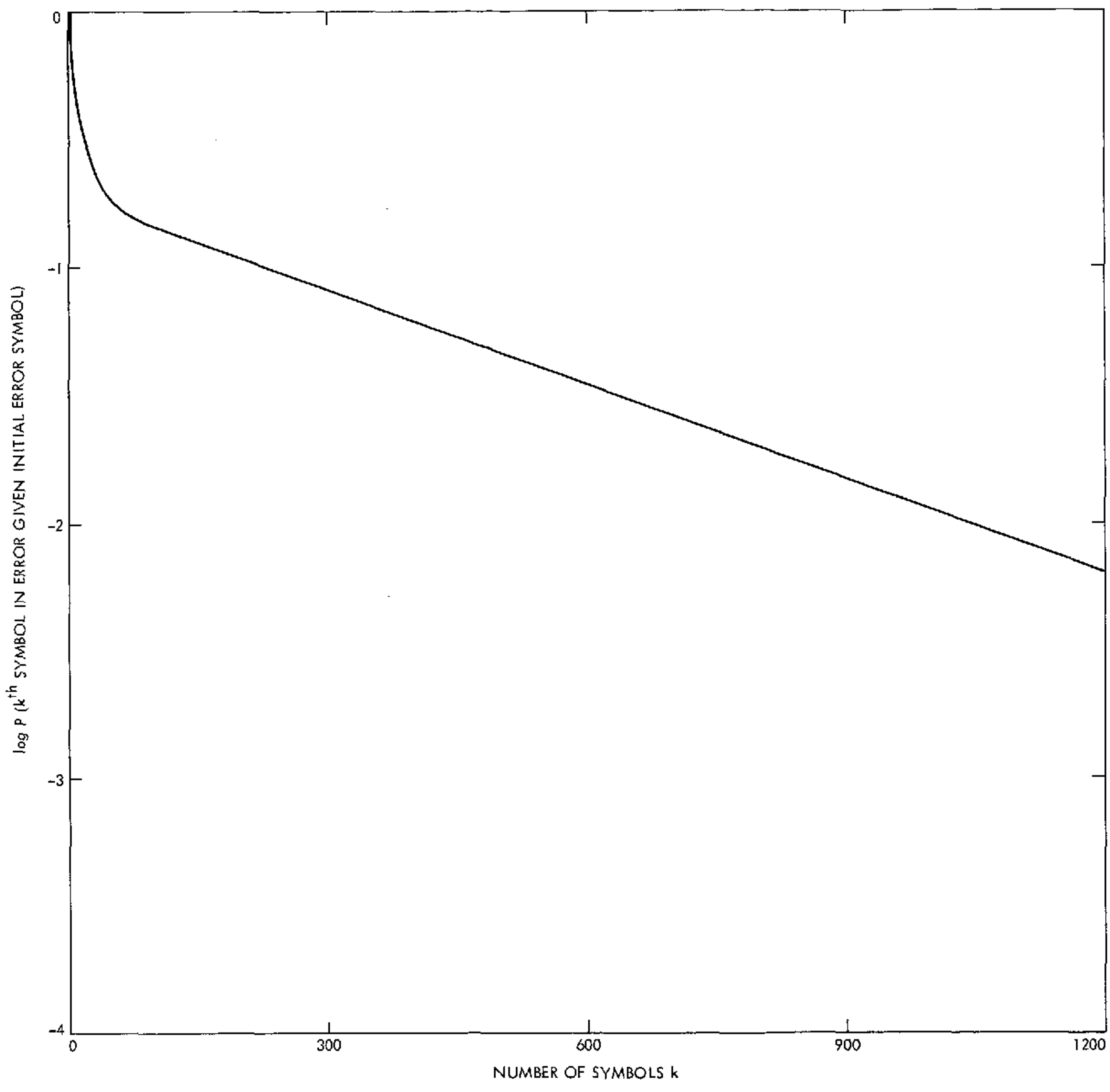

Fig. 23. Auto-correlation of symbol errors (overall 50 kbps; symbol length = 6 bits

iv. Sync Acquisition and Maintenance Probabilities

We are now ready to consider in some detail the problem of sync acquisition and maintenance on the GCF. Of course it is understood that all we can do here is to exhibit the probabilities of each of a number of strategies that are now being proposed for reacquiring synchronization once it is lost and for maintaining it after it is acquired. The hardware problems are beyond the scope of this work!

Specifically we intend to compare the performances of two strategies both based on using a prefix sequence of 24 bits in each of the 1200-bit blocks.

1. The first strategy proposes to accept sync if there are not more than 3-bit errors in the prefix sequence.
2. The second accepts sync if there is at most only one error symbol in the 24-bit prefix considered as four 6-bit symbols.

Our criterion of comparison shall be the efficiency of each of the algorithms in reacquiring a lost sync within a frame of 1200 bits after it is lost and of maintaining it once it is reacquired. Then we shall explore ways of improving the algorithms.

Certainly if the 24-bit prefix of a block agrees exactly with the sync sequence we would have no doubt but that we have acquired synchronization. It is when this agreement is not perfect that we are forced to make a decision as to whether the errors causing the disagreement are due only to the channel noise or to the fact that the 24-bit sequence was not originally the sync sequence. At the extremely high bit rate which will be in use it is not unacceptable to search every 24-bit segment of the incoming data until one of the segments "looks enough" like the sync sequence that we are reasonably sure that we have located the start of a new block. How many errors in a 24-bit segment should we attribute bit positions. And if we are told that the sync sequence contains some errors all we can really assume is that any one of its bits could be the one in error. So that each bit is as likely to be in error as any other bit and then it is as likely to be in error as it is not. Thus we see that the probability of false detection does not depend on the error rate of the channel. Hence to compute this probability we can treat the prefix 24-bit sequence as composed of independent equally likely bits.

We draw several conclusions from our tables.

1. Table 23(a): For $n = 24$, even in the worst (Red) error mode, the algorithm for acquiring sync, once it has been lost, which looks for a 24-bit sequence that looks like the sync sequence up to only 1 bit error would fail to detect sync less than 0.1% of the time within one block after the loss of sync. And in only about 0.18% of the time will this algorithm lock onto the wrong synchronization. However, the algorithm that allows up to 3 errors in the sync sequence will lock onto the wrong synchronization in over 16% of the time although it will hardly fail to identify the true sync sequence. Table 23(b) tells the story for the wide-band 50 kbps circuit.
2. On the other hand the algorithm which looks at the 24-bit sync sequence as four 6-bit symbols and locks onto synchronization if not more than one symbol is in error will lock onto the wrong sync in less than 2% of the time and is equally as efficient as the first algorithm in not failing to identify true synchronization. This algorithm is equally strong for the wide-band channel (see Tables 24(a) and (b)).

We can explain our conclusions in 1 and 2 by the fact that we are dealing with a burst noise channel which is very good indeed during the good modes (see the high thruput in each case analyzed - Tables 23-24). When the errors do to the action of the channel noise and still be well protected against locking onto wrong line sync? In other words how well must a segment look enough like the sync sequence? For if we refuse to lock onto sync any time the 24 bit segment tested disagrees with the sync sequence we may miss acquiring the sync because of the noise in the channel. We call such a mistake FAILURE TO DETECT sync. On the other hand if our algorithm is not stringent enough, in a frame of 1200 bits a 24-bit segment which is not the sync sequence might be accepted as one. This mistake is referred to as FALSE DETECTION of sync.

Probability of failure to detect sync (Tables 23-26) is given, in the case of the first algorithm, by

$$\Pr(\text{failure to detect using the bit count}) = \sum_{k>t} P(k,n) \tag{43}$$

where $P(k,n)$ computed in (37) is the probability of getting k bit errors in an n-bit sync sequence and k_o is the detection level for the algorithm. For the second algorithm this probability is given by:

$$\Pr(\text{failure to detect using algorithm 2}) = \sum_{k>t} P^s(k,n) \tag{44}$$

where $P^s(k,n)$ given in (41) is the probability of k symbol errors in an n-symbol sync sequence. We present these probabilities for each algorithm and each test run, each of the Red, Amber and Green error modes and the averaged high-frequency end wide-band channels. To compute the probability of false detection we use the fact that if our algorithm allows up to k errors in the sync sequence then there are times we acquire sync that we are really choosing any one of the other sequences of 0's and 1's that differ from the sync sequence in $\leq t$

occur however they do so in bunches. The second algorithm uses this fact by allowing up to five errors provided they all occur in only one symbol.

3. We increase the length of the sync sequence to 30 and 36 bits to see when the first algorithm can be efficient (see Tables 25 and 26). At $n = 30$ bits we can very safely allow up to 3 errors in the sync sequence. This procedure will lock onto the wrong sync within one block of sync loss in less than 0.5% of the time while it provides ample protection against failure to identify the right sync sequence. Still for $n = 30$, even allowing up to four errors will lead to wrong identification in less than 4% of the time compared to 16% for allowing just one error at $n = 24$. See the rest of Table 24 for the results for the Amber, Green and the averaged channels. As expected the efficiency of the first algorithm improves with longer sync sequence. Look at the case for $n = 36$.

4. In the case when the sync sequence is taken to be 5 symbols (30 bits) instead of four it is hardly possible to lock onto the wrong synchronization with the algorithm that allows up to one symbol error. Within a block of loss of sync this algorithm will not fail to identify the sync sequence if there is one and will lock onto the wrong sync with about the same probability (<0.0009). But to allow up to 2 symbol errors will detect the wrong sync in about 5% of the time (see Table 26).

We designate by THRUPUT the probability that the sync sequence passes through error-free and it is shown in Tables 23-24. Then the probability that the sync is maintained once it is acquired which depends on the criterion for sync acquisition being used is the thruput plus the probability of getting $\leq t$ errors in the sync sequence. This probability is very high indeed for each of the algorithms examined. Thus we have based our comparison only on their strength of reacquiring synchronization within a frame of 1200 bits after it is lost.

(v) Conclusion

For a sync sequence of length $n = 24$ bits the algorithm that allows not more than one symbol error is preferred to the one that allows up to three bit errors. By looking at symbols instead of individual bits we can reduce the probability of false detection of sync sequence from 16% to less than 2% with correspondingly lower probability of ever failing to lock onto the right sync.

We are even much better protected against ever making a wrong decision if we allow the sync sequence to be up to 5 symbols long. In this case the probability of ever making either type of error is less than 0.0009. To achieve this reliability we cannot allow more than two bit errors in the 30-bit sync sequence if we employ the bit count as our criterion of reacquiring synchronization once it is lost.

It is therefore suggested that we look into the possibility of using a sync sequence of 5 symbols so as to take advantage of the efficiency of the second algorithm.

Table 23(a). Probabilities of failure to detect and of false detection of sync for algorithm which looks at 24-bit prefix and allows ≤t errors. (H-F 4.8 kbps)

		Detection Level				Prefix Thruput
	k	1	2	3	4	
Prob. of false detect		.00175	.021	.163	.9078	
PROB. OF FAILURE TO DETECT	Red	.00093	.00083	.00072	.00062	.99896
	Amber	.00012	.00010	.87E-4	.71E-4	.999858
	Green	.15E-4	.12E-4	.96E-5	.74E-5	.999981
	Av. 48	.00017	.00015	.00013	.00011	.999806
	1	.00229	.00189	.00152	.00118	.99728
	2	.000024	.000019	.000014	.0000099	.999971
	3	.000026	.000025	.000023	.000022	.999973
	4	.00018	.00016	.00014	.00012	.999802
	5	.00015	.000114	.000087	.00006	.999816
	6	.000032	.000029	.000026	.000023	.999965
	7	.00016	.00015	.00013	.00012	.99982
	8	.000016	.000012	.000009	.000007	.999981
	9	.44E-6	.88E-7	.14E-7	.19E-8	.999998
	10	.77E-5	.61E-5	.47E-5	.34E-5	.999990
	11	.66E-4	.61E-4	.55E-4	.50E-4	.999928
	12	.10E-4	.10E-4	.98E-5	.95E-5	.999989
	13	.16E-4	.15E-4	.15E-4	.14E-4	.999983
	14	.92E-4	.78E-4	.65E-4	.52E-4	.999894
	15	.21E-3	.17E-3	.12E-3	.82E-4	.999734
	16	.11E-4	.39E-5	.12E-5	.28E-6	.999975
	17	.49E-4	.45E-4	.40E-4	.35E-4	.999945
	18	.14E-4	.13E-4	.12E-4	.10E-4	.999984
	19	.45E-4	.36E-4	.28E-4	.22E-4	.999944
	20	.60E-4	.23E-4	.73E-5	.19E-5	.999874
	21	.62E-3	.60E-3	.56E-3	.53E-3	.999347
	22	.14E-3	.80E-4	.41E-4	.18E-4	.999789
	23	.00044	.00041	.00039	.00036	.999542
	24	.00015	.0001	.00006	.000029	.999795
	25	.87E-4	.79E-4	.70E-4	.61E-4	.999903
	26	.00028	.00023	.00019	.00016	.999679
	27	.00025	.00022	.00018	.00015	.999707
	28	.00010	.98E-4	.92E-4	.87E-4	.99989
	29	.49E-4	.45E-4	.40E-4	.36E-4	.999946

Table 23(b). Probabilities of failure to detect and of false detection of sync for algorithm which looks at 24-bit prefix and allows ≤ t errors (W-B 50 kbps)

		Detection Level				
	k	1	2	3	4	Prefix Thruput
Prob. of false detect		.00175	.021	.163	.9078	
PROB. OF FAILURE TO DETECT	Av. 50 kbps	.00015	.00013	.00011	.96E-4	.99983
	1	.00013	.00012	.00012	.00011	.999862
	2	.81E-4	.78E-4	.74E-4	.70E-4	.999916
	3	.35E-4	.33E-4	.32E-4	.30E-4	.999964
	4	.54E-5	.45E-5	.37E-5	.30E-5	.999993
	5	.67E-5	.65E-5	.62E-5	.59E-5	.999993
	6					
	7	.70E-5	.66E-5	.61E-5	.57E-5	.999992
	8	.61E-5	.58E-5	.56E-5	.53E-5	.999993
	9	.73E-5	.69E-5	.65E-5	.62E-5	.999992
	10	.49E-4	.47E-4	.44E-4	.42E-4	.99995
	11	.71E-4	.69E-4	.66E-4	.64E-4	.99993
	12	.0011	.00096	.00078	.00059	.998706

Table 24(a). **Probabilities of failure to detect and of false detection of sync for algorithm which looks at a prefix of four 6-bit symbols and allows ≤t errors (H-F 4.8 kbps)**

		Detection Level k 1	2	3	Prefix Thruput
Prob. of false detect		.0177	1	1	
PROB. OF FAILURE TO DETECT	Red	.00079	.000546	.00029	.998974
	Amber	.0001	.63E-4	.29E-4	.999858
	Green	.12E-4	.96E-4	.26E-5	.999982
	Av. 4.8 kbps	.00015	.96E-4	.47E-4	.999803
	1	.00186	.0011	.44E-3	.997276
	2	.193E-4	.12E-4	.44E-5	.999972
	3	.27E-4	.227E-4	.16E-4	.999968
	4	.00016	.00011	.623E-4	.999798
	5	.00011	.586E-4	.225E-4	.999814
	6	.31E-4	.225E-4	.13E-4	.999961
	7	.00015	.00012	.68E-4	.999821
	8	.13E-4	.791E-5	.30E-5	.999981
	9	$x10^{-6}$	$x10^{-7}$	$x10^{-8}$	.999997
	10	.74E-5	.40E-5	.15E-5	.999988
	11	.49E-4	.38E-4	.24E-4	.99994
	12	.75E-5	.69E-5	.58E-5	.999992
	13	.19E-4	.17E-4	.14E-4	.999979
	14	.74E-4	.48E-4	.22E-4	.9999
	15	.00018	.00011	.38E-4	.99974
	16	.85E-5	.19E-5	.18E-6	.999976
	17	.40E-4	.32E-4	.18E-4	.99995
	18	.12E-4	.834E-5	.53E-5	.99997
	19	.34E-4	.18E-4	.68E-5	.999943
	20	.46E-4	.11E-4	.11E-5	.999875
	21	.00059	.00051	.00038	.99933
	22	.00011	.45E-4	.88E-5q	.999786
	23	.00040	.00033	.00023	.99955
	24	.00013	.59E-4	.14E-4	.999798
	25	.784E-4	.61E-4	.35E-4	.99999
	26	.00022	.00014	.63E 4	.99968
	27	.00021	.00013	.65E-4	.99971
	28	.96E-4	.82E-4	.624E-4	.99989
	29	.44E-4	.351E-4	.22E-4	.999946

Table 24(b). Probabilities of failure to detect and false detection of sync for algorithm which looks at a prefix of four 6-bit symbols and allows ≤t errors (W-B 50 kbps)

		Detection Level k: 1	2	3	Prefix Thruput
Prob. of false detect		.0177	1	1	
PROB. OF FAILURE TO DETECT	Av. 50 kbps	.00012	.79E-4	.36E-4	.999845
	1	.00011	.95E-4	.72E-4	.9998686
	2	.812E-4	.69E-4	.52E-4	.999908
	3	.35E-4	.292E-4	.22E-4	.9999608
	4	.11E-5	.611E-6	.26E-6	.999997
	5	.89E-5	.782E-5	.63E-5	.9999899
	6	.39E-5	.33E-5	.261E-5	.9999955
	7	.47E05	.37E-5	.27E-5	.999994
	8	.79E-5	.69E-5	.55E-5	.999991
	9	.64E-5	.53E-5	.42E-5	.999993
	10	.48E-4	.41E-4	.30E-4	.999945
	11	.55E-4	.49E-4	.40E-4	.999939
	12	.001	.00072	.00031	.99867

Table 25. Probabilities of failure to detect and of false detection of sync for algorithm which looks at a prefix of length n = 24, 30 or 36 bits and allows ≤k errors.

n	k	Prob. of False Detect	Prob. of Failure to Detect Red	Amber	Green	Av. 4.8	Av. 50
24	1	.00175	.00093	.00012	.15E-4	.00017	.00015
	2	.021	.00083	.00010	.12E-4	.00015	.00013
	3	.163	.00072	.87E-4	.96E-5	.00013	.00011
	4	.9078	.00062	.71E-4	.74E-5	.00011	.96E-4
	5	.9078	.00062	.71E-4	.74E-5	.00011	.96E-4
30	2	.00051	.00092	.00012	.14E-4	.00017	.00015
	3	.0049	.00082	.00010	.12E-4	.00015	.00013
	4	.0348	.00073	.86E-4	.93E-5	.00013	.00012
	5	.1901	.00063	.71E-4	.73E-5	.00011	.99E-4
	6	.8371	.00054	.58E-4	.56E-5	.9E-4	.82E-4
	7	1	.00045	.46E-4	.42E-5	.73E-4	.66E-4
36	3	.00013	.00091	.00011	.13E-4	.00016	.00015
	4	.0011	.00082	.99E-4	.11E-4	.00014	.00013
	5	.0075	.00073	.85E-4	.89E-5	.00013	.00012
	6	.041	.00064	.71E-4	.72E-5	.00011	.00010
	7	.182	.00055	.59E-4	.56E-5	.92E-4	.86E-4
	8	.694	.00047	.48E-4	.43E-5	.77E-4	.71E-4
	9	1	.00039	.38E-4	.32E-5	.62E-4	.57E-4

Table 26. Probabilities of failure to detect and of false detection of sync for algorithm which looks at a prefix of n 6-bit symbols (n = 4, 5, 6) and allows ≤k errors.

n	k	Prob. of false Detect	Prob. of Failure to Detect Red	Amber	Green	Av. 4.8	Av. 50
4	1	.0177	.00079	.0001	.12E-4	.00015	.00012
	2	1	.000546	.63E-4	.64E-5	.96E-4	.79E-4
5	1	.00034	.00089	.00012	.14E-4	.00017	.00014
	2	.044	.00066	.79E-4	.84E-5	.00012	.0001
	3	1	.00044	.47E-4	.45E-5	.75E-4	.62E-4
6	1	.64E-5	.00097	.00013	.15E-4	.00018	.00015
	2	.0010	.00076	.93E-4	.10E-4	.00014	.00012
	3	.086	.00056	.62E-4	.61E-5	.97E-4	.83E-4
	4	1	.00036	.35E-4	.31E-5	.59E-4	.48E-4

Section VII

BURST DISTRIBUTION

From what has been said up to now it is clear that the single most important distribution on the GCF is the Burst Distribution. Indeed it is precisely because of this reason that attempts are being made to design forward and feedback error correcting strategies that have the capability of removing these bursts of errors for improved communication. The fact is that absolutely every step must be taken to understand the nature of these bursts: how long they are, how dense the errors within them are, and particularly how many standard 1200-bit blocks are affected each time the channel enters into this bursty mode. These and related questions will be answered in this section.

But before we start, let us try to fix ideas of what exactly we shall refer to as a burst of errors. We define a burst as a sequence of bits (a) beginning and ending with an error, (b) separated from the nearest preceding and following error by a gap of no less than some number, say G, called the guardspace, and (c) containing within it no gap equal to or greater than the guardspace.

Immediately a number of questions spring to mind. For example, how large must a guardspace be? Or for a given set of data what is the criterion for choosing an optimum value of G? The fact is that since optimality of G can be defined only with respect to a set of criteria which in turn are based on what we consider important, a value of G which is optimal in one sense may necessarily not be optimal in some other. We shall illustrate this point with two examples:

1. Following Stern's intuitive reasoning in [15] the optimum G should separate the data into bursts with a density of errors which is much higher than the "background" error rate. The background error rate is used here

to refer to the error rate which would result if each sequence of bits defined as a burst were called and replaced with a single bit error. The proportion of bursts in a test, the density of errors in a burst and the distribution of burst lengths all should not be too sensitive to the G that is optimum. By considering the variation of the background error rate with G alone it is possible to find a range of G values that leaves the background error rate constant. Take as optimum G any value in this range.

2. We are primarily interested in choosing as guardspace that G that will divide the data into bursts a high proportion of which are less than the burst correcting capability of the code employed on the channel. To be more precise let b be the burst correcting capability of a code C. By this we mean relative to the guardspace G, b is the largest integer for which every noise sequence containing only bursts of length b or less is correctly decoded. See Gallager [16]. So in this case we will be interested in choosing a G for which b is maximal for the code C. If for example, the capability of the code C is as high as $\frac{1}{3} G$ i.e., $b = \frac{G}{3}$, then using the relation

$$\frac{G}{b} \geq \frac{1+R}{1-R} \tag{45}$$

in [16], which connects the rate R, G and B, we see that the rate R, cannot be more than 0.5. Which is too low of course. To achieve a rate R of 0.9, b cannot be more than $G/19$. That is, the error correcting capability of a code with rate $R = 0.9$ cannot exceed $b = G/19$.

Table 27(a). Optimal guardspace, 4800 bps data

	G = 400; b = 133					G = 3600; b = 1200					G = 4800; b = 1600				
	# of burst	# ≤ b	# > b	max. burst	# of errors in max. burst	# of burst	# ≤ b	# > b	max. burst	# of errors in max. burst	# of burst	# ≤ b	# > b	max. burst	# of errors in max. burst
1	322	151	171	6133	141	100	79	21	217362	3550	85	60	25	217362	3550
2	11	10	1	150	63	11	11	0	150	63	11	11	0	150	63
3	8	5	3	535	246	7	7	0	1158	9	7	7	0	1158	9
4	93	79	14	1077	126	89	87	2	6164	477	86	81	5	6164	477
5	122	90	32	654	27	109	103	6	7433	147	109	103	6	7433	147
6	23	18	5	565	272	20	20	0	884	14	19	18	1	4780	32
7	109	87	22	861	299	76	62	14	4787	15	66	47	19	12866	309
8	22	21	1	148	66	13	8	5	3667	12	13	10	3	3667	12
9	10	9	1	435	6	8	7	1	1615	15	8	7	1	1615	13
10	7	1	6	1740	70	5	3	2	5082	213	5	3	2	5082	213
11	1	1	0	47	22	-	-	-	-	-	-	-	-	-	-
12	8	4	4	570	164	5	3	2	2432	352	5	4	1	2432	352
13	28	20	8	868	17	15	11	4	6205	143	13	10	3	14291	248
14	214	201	13	1249	36	174	154	20	9268	33	167	149	18	15562	39
15	23	23	0	27	6	23	23	0	27	6	23	23	0	27	6
16	37	33	4	783	324	27	21	6	4559	51	26	20	6	5456	43
17	4	2	2	738	109	4	4	0	738	109	4	4	0	738	109
18	31	27	4	389	15	20	18	2	3654	24	19	17	2	6359	3
19	37	36	1	143	6	28	23	5	3760	12	26	20	6	4520	9
20	34	12	22	1347	537	21	17	4	8805	1117	20	15	5	8805	1117
21	29	5	4	305	28	18	13	5	4742	12	18	13	5	4742	12

Table 27(a) - Continued

	G = 400; b = 133					G = 3600; b = 1200					G = 4800; b = 1600				
	# of burst	# ≤ b	# > b	max. burst	# of errors in max. burst	# of burst	# ≤ b	# > b	max. burst	# of errors in max. burst	# of burst	# ≤ b	# > b	max. burst	# of errors in max. burst
22	102	39	63	2506	535	55	33	22	13718	906	50	31	19	22459	979
23	47	37	10	5157	346	27	17	10	6178	367	27	18	9	6178	367
24	18	17	1	518	236	16	14	2	2126	12	16	14	2	2126	12
25	213	186	27	1265	444	183	169	14	14761	105	180	164	16	14761	105
26	109	87	22	1944	295	89	76	13	6355	162	86	72	14	1167	289
27	21	19	2	611	286	19	18	1	2464	6	19	18	1	2464	6
28	13	9	4	1758	211	11	10	1	5588	223	11	10	1	5588	223

Table 27(b). Optimal guardspace, 50 kbps wideband data.

	G = 400; b = 133					G = 3600; b = 1200					G = 4800; b = 1600				
	# of burst	# ≤ b	# > b	max. burst	# of errors in max. burst	# of burst	# ≤ b	# >b	max. burst	# of errors in max. burst	# of burst	# ≤ b	# >b	max. burst	# of errors in max. burst
1	34	17	17	473	248	30	27	3	378	80	28	25	3	7018	112
2	13	4	9	2008	681	11	8	3	4651	98	11	8	3	4651	98
3	15	6	9	302	193	12	10	2	3570	160	12	10	2	3570	160
4	2	0	2	161	23	2	2	0	161	23	2	2	0	161	23
5	3	2	1	210	95	3	3	0	210	95	3	3	0	210	95
6	2	1	1	266	61	2	2	0	266	92	2	2	0	266	92
7	2	1	1	176	90	2	2	0	176	90	2	2	0	176	90
8	2	0	2	346	142	2	2	0	346	142	2	2	0	346	142
9	3	2	1	319	65	3	3	0	319	65	3	3	0	319	65
10	21	9	12	326	143	19	18	1	3507	155	19	18	1	3507	155
11	17	7	10	687	160	12	9	3	3506	332	12	9	3	3506	332
12	100	50	50	13636	2361	27	16	11	70931	5603	26	16	10	124744	10477

So our criterion for choosing a G which is optimum for a code C with rate R is to choose that G for which the lengths of a desired proportion of the bursts are at most b bits where $b \leq \frac{1-R}{1+R} G$. We cannot allow G to be too large even though theoretically that would enable us to correct all the bursts. The compromise is to find an implementable code having a desired R and then a G giving a maximal b for that code and for the frame size (or implementable multiple thereof) on the channel. For the GCF the frame-size is 1200 bits.

We do not want to leave the reader with the impression that b can always be used as criterion of the effectiveness of a code against burst noise. For example, on a channel where long bursts containing relatively few errors are far more likely than short bursts containing many errors one would prefer a code capable of correcting the likely longer bursts at the expense of the less likely short bursts.

As an illustration of the magnitude of G we are talking about I have fixed a $b = G/3$ and then found the value of G for which a high proportion of the bursts have lengths less than or equal to b. The results are contained in Tables 27(a) and (b). In both the high-speed and wideband circuits a guardspace of 400 bits is too short for identifying bursts. It does seem however that in both cases a G of 3600 bits is adequate. For this value of optimal G the bursts are longer in the 4800 bps data (maximal burst length is 217,362 bits) than in the 50 kbps data with a maximal burst length of 70,931 bits.

We now find the average length of the bursts, the density of errors within them and how many 1200-bit blocks are affected each time the channel enters a burst. Specifically we shall calculate the

(i) distribution and mean of burst lengths

(ii) P(k errors in a given burst of length n) and its mean

and

(iii) the block burst distribution.

i. Distribution and Mean of Burst Lengths

Let

$$L(n) = P(\text{burst of length } n).$$

Then as shown in Proposition 1 of Appendix VI, L(n) is given by

$$L(n) = \begin{cases} 0 \quad \text{for} & n \le 0 \\ U(G)\bar{L}(n); & n \ge 1 \end{cases} \tag{46}$$

where

$$\bar{L}(n) = \sum_{\ell=0}^{\min(G-1,n-2)} V(\ell)\bar{L}(n-\ell-1); \quad n \ge 2$$

$$\bar{L}(1) = 1$$

and $U(k)$, $V(k)$, $k \ge 0$ are given by (11).

Typical test runs give too few bursts to make comparison between the model and empirical distributions realistic. For instance for G = 400, 15 of the 29 error runs at 4800 bps have less than 30 bursts each; for G = 3600 bits, 19 of the runs each has less than this number. It therefore seems that the appropriate burst distribution to compare with the data is a variant of (46) which is a function of the total length of bits transmitted in the particular run. Nevertheless comparative graphs of model and empirical distributions are plotted (typical ones are shown in Figures 28, 29 and 31).

The mean burst length is given by

$$\sum_{n\geq 1} nL(n) = U(G) \sum_{n\geq 1} n\bar{L}(n) \tag{47}$$

$$= U(G) \frac{1+V_G'(1)}{(1-V_G(1))^2}$$

and the variance is

$$U(G)R''(1) + U(G)R'(1) - [U(G)R'(1)]^2 \; ; \tag{48}$$

where

$$V_G(1) = \sum_i \frac{c_i(1-p_i^G)}{p_i}$$

$$V_G'(1) = \sum_i \frac{c_i}{1-p_i} \{1 - p_i^G - G(1-p_i)p_i^{G-1}\}$$

and, $R'(1)$ and $R''(1)$ are given by expressions (VI.13) and (VI.14) in the Appendix.

The wide variation in burst lengths in each of the error runs (between 1 bit and the maximum lengths shown in Table 27(a) and (b)) explains the relatively low mean burst lengths and high standard deviation in each of the error groups (Table 28). It also explains why the (model) standard deviation is higher than the mean, it being possible to have no burst at all in a run (indeed)

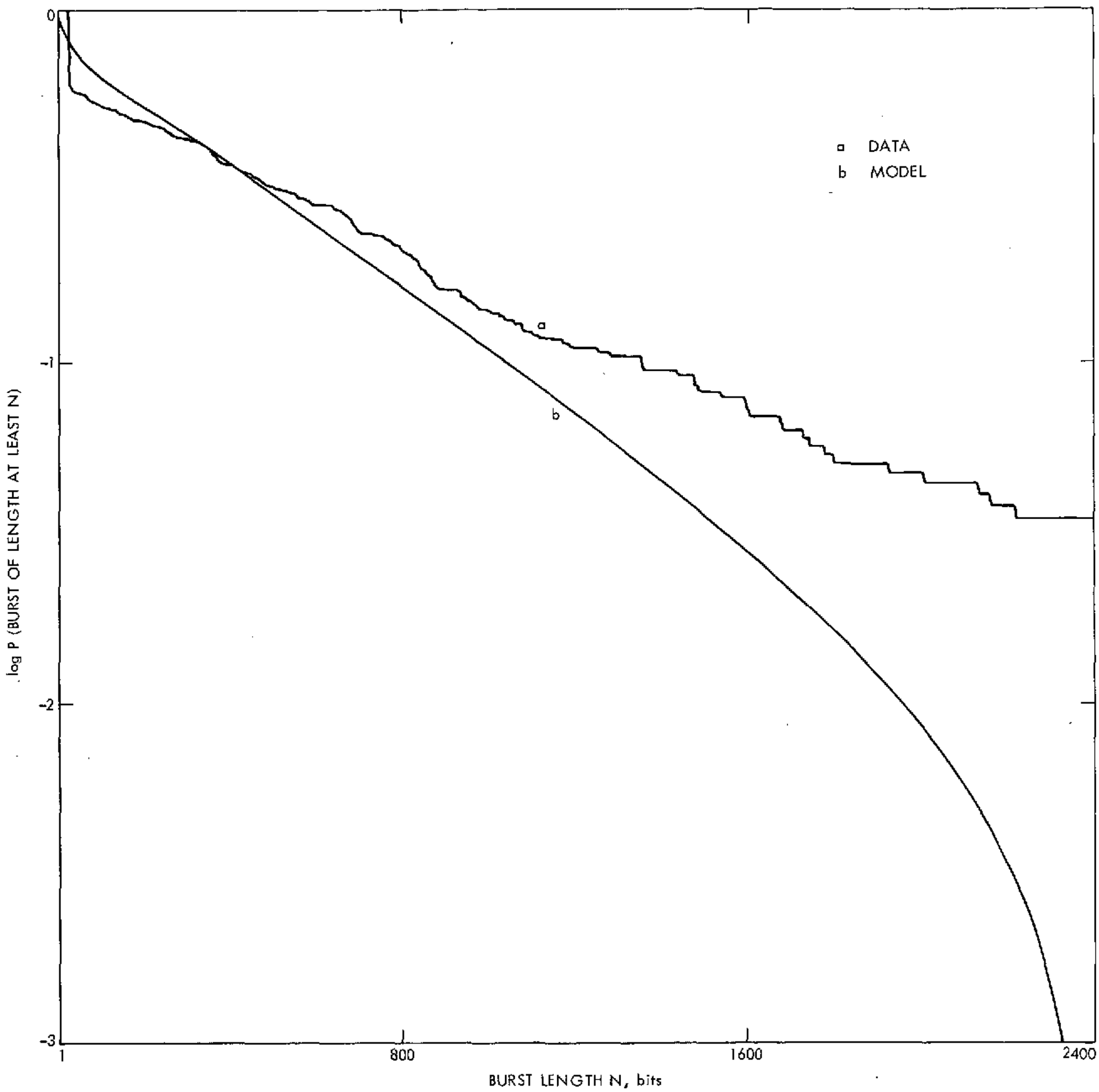

Fig. 24. Distribution of burst lengths (4.8 kbps line; Red group; G = 400)

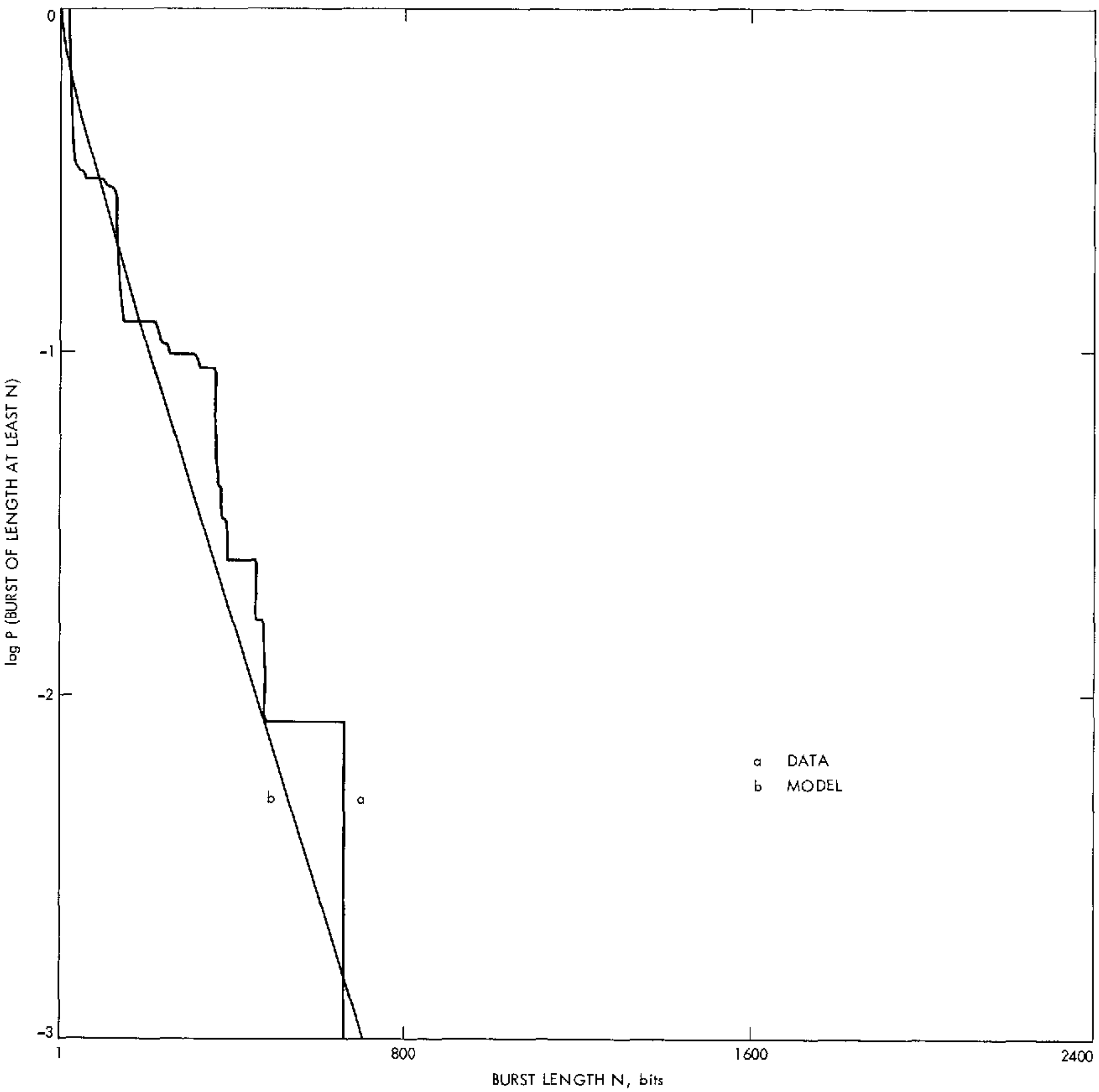

Fig. 25, Distribution of burst lengths (4.8 kbps line; Amber group; G = 400)

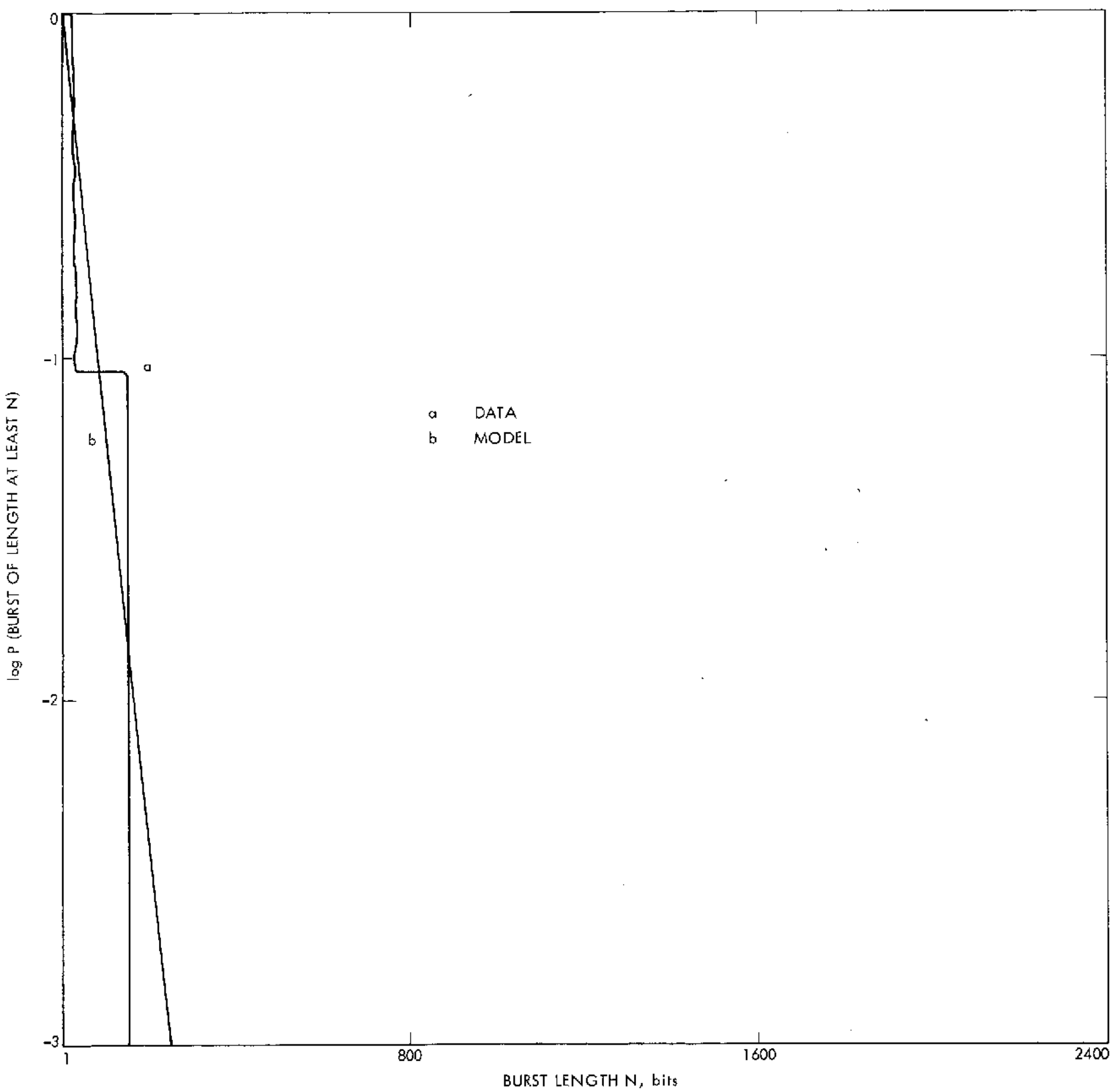

Fig. 26. Distribution of burst lengths (4.8 kbps line; Green group; G = 400)

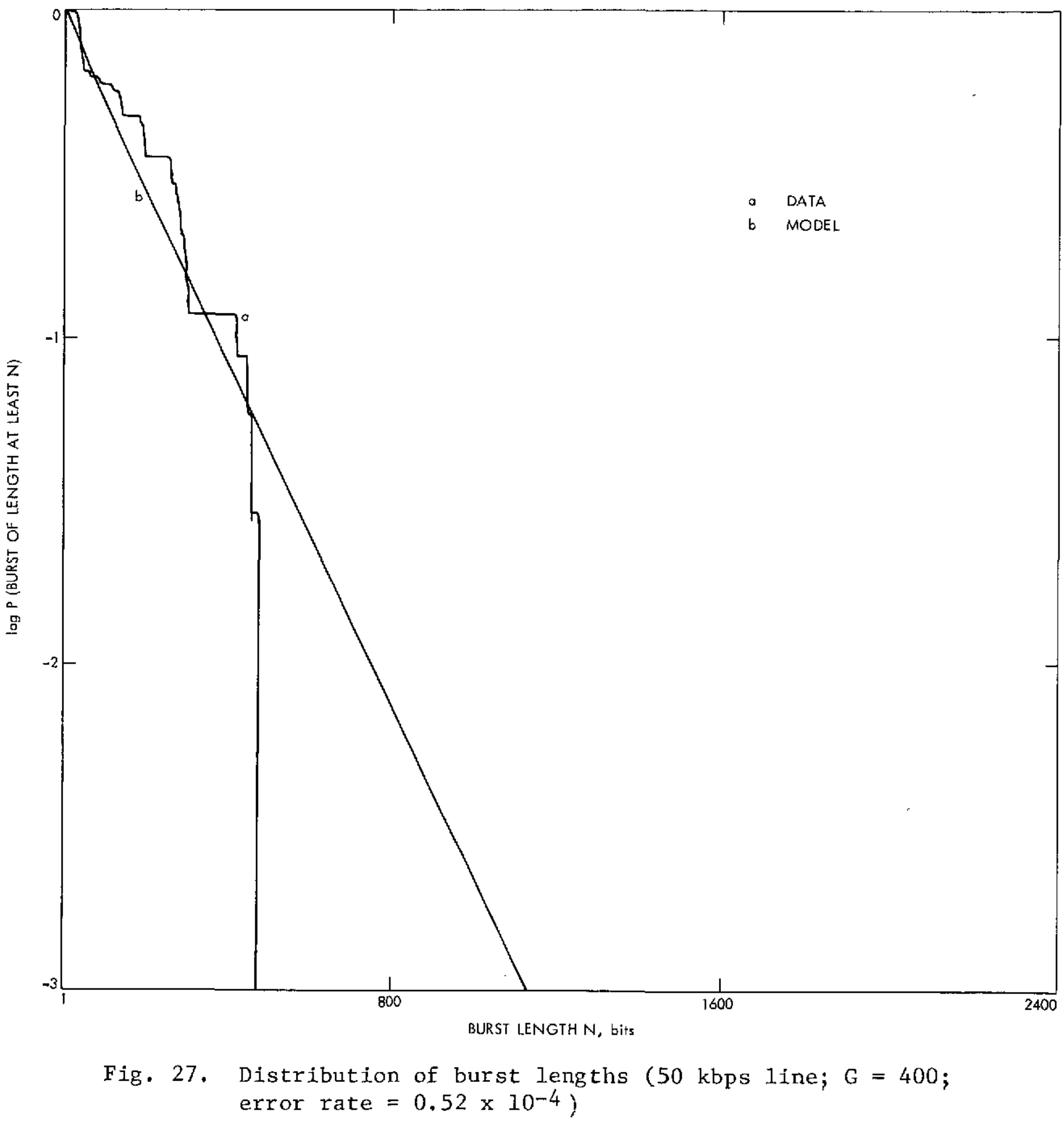

Fig. 27. Distribution of burst lengths (50 kbps line; G = 400; error rate = 0.52×10^{-4})

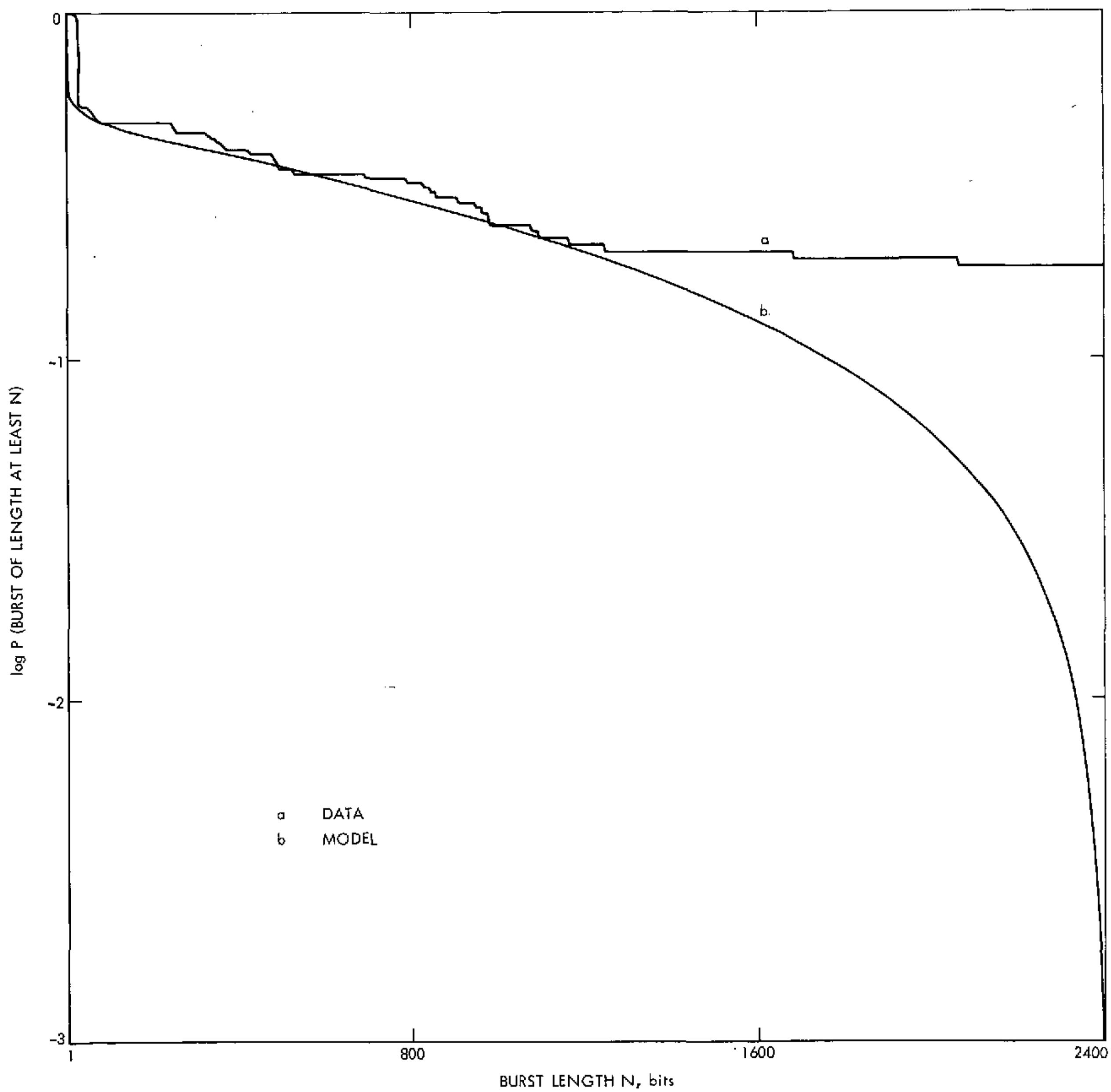

Fig. 28. Distribution of burst lengths (4.8 kbps line; Red group; G = 3600)

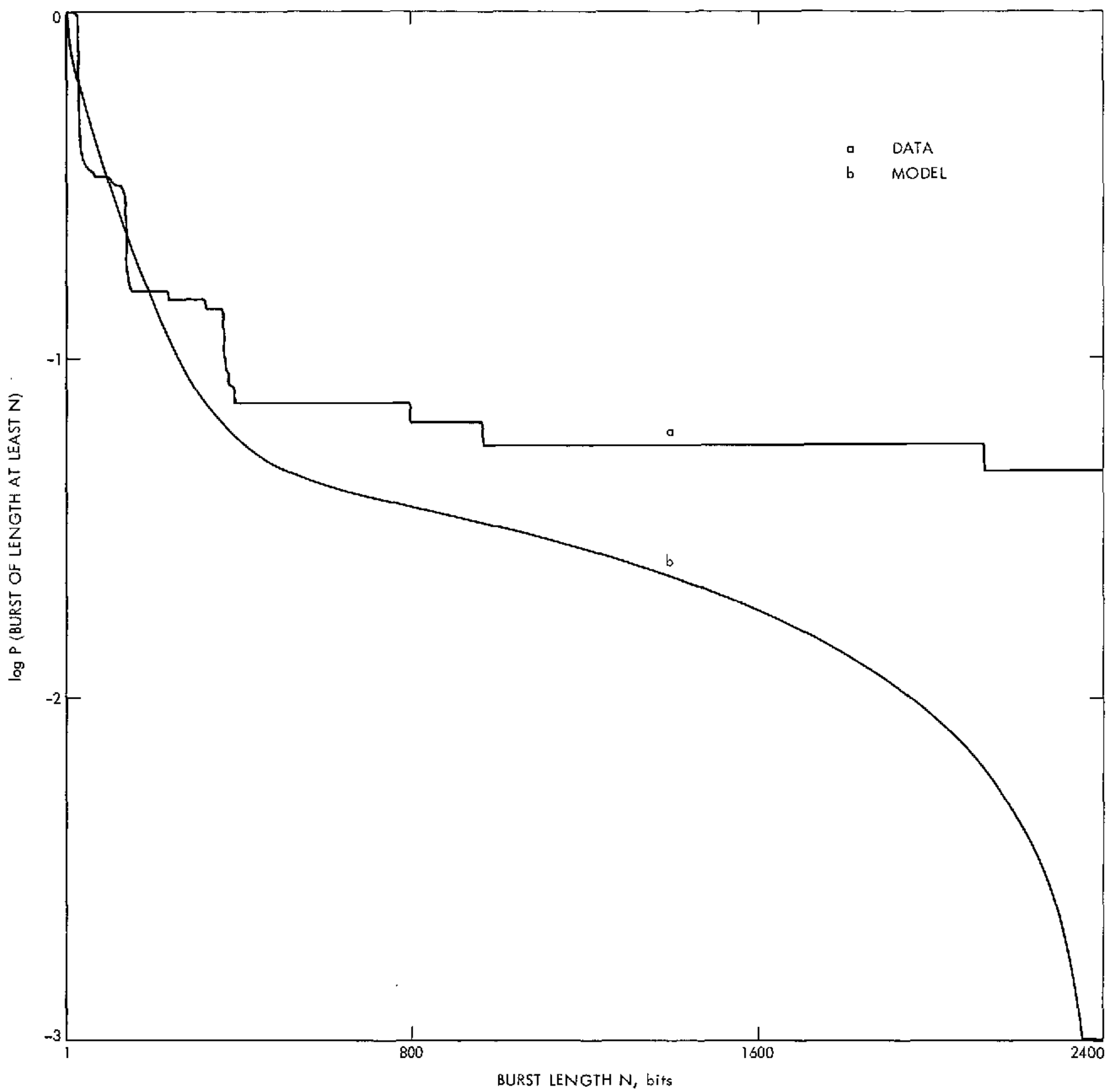

Fig. 29. Distribution of burst lengths (4.8 kbps line; Amber group; G = 3600)

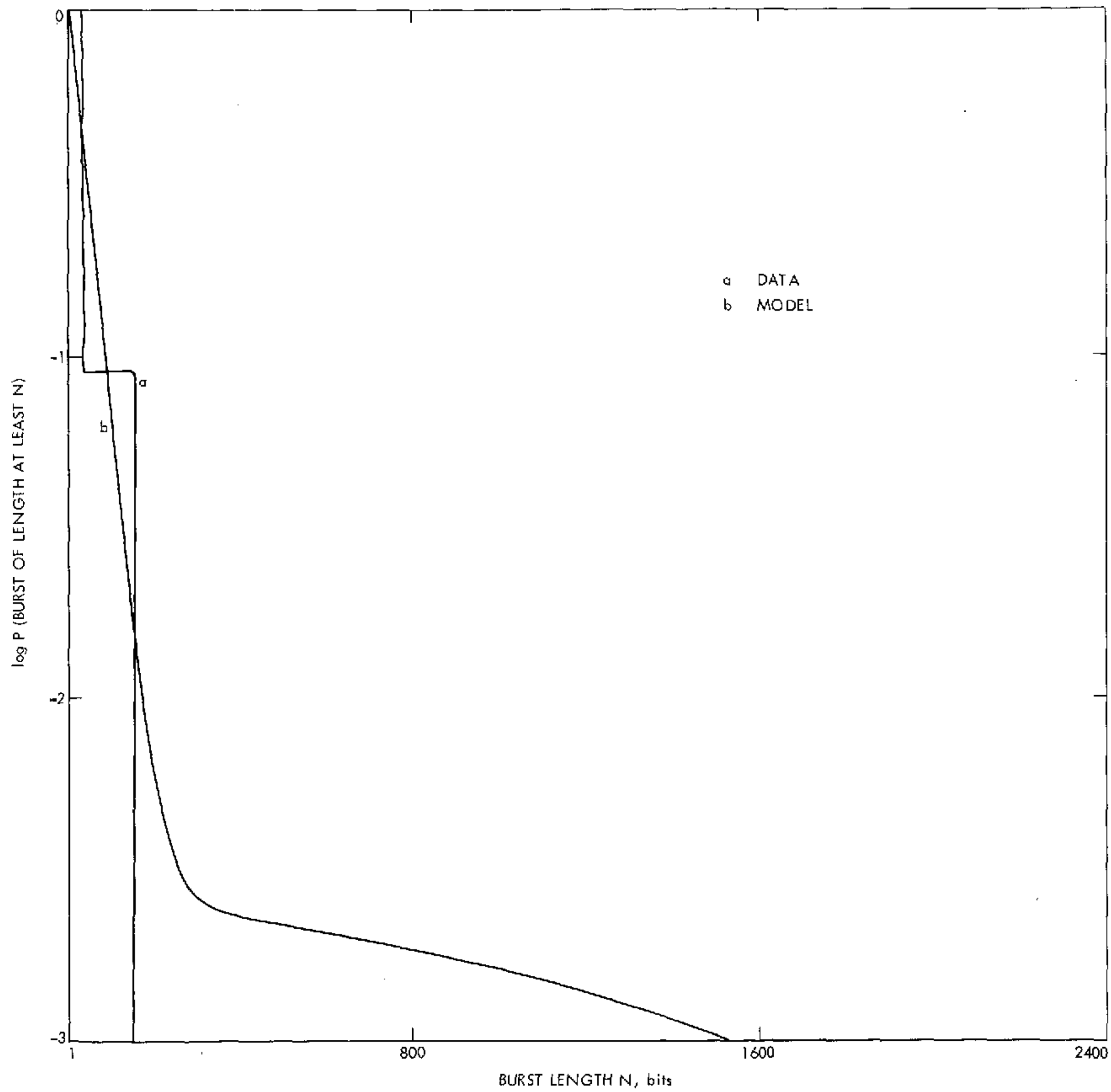

Fig. 30. Distribution of burst lengths (4.8 kbps line; Green group; G = 3600)

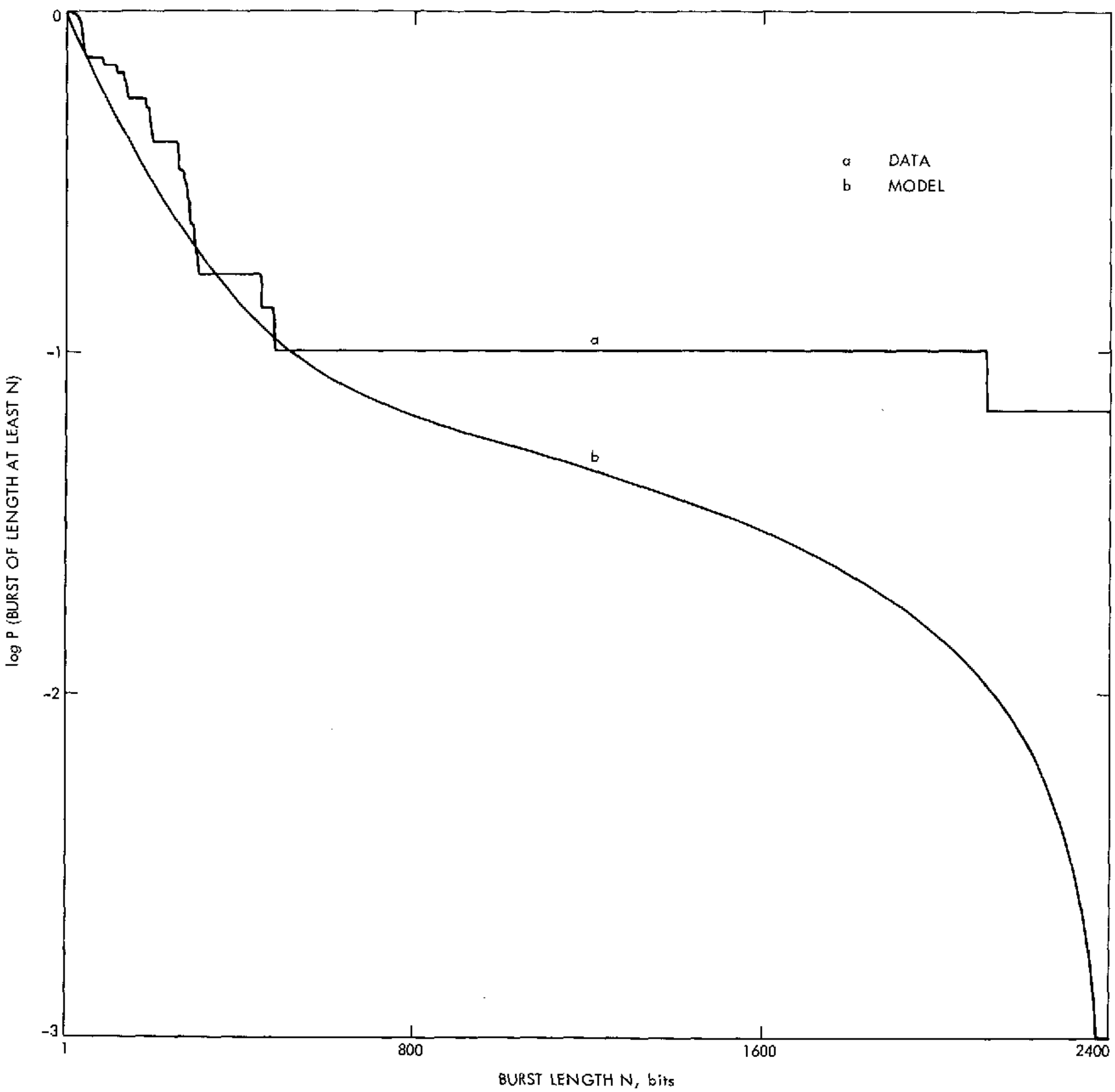

Fig. 31. Distribution of burst lengths (50 kbps line; G = 3600; error rate = 0.52×10^{-4})

there are two such runs without errors). The discrepancy between the data and model mean burst lengths can be explained in terms of the modem fixed errors we talked about earlier; note that the data values are therefore expected to be higher than their predicted values (as Table 28 shows).

Table 28. Mean burst length, 4800 bps circuit, G = 400 bits

Group	Data	Model	
	mean length	mean length	standard deviation
Red	389.0	340.0	402
Amber	94.0	70.0	82
Green	50.0	41.0	41
Average 4800 bps	185.7	135.0	177

As Tables 27(a) and (b) show, most of the bursts occurring are very short. In both the 4800 bps high-speed and the 50 kbps wideband data, a high percentage of the bursts is less than 133 bits in length (using a guardspace of 400 bits). We note here also the effect of the modem fixed errors which, at this value of G, gives a large number of bursts exactly 24 bits in length (each random error causes errors in exactly 18 and 23 bit positions away from it) and containing exactly three errors.

ii. Distribution of Errors in a Burst and Its Mean

Using a guardspace of only 40 bits it has been shown in [1] for the 4800 bps data that the ratios bad/good bits in the bursts average to 41%, 44% and 45% for respectively the Green, Amber and Red error groups with overall average of about 44%. Let us see here how the guardspace affects this error density in the bursts.

To do this we find the distribution of errors in a burst. Then by Proposition 2 of the Appendix we have

$$P(k \text{ errors in a given burst of length } n) = \frac{\overline{Q}(k,n)}{\overline{L}(n)}$$

where

$$\overline{Q}(k,n) = \begin{cases} 0 & \text{if} \quad k = 0,\ n \geq 0; \quad \text{or} \quad k > n \\ 1 & \text{if} \quad n = k = 1 \\ \sum\limits_{\ell=0}^{\min(G-1,n-k)} V(\ell)\overline{Q}(k-1,\ n-\ell-1); & n \geq k \geq 2 \end{cases} \tag{49}$$

and $\overline{L}(n)$ is given by (46).

The mean number of errors in a burst of length n, $\overline{K}_n$, is given by

$$\overline{K}_n = \begin{cases} 1 & \text{if} \quad n = 1 \text{ or } 2 \\ \dfrac{\sum\limits_{k=1}^{n} \overline{L}(k)\overline{L}(n-k+1)}{\overline{L}(n)}; & n \geq 3 \end{cases} \tag{50}$$

We pick an $n = 2400$ and plot the graphs of the probability of errors in a burst of length n for $G = 400$ and 3600 bits and each of the error groups. Typical plots are shown in Figures 32, 33 and 34. The mean number of errors for this length of burst and the bad/good ratio are shown in Table 29 which also shows that the bad/good ratio decreases with increasing guardspace as should be expected.

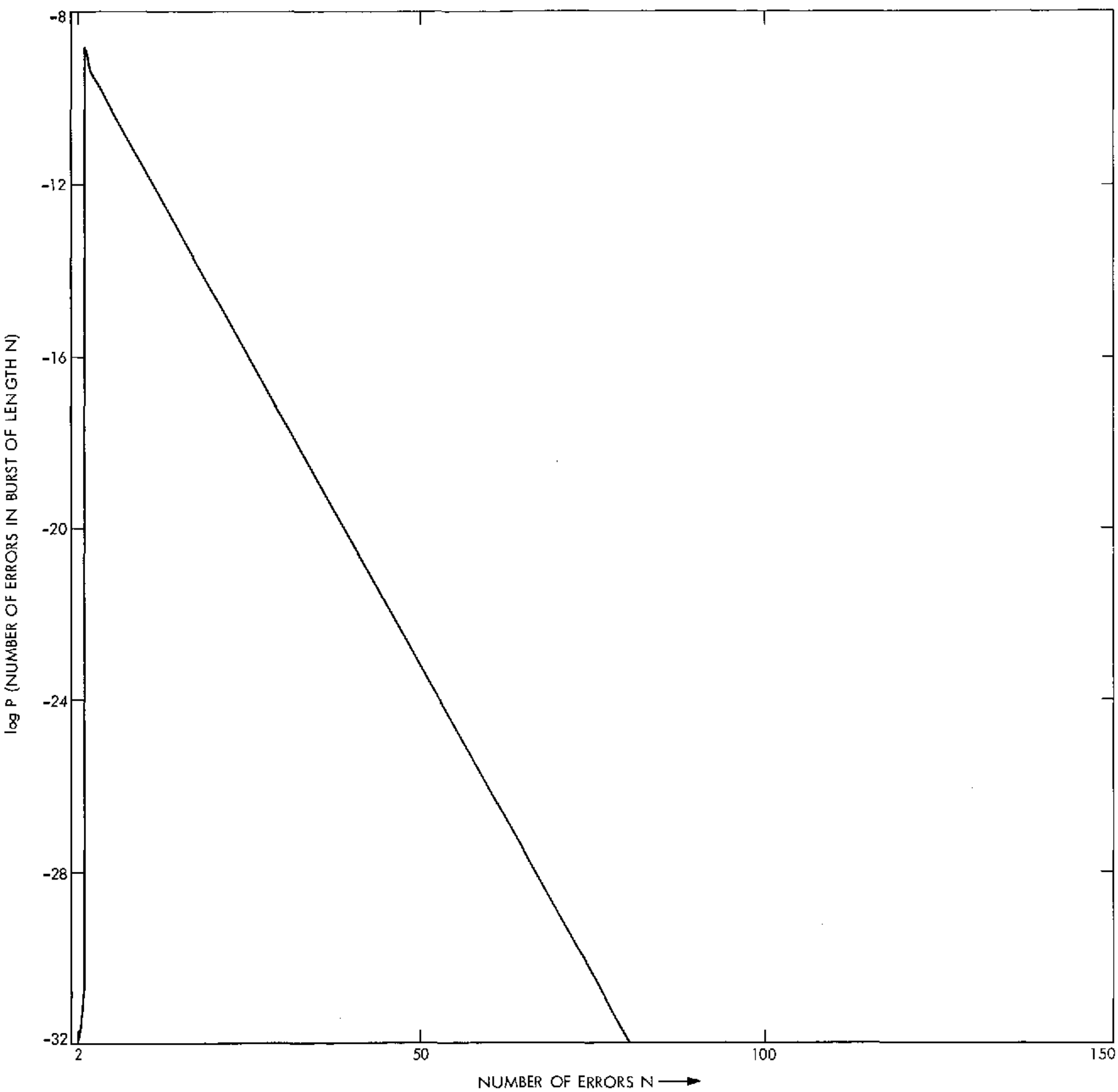

Fig. 32. Distribution of errors in a burst (averaged 4,8 kbps line; burst length = 2400 bits; G = 400)

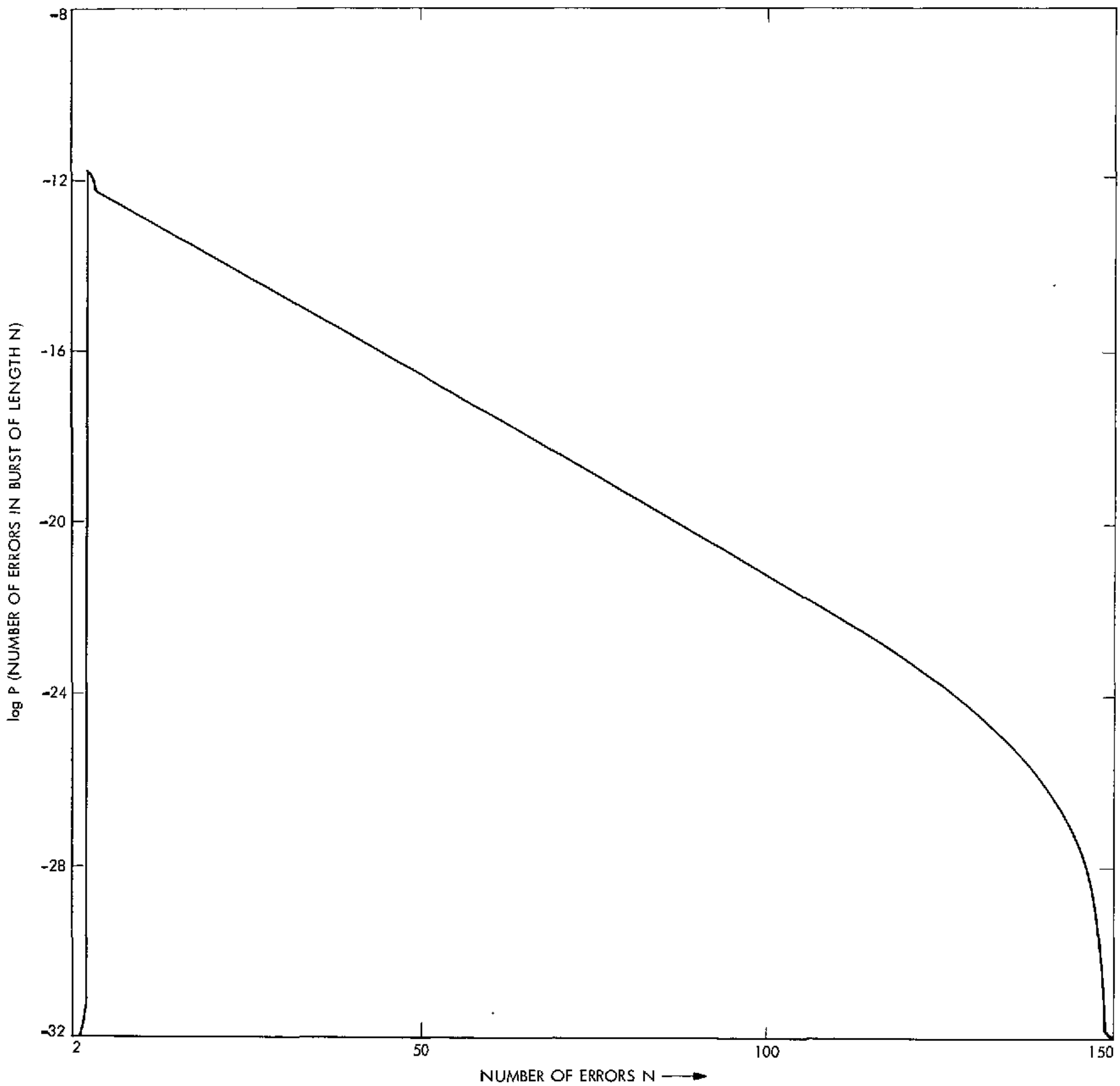

Fig. 33. Distribution of errors in a burst (overall 50 kbps line; burst length = 2400 bits; G = 400)

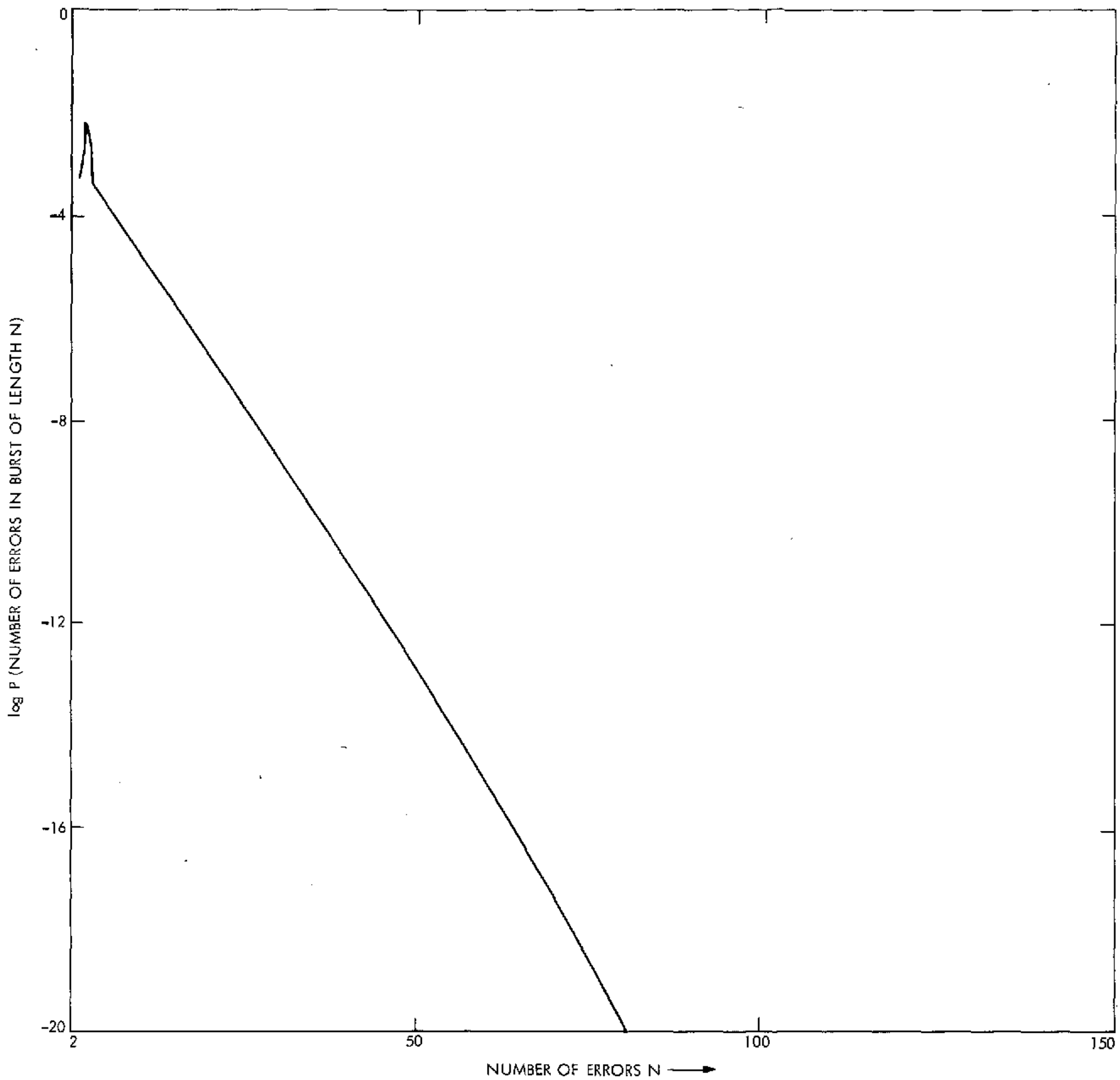

Fig. 34. Distribution of errors in a burst (averaged 4.8 kbps line; burst length = 2400 bits; G = 3600)

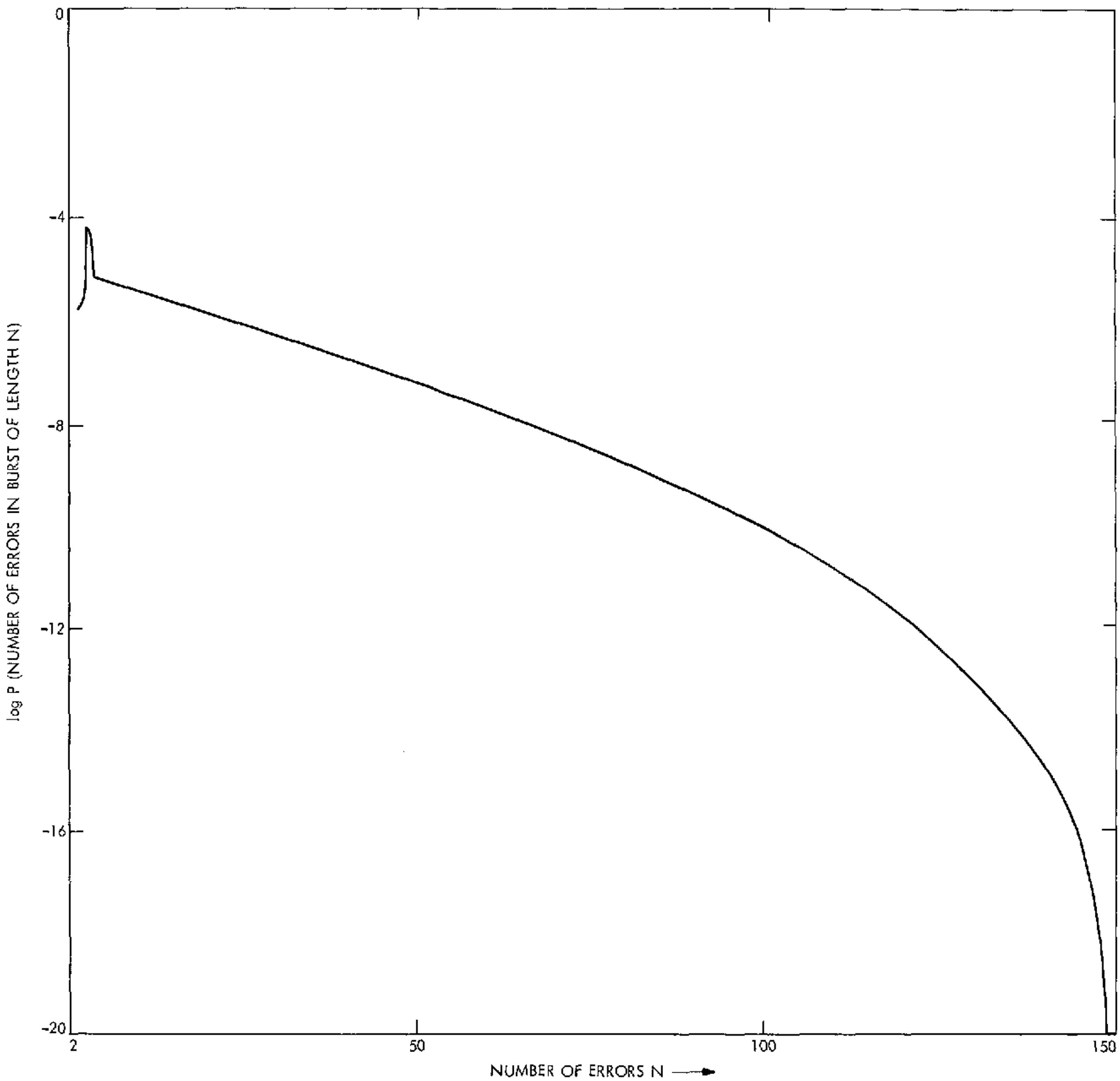

Fig. 35. Distribution of errors in a burst (overall 50 kbps line; burst length = 2400 bits; G = 3600)

Table 29. Mean number of errors, $\overline{K}_n$, in burst of length n = 2400

Group	G = 400			G = 3600		
	$\overline{K}_n$	Density	Bad/Good	$\overline{K}_n$	Density	Bad/Good
Red	239	0.01	0.11	132	0.055	0.058
Amber	254	0.11	0.12	36	0.015	0.015
Green	371	0.155	0.183	22	0.009	0.009
Av. 4800 bps	211	0.088	0.096	68	0.028	0.029
Av. 50 kbps	302	0.126	0.144	199	0.083	0.09

iii. Block Burst

Lastly we ask for the distribution of block bursts. By block-burst we mean a string of blocks starting and ending with an error block, seperated from the nearest preceding and following error block by a gap (of error-free blocks) of not less than the block guardspace and containing within it no gap equal to or greater than the block guardspace.

If d = block guardspace, $L^s(n)$ = probability of a block-burst of length n (each block is s bits long), then as shown in the Appendix

$$L^s_{(n)} = \begin{cases} 0 & \text{for} \quad n \le 0 \\ U^s(d)\overline{L}^s(n); & n \ge 1 \end{cases} \tag{51}$$

where $U^s(d)$ and $\overline{L}^s(n)$ are as given in (152) and (153).

Taking S = 1200 bits and using a block guardspace d = 10 blocks we find that there is 5% chance of getting a block-burst extending to 10 blocks or more in the Red error mode. This probability reduces to less than 1% in the low (Green) error mode. Graphs of this distribution for the different error modes are shown in Figures 36, 37 and 38 .

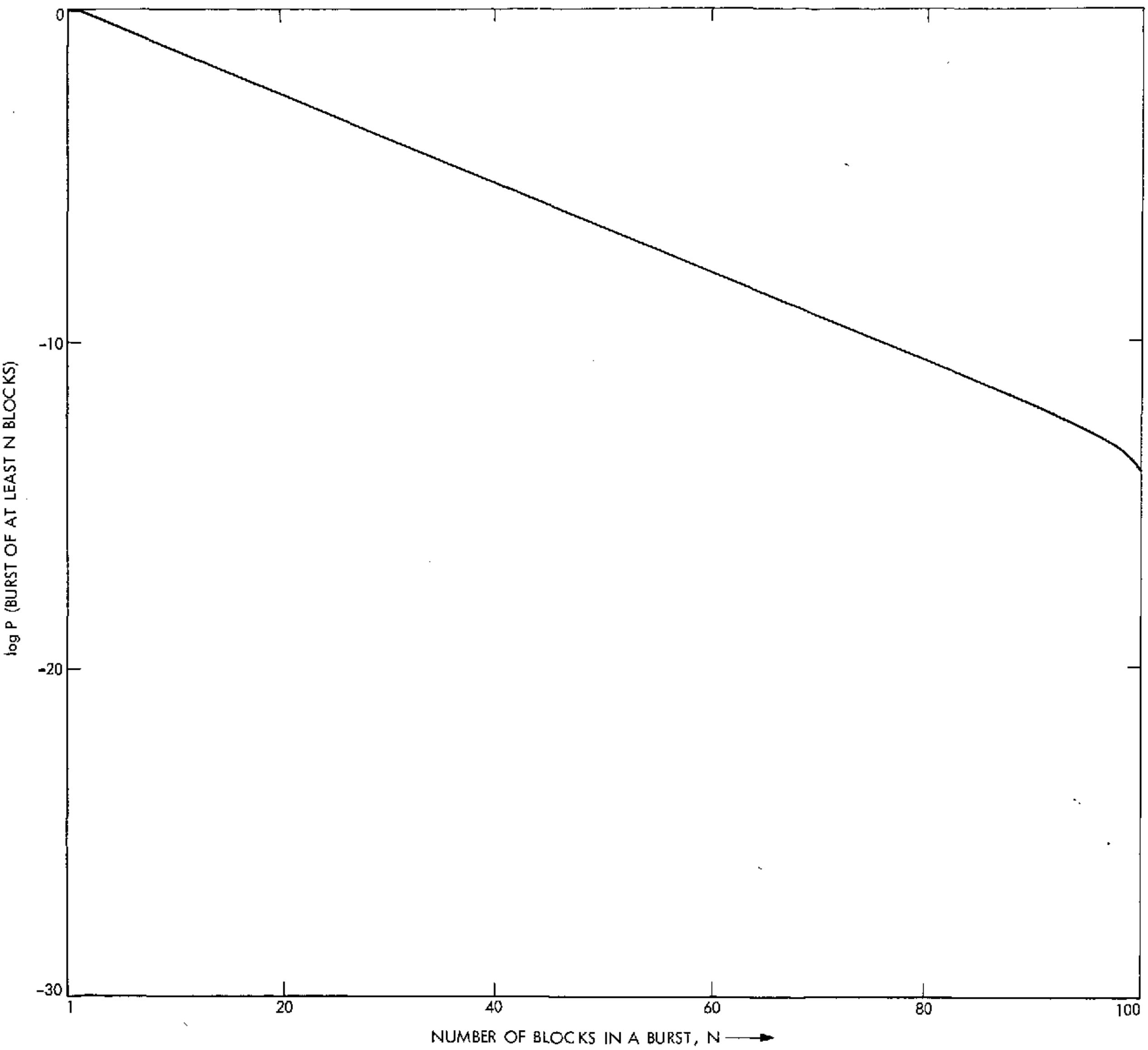

Fig. 36. Block burst (4.8 kbps line; Red group; guardspace = 10 blocks)

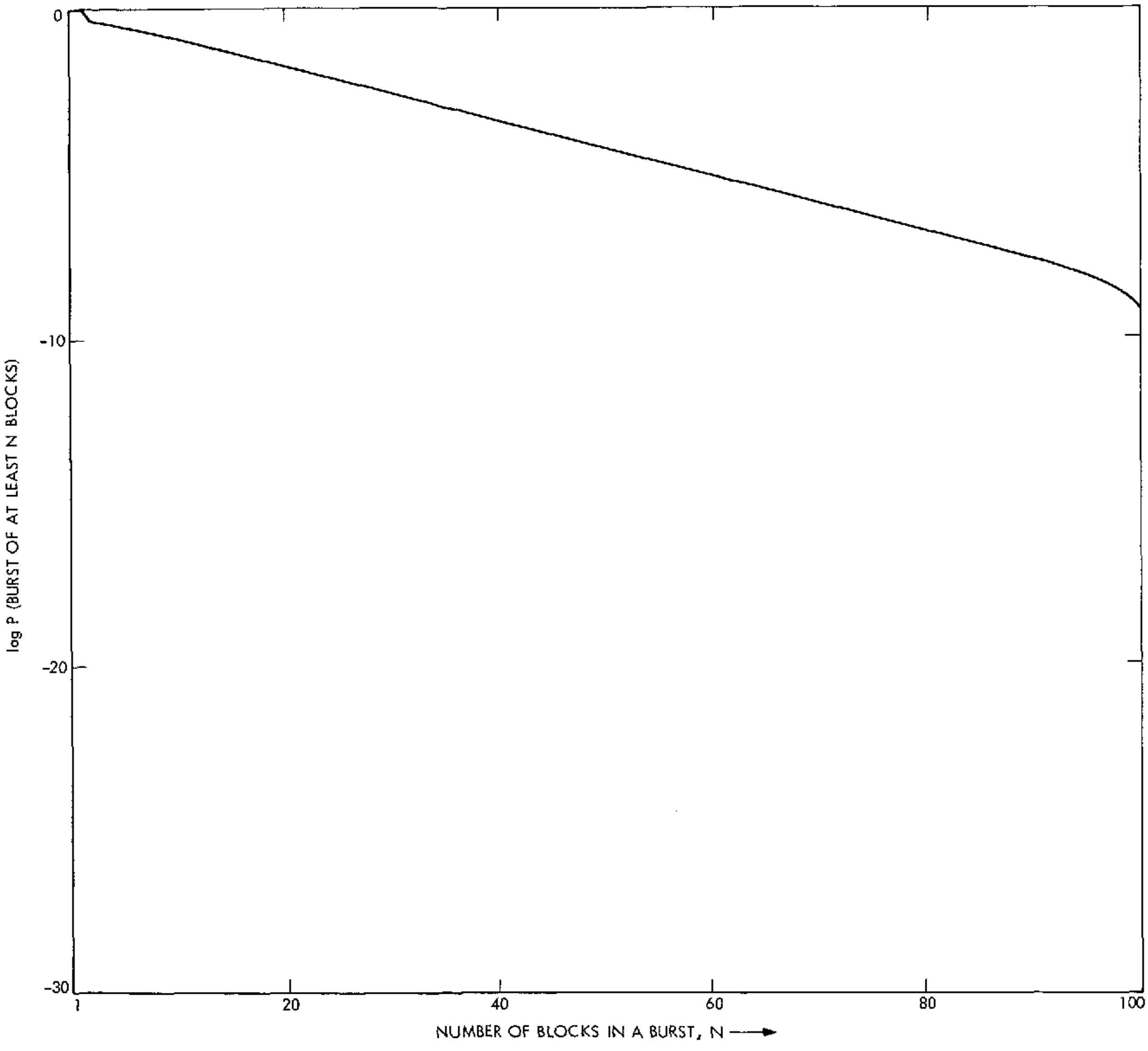

Fig. 37. Block burst (4.8 kbps line; Amber group; guardspace = 10 blocks)

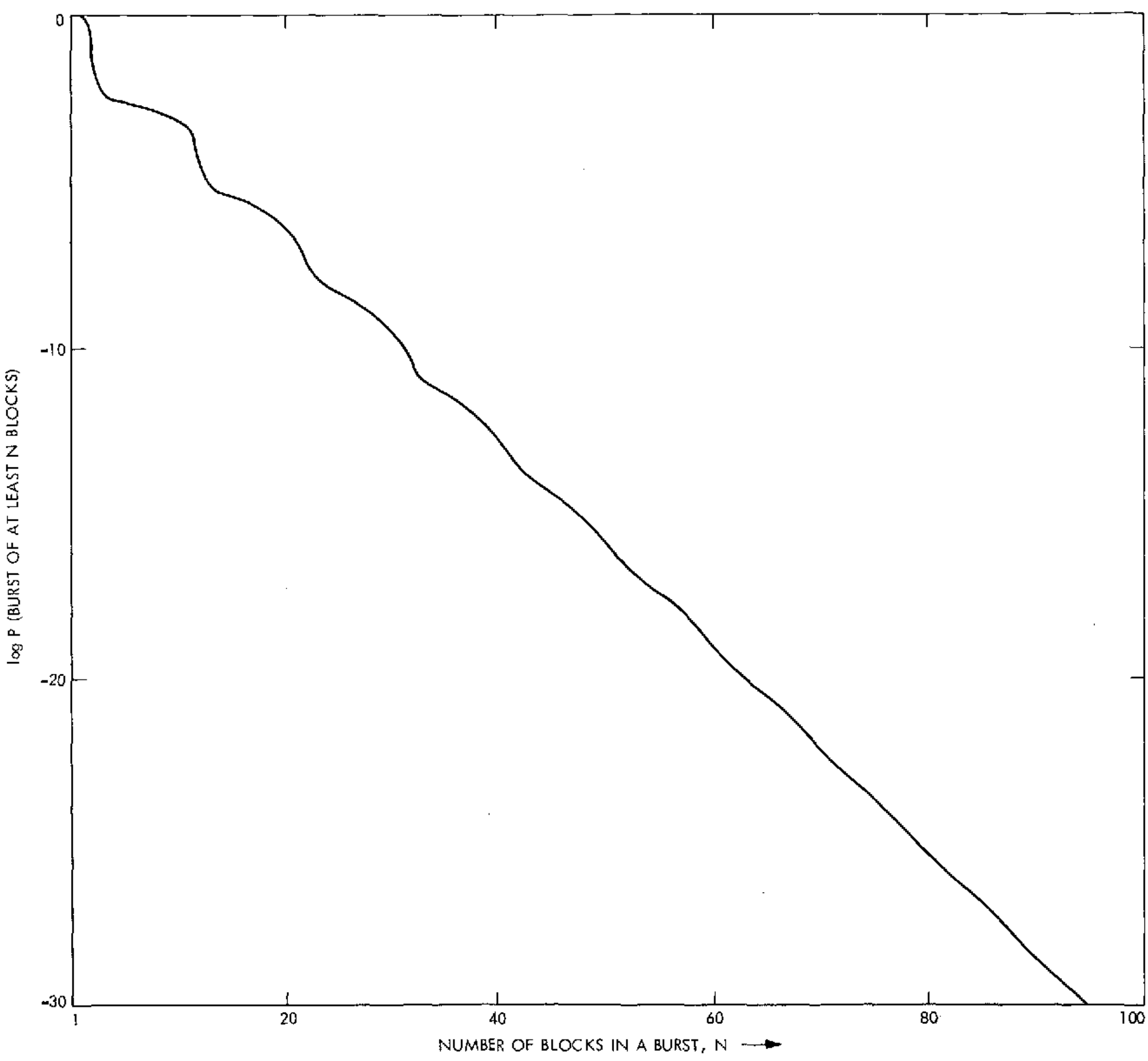

Fig. 38. Block burst (4.8 kbps line; Green group; guardspace = 10 blocks)

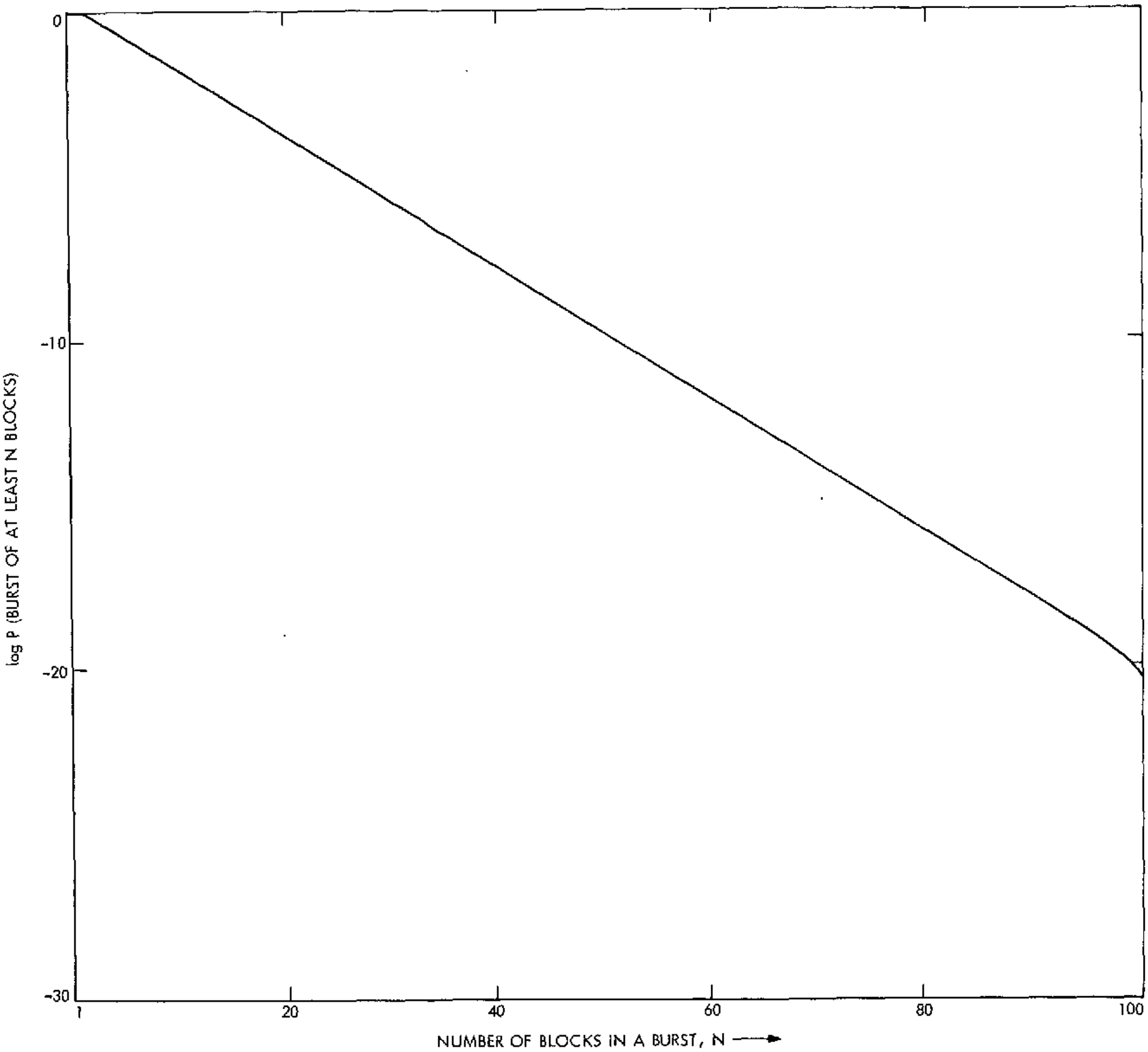

Fig. 39. Block burst (50 kbps line; guardspace = 10 blocks)

Section VIII

CONCLUSION AND REMAINING PROBLEMS

Although we have demonstrated that the simplified five-state Markov chain (Figure 3) can be used successfully to model the errors occurring on the GCF we still want to insist that a more realistic model should incorporate the varying line conditions caused by users coming onto or dropping off the line at different times. As indicated in Section I this can be done by placing appropriate probability distributions both on the varying number of users on the channel which causes the changes in line conditions and the times between these changes.

In our model we have taken note of these line changes by dividing the wide variations in bit error-rate into three main error groups: the Red, Amber and Green modes. All statistics calculated for each group are valid for such a group, as it has been shown in this study. That is why the emphasis has not been on the averaged channel parameters for either the high-speed or the wide-band circuit since they cannot be expected to depict all the three error groups with any measure of reliability. Our success in modelling rather accurately not only the individual test runs but also groups of runs gives us a lot of confidence in all the predictions based on those model parameters.

Of course the problem still remains as to which of the Red, Amber or Green group of statistics to employ at any given time. Our answer? You guessed right! Be conservative, take a pessimistic view of the channel by using the Red-group model statistics in designing all error correcting (and detecting) schemes not only because of the protection it provides but also because this is the error group that our model depicts most accurately. In our model, the higher the error-rate the closer the agreement between the model and empirical statistics. This is as it should be since we are interested only in the bursty mode of the channel.

Statistics, other than those displayed in these pages, exist for estimating the performance of all error correcting (and detecting) codes and for constructing feedback (and retransmission) schemes. Indeed work using some of these statistics has already begun.

We still want to think that the highlight of this study is not the construction of the five-state Markov chain as model for the GCF. Rather it is the development of the Maximum Likelihood procedure for estimating the model parameters using an iterative method. This procedure, as has been shown, is applicable to any finite Markov chain.

It is hoped that the reader has not been left with the impression that all the problems attendant to the model have been solved! Aside from the problem of constructing a model with the fluctuating line conditions built into it along the line sketched above, there are other problems of both theoretical and practical interest that are not totally solved yet. We list a few below.

1. One difficulty we have with our model is finding a physical justification for allowing only one error-state B in which errors occur with probability one. It is fairly well-known that the error-causing mechanism does not reverse the bit each time the channel enters into a burst. Our explanation was that a bursty state is represented by transitions between states B and G_4, sojourn in either of them being allowed. A direct way to model what happens in the physical channel would be to allow two or more states in which errors can occur. Let errors occur in one of the

states with probability 1/2 (given that the state is reached) to account for those times when, in the midst of otherwise good transmissions, random errors occur. The other error-state is the burst state in which errors occur with some probability $0 < h < 1$ to be considered as one of the channel parameters to be estimated.

The problem is that the expression (84) for the capacity of the channel is valid only for the case of one error state. Although (74) is true, in general, the capacity cannot be evaluated directly from it. The problem of finding analytic expression for the capacity when there is more than one error state is still open.

Our attempt has been directed at finding bounds on the entropy H in (74) by finding upper and lower bounds on the function $h(z_1,\ldots,z_n)$ in (80) and showing that these bounds are close enough for large n.

2. Blackwell and Koopmans [8] showed that for a 4×4 irreducible aperiodic Markov matrix M satisfying some mild conditions, the function P(M,z) in (10) of the error sequence $z = \{z_n\}$ can be fixed (the same) for different matrices M. In the terminology of our model this means that if we had used a four-state Markov model satisfying the specified conditions it is possible to have different sets of Maximum Likelihood estimates of the model parameters giving the same joint probability distribution, P(M,z), for the same error sequence z. Then our interest in knowing all such transition matrices becomes apparent. For two seemingly different sets of estimates may indeed be equivalent or two different test runs may be recognized as two samples from the same underlying distribution.

Blackwell and Koopmans found a finite set of functions $f_1, \cdots, f_k$ each defined on the set of all irreducible aperiodic Markov matrices M such that M_1 and M_2 are equivalent (in the sense that $P(M_1, z) = P(M_2, z)$) if and only if $f_i(M_1) = f_i(M_2)$. Such functions f they call A COMPLETE SET OF INVARIANTS. For the 4x4 matrix there are only eight quite-easily-checked such functions (probabilities).

Although the conditions for their result do not all apply to our case there is enough structure in our model to enable us to find the complete set of invariants. We shall return to this problem in a seperate paper.

3. The buffer problem in feedback retransmission.

To people familiar with digital communication this is not a new problem. It has been considered in different forms by different people; indeed a number of schemes are now being developed to reduce the buffer sizes both at the transmitter and the receiver when feedback retransmission method is employed for error correction on the CCF. We describe one such scheme here.

Imagine we have, along with our channel, a reverse (feedback) link, from the receiver to the transmitter, of low capacity available for our use. Let data be transmitted along the channel at constant rate R, data being supplied to the transmitter at rate $R_T < R$. The receiver delivers the received data blocks in sequence; each time a block is received with an error in it the receiver sends to the transmitter through the feedback link a request for retransmission. A copy of the error block and all subsequent blocks received are stored in a buffer at the receiver until the error block is retransmitted and received correctly. This is how the receive buffer fills up. Let T be the loop time, i.e., the time for a transmitted block to reach the receiver and a retransmission request relayed to the transmitter if the block is received with an error in it. It therefore follows that all the blocks entering the transmitter, at constant rate R_T, after each block is transmitted will have to be retained at the transmitter buffer for at least time T by which time the transmitter will know whether or not the block has passed through error-free. We assume the requested retransmission is done as soon as possible after the request is received.

During long block-bursts when successive blocks are hit with errors many blocks of the incoming data may have to be stored at the transmitter buffer. On the otherhand at low block error-rate when only a few blocks are received in error or when a single block is received in error after repeated retransmission, the transmitter buffer does not need to store more than just a few blocks while data is piling up high at the receive buffer.

The problem is to find the distribution of the number of blocks that will be stored in the buffers at the transmitter and the receiver, using our model statistics for the channel parameters. Or equivalently, we may fix a buffer length at N blocks, say, and ask the probability that the number of blocks that have to be stored will exceed N.

While transmission is going on, good data are being delivered to the user in sequence as they arrive at the receiver. If this cannot be done because of requested retransmissions of blocks received in error, the user can wait for a maximum of 8 seconds. After this time all the data in the receive buffer, both good and bad blocks, are delivered to the user. What is the block error-rate of the data eventually delivered to the user?

References

1. McClure, J. P., "4800 bps High Speed Data Error Statistics", JPL Inter-office Memo., Jan. 1973.

2. Adeyemi, O., "An Information-Theoretic Model for the Ground Communication Facility Line", DSN Technical Report 32-1526, Nov-Dec 1972.

3. Berger, J. M. and Mandelbrot, B., "A New Model for Error Clustering in Telephone Circuits", IBM J. Res. and Dev., Vol. 7, pp. 224-236, July 1963.

4. Gilbert, E. N., "Capacity of a Burst-Noise Channel", Bell Syst. Tech. J., Vol. 39, pp. 1253-1265, Sept. 1960.

5. Berkovits, S., Cohen, E. L. and Zierler, N., "A Model for Digital Error Distribution", Tech. Memo, Mitre Corp., Bedford, Mass.

6. Sussman, S. M., "Communication Channel Attributes as Related to Error Control", IEEE Annual Communication Convention, 1965, pp. 5-13.

7. Baum, L. E. and Welch, L. R., "A Statistical Estimation Procedure for Probabilistic Functions of Finite Markov Processes" IDA-CRD Log No. 8664.

8. Blackwell, D. and Koopmans, L., "On the Identifiability Problem for Functions of Finite Markov Chains", Ann. Math. Stat., Vol. 28, pp. 1011-1015.

9. Kunisawa, K., Makabe,H. and Morimura, H., "Notes on the Confidence Bands of Population Distributions", Reports of Stat. Appl. Res. JUSE, Vol. 4 (1955) pp. 18-20.

10. Kunisawa, K., Makabe, H. and Morimura, H., "Tables of Confidence Bands for the Population Distribution Function", Reports of Stat. Appl. Res. JUSE, Vol. 1 (1951), pp. 23-44.

11. Darling D. A., "The Kolmogorov-Smirnov, Cramer-Von Mises Tests", Ann. Math. Stat., Vol. 28, pp. 823-838.

12. Elliot, E. O., "Estimates of Error Rates for Codes on Burst-Noise Channels", Bell Sys. Tech. J., 42 (1963) pp. 1977-1997.

13. McEliece, R. J., "The Problem of Synchronization of Noisy Video", DSN Technical Report 32-1526.

14. Adeyemi, O., "Sync Acquisition and Maintenance on the GCF", Interoffice Memo #331-74-40A.

15. Stern, D. C., "Statistical Analysis of Errors Occurring in the Transmission of High-Speed Digital Data", Master's Thesis, Dept. of Engineering, University of Maryland, 1970.

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

17. Feller, W., An Introduction to Probability Theory and its Application, Vol. 1, 2nd Ed., John Wiley, New York (1959).

18. Rumsey, H., Oral communications.

19. McClure, J. P., "High Speed Data Outage and Block Burst Distribution", Interoffice Memo #3380-73-175.

Appendix I

In this appendix details of the maximum likelihood estimation procedure used to obtain the optimizing parameter set from the data will be outlined.

The steady state probabilities of the finite state Markov chain (M·C) with transition matrix M is given by u_i, $i = 1, 5$ where

$$M = \begin{pmatrix} p_1 & 0 & 0 & 0 & 1-p_1 \\ 0 & p_2 & 0 & 0 & 1-p_2 \\ 0 & 0 & p_3 & 0 & 1-p_3 \\ 0 & 0 & 0 & p_4 & 1-p_4 \\ c_1 & c_2 & c_3 & c_4 & q \end{pmatrix}$$

and

$$u_i = \frac{c_i u_5}{1-p_i} ; \qquad i = 1, 4$$

$$u_5 = \left(1 + \sum_{i=1}^{4} \frac{c_i}{1-p_i}\right)^{-1} \tag{52}$$

Since errors occur only in state B, the bit error probability is given by

$$P_1 = u_5 .$$

The fact that

$$U(k) = \sum_{i=1}^{4} c_i p_i^{k-1} \qquad k \geq 0$$

implies

$$\sum_{i=1}^{4} \frac{c_i}{p_i} = U(0) = 1 \ .$$

Now let us show how to get a MLE of p_1, say, from the data. Suppose, as we did in (7), we assume that for the process to be in state G_1 the length of the gap is at least k_o bits (G_1 is the best error-free state). Then since

$$U(k_o) = P(0^{k_o}|1)$$

$$= \sum_{k \geq k_o} V(k) \ ; \qquad V(k) = P(0^k 1|1)$$

the conditional probability of getting a gap of length $k \geq k_o$ is given by

$$\frac{V(k)}{U(k_o)} = cp_1^{k-1}$$

where

$$c = \frac{c_1(1-p_1)}{U(k_o)} \ ; \qquad U(k_o) = c_1 p_1^{k_o - 1} \ .$$

Suppose there are ℓ gaps of lengths $k_1, k_2, \ldots, k_\ell$ such that $k_i \geq k_o$, $i = 1, \ldots, \ell$. Then the joint conditional probability of getting the ℓ gaps is given by:

$$P(k_1, k_2, \ldots, k_\ell | k_o) = c^\ell p_1^{\sum_1^\ell k_i - \ell} .$$

It is desired to maximize this probability. We see that

$$\frac{\partial}{\partial p_1} P(k_1, k_2, \ldots, k_\ell | k_o) = \frac{\partial}{\partial p_1}\left[\left(\frac{1-p_1}{p_1^{k_o}}\right)^\ell p_1^{\sum_1^\ell k_i}\right]$$

$$= p_1^{\sum_1^\ell k_i} \ell \left(\frac{1-p_1}{p_1^{k_o}}\right)^{\ell-1}\left[\frac{-p_1^{k_o} - (1-p_1)k_o p_1^{k_o - 1}}{p_1^{2k_o}}\right]$$

$$+ \left(\frac{1-p_1}{p_1^{k_o}}\right)^\ell \left(\sum_1^\ell k_i\right) p_1^{\sum_1^\ell k_i - 1} = 0$$

if

$$\ell\left[- 1 - \left(\frac{1-p_1}{p_1}\right)k_o\right] + \left(\frac{1-p_1}{p_1}\right)\left(\sum_1^\ell k_i\right) = 0$$

which implies

$$p_1 = \frac{\sum_1^\ell k_i - \ell k_o}{\sum_1^\ell k_i - \ell k_o + \ell} . \qquad (53)$$

So in general

$$\hat{p}_i = \frac{\sum_{j=1}^{\ell_i} k_{ji} - \ell_i \bar{k}_i}{\sum_{j=1}^{\ell_i} k_{ji} - \ell_i \bar{k}_i + \ell_i} \; ; \qquad i = 1, 4 \; . \tag{54}$$

And because ℓ_i = number of times process enters state i,

$$\hat{c}_i = \frac{\ell_i}{N_e}$$

$$\hat{q} = \frac{N_{11}}{N_e}$$

where k_{ji}, $\bar{k}_i$, N_ℓ, and N_{11} are, as defined in Section on Parameter estimation, given by:

k_{ji} = length of gap j, j = 1, ℓ_i in state i, i = 1, 4

$\bar{k}_i$ = the threshold to state i

N_e = number of errors in the run

and N_{11} = number of gaps of length zero.

Suppose we are given a sample of size N of observations from a finite state M.C in the form of an error sequence $Z = \{z_n\}_{n=1,N}$. Our aim is to determine the transition matrix M from the sequence Z. The general form of the following procedure was proposed by Baum and Welch in [7] but in that paper no mention was made of a way to get the initial estimates with which to start their iterative method. Our iterative method for getting the matrix M will use the $\hat{p}_i$ and $\hat{c}_i$ obtained above as initial estimates.

The probability of getting the sequence $z_1, z_2, \ldots, z_n$ (using the structure in the 31 samples) if M is the transition matrix can be written as:

$$P(M, Z) = P_1 U(\ell) U(L) \prod_{j=1}^{N_e-1} V(\ell_j) \tag{55}$$

subject to $U(0) = 1$.

(Our notation here is as used in the section on Parameter Estimation.)

Let M^o denote the true value of M. We would like to take as our estimate $\hat{M}$ of M^o a value of M which maximizes $P(M, Z)$. It is possible for distinct $\hat{M}$ to yield this maximum (see Blackwell and Koopmans [8]). But we shall content ourself with getting any one of such $\hat{M}$.

Now, the M maximizing $P(M, Z)$ also maximizes

$$\log \frac{P(M,Z)}{P_1} = \log \sum_{i=1}^{4} c_i p_i^{\ell-1} + \log \sum_i c_i p_i^{L-1} + \sum_{j=1}^{N_e-1} \log \sum_i c_i (1-p_i) p_i^{\ell_j-1} \tag{56}$$

subject to

$$U(0) = \sum \frac{c_i}{p_i} = 1 .$$

Put

$$\bar{c}_i = \frac{c_i(1-p_i)}{p_i}$$

and let

$$\alpha_1 = \sum_i \frac{\bar{c}_i p_i^{\ell}}{1-p_i} \; ; \qquad \alpha_L = \sum_i \frac{\bar{c}_i p_i^{L}}{1-p_i} \; ; \qquad \alpha_j = \sum_i \bar{c}_i p_i^{\ell_j} \; . \tag{57}$$

Then (56) can be written as:

$$\log \frac{P(M,Z)}{p_1} = \log\alpha_1 + \log\alpha_L + \sum_{j=1}^{N_e-1} \log\alpha_j \tag{58}$$

subject to the condition $\underset{\sim}{G}$; $\sum \frac{\bar{c}_i}{1-p_i} = 1$.

Using method of Lagrange's multipliers we have

$$\nabla \log \frac{P(M,Z)}{p_1} = \lambda \nabla \underset{\sim}{G}$$

where ∇ is the differential operator. That is

$$\frac{\partial}{\partial \bar{c}_i} \log \frac{P(M,Z)}{p_1} = \frac{\frac{p_i^{\ell}}{1-p_i}}{\alpha_1} + \frac{\frac{p_i^{L}}{1-p_i}}{\alpha_L} + \sum_j^{N_e-1} \frac{p_i^{\ell_j}}{\alpha_j} = \frac{\lambda}{1-p_i} \tag{59}$$

and

$$\frac{\partial}{\partial p_i} \log \frac{P(M,Z)}{p_1} = \frac{\ell \bar{c}_i (1-p_i) p_i^{\ell-1} + \bar{c}_i p_i^{\ell}}{(1-p_i)^2 \alpha_1} + \frac{L\bar{c}_i (1-p_i) p_i^{L-1} + \bar{c}_i p_i^{L}}{(1-p_i)^2 \alpha_L} + \sum_j \ell_j \frac{\bar{c}_i p_i^{\ell_j - 1}}{\alpha_j}$$

$$= \lambda \frac{\bar{c}_i}{(1-p_i)^2} \qquad (60)$$

(59) x $\sum \bar{c}_i$ gives

$$\lambda = N_e + 1 \qquad (61)$$

Dividing (59) x $\bar{c}_i$ by (60) x p_i we get

$$\frac{\dfrac{\bar{c}_i p_i^{\ell}}{(1-p_i)\alpha_1} + \dfrac{\bar{c}_i p_i^{L}}{(1-p_i)\alpha_L} + \sum_j \dfrac{\bar{c}_i p_i^{\ell_j}}{\alpha_j}}{\dfrac{\ell \bar{c}_i p_i^{\ell}}{(1-p_i)\alpha_1} + \dfrac{\bar{c}_i p_i^{\ell+1}}{(1-p_i)^2 \alpha_1} + \dfrac{L\bar{c}_i p_i^{L}}{(1-p_i)\alpha_L} + \dfrac{\bar{c}_i p_i^{L+1}}{(1-p_i)^2 \alpha_L} + \sum_j \ell_j \dfrac{\bar{c}_i p_i^{\ell_j}}{\alpha_j}} = \frac{1-p_i}{p_i} \qquad (62)$$

Denote the numerator and denominator of the LHS of (62) by D_1 and D_2 respectively. Then from (62) we have

$$p_i = \frac{D_2}{D_1 + D_2} \qquad (63)$$

and

$$1 - p_i = \frac{D_1}{D_1+D_2} \quad .$$

Since (59) x $\overline{c}_i$ using (61) is

$$D_1 = \frac{(N_e+1)\overline{c}_i}{1-p_i}$$

we can write for $\overline{c}_i$ using (63):

$$\overline{c}_i = \frac{D_1^2}{(N_e+1)(D_1+D_2)} \quad . \tag{64}$$

The expressions for p_i and $\overline{c}_i$ in (63) and (64) are the ones to use to iterate to get the optimum values of these parameters using as initial values the raw estimates obtained in (54).

To get a single transition matrix M we consider the n runs as independent samples from the channel with transition M. Then the joint probability P(M, Z) of getting the sequence $z_1, z_2, \ldots, z_N$, where $N = \sum_{j\geq 1}^{n} N_j$ is the total sum of bits in all the samples, is given by:

$$P(M, Z) = P_1^n \prod_{k=1}^{n} U(\ell_{1k})U(L_k) \prod_{j=1}^{N_{ek}-1} V(\ell_{jk}) \tag{65}$$

$$U(0) = 1 \quad .$$

The notation here is as used for equation(17). Our aim here also is to maximize $P(M, Z)$. Using the method of Lagrange's multipliers subject to the condition

$$1 = U(0) = \sum \frac{c_i}{p_i}$$

we obtain

$$p_i = \frac{\bar{D}_2}{D}$$

and

$$\bar{c}_i = \frac{\bar{D}_1^2}{\left(\sum_{k=1}^{n} N_{ek} + n\right) D} \tag{66}$$

$$c_i = \frac{\bar{c}_i p_i}{1 - p_i}$$

where N_{ek} = number of errors in sample k

$$\bar{D}_1 = \sum_{k=1}^{n} \left[\frac{\bar{c}_i p_i^{\ell_{1k}}}{(1-p_i)\alpha_{1k}} + \frac{\bar{c}_i p_i^{L_k}}{(1-p_i)\alpha_{L_k}} + \sum_{j=1}^{N_{ek}-1} \frac{\bar{c}_i p_i^{\ell_{jk}}}{\alpha_{jk}} \right]$$

$$\bar{D}_2 = \sum_{k=1}^{n} \left[\frac{\ell_{1k} \bar{c}_i p_i^{\ell_{1k}}}{(1-p_i)\alpha_k} + \frac{\bar{c}_i p_i^{\ell_{1k}+1}}{(1-p_i)^2 \alpha_{1k}} + \frac{L_k \bar{c}_i p_i^{L_k}}{(1-p_i)\alpha_{1k}} + \frac{\bar{c}_i p_i^{L_k+1}}{(1-p_i)^2 \alpha_{L_k}} + \sum_{j=1}^{N_{ek}-1} \ell_{jk} \frac{\bar{c}_i p_i^{\ell_{jk}}}{\alpha_{jk}} \right]$$

and

$$D = \bar{D}_1 + \bar{D}_2 .$$

Appendix II

We shall indicate here the method used to find the n-step transition matrix M^n of the channel M. Here also, as in Reference [2], we shall follow the method in Feller [17].

Let X and Y be the right and left eigen vectors of M with eigen value $\lambda \neq 0$, $s = \frac{1}{\lambda}$. That is

$$sMX = X \qquad (67)i$$

$$sYM = Y \qquad (67)ii$$

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_5 \end{pmatrix}; \qquad Y = (y_1, y_2, \ldots, y_5)$$

Writing (i) and (ii) out we have:

$$\begin{aligned}
x_i &= \frac{s(1-p_i)}{1-sp_i} x_5 \qquad i = 1, 4 \\
x_5 &= s \sum_{i=1}^{4} c_i x_i + sqx_5 ; \qquad q = 1 - \sum_{i=1}^{4} c_i . \\
y_i &= \frac{sc_i}{1-sp_i} y_5 \\
y_5 &= s \sum_{i=1}^{4} (1-p_i) y_i + sqy_5 .
\end{aligned} \qquad (68)$$

Since $x_5(y_5)$ is determined up to arbitrary multiplicative constant, we may put $x_5(y_5) = 1$. (Note that $x_5(y_5) \neq 0$, for that would make $x_i \equiv 0$ for any s; also that (i) and (ii) have the same eigen values. So it suffices to find these solutions using (67)i). Then

$$1 = \sum_{i=1}^{4} c_i s^2 \frac{(1-p_i)}{1-sp_i} + s\left(1 - \sum_{i=1}^{4} c_i\right). \tag{69}$$

Substituting $\lambda = \frac{1}{s}$ we can write (69) as:

$$\lambda - 1 = \sum_{i=1}^{4} c_i \left[\frac{1-\lambda}{\lambda-p_i}\right] \tag{70}$$

which shows that $\lambda = 1$ is a root. Removing this root we have left:

$$\sum_{i=1}^{4} \frac{c_i}{\lambda-p_i} + 1 = 0 \tag{71}$$

or

$$\lambda\left\{\lambda^3 - \lambda^2\left(\sum_{i=1}^{4} p_i - \sum_{i=1}^{4} c_i\right) + \lambda\left[\frac{c_1}{p_1}(p_2p_3+p_2p_4+p_3p_4) + \frac{c_2}{p_2}(p_1p_3+p_1p_4+p_3p_4) + \frac{c_3}{p_3}(p_1p_2+p_1p_4+p_2p_4) + \frac{c_4}{p_4}(p_1p_2+p_1p_3+p_2p_3)\right] - \left(\frac{c_1}{p_1} p_2p_3p_4 + \frac{c_2}{p_2} p_1p_3p_4 + \frac{c_3}{p_3} p_1p_2p_4 + \frac{c_4}{p_4} p_1p_2p_3\right)\right\} = 0 .$$

Hence $\lambda = 0$ is another root; and if we write the cubic equation as

$$\hat{a}\lambda^3 + \hat{b}\lambda^2 + \hat{c}\lambda + \hat{d} = 0$$

standard method for solving cubics gives the three roots x_1, x_2, x_3 as:

$$\begin{aligned}
x_1 &= -\left\{(A - B)^{\frac{1}{3}} + (B + A)^{\frac{1}{3}}\right\} + \frac{\hat{b}}{3} \\
x_2 &= \frac{1}{2}\left\{(A - B)^{\frac{1}{3}} + (B + A)^{\frac{1}{3}}\right\} + \frac{i\sqrt{3}}{2}\left\{(B + A)^{\frac{1}{3}} - (A - B)^{\frac{1}{3}}\right\} - \frac{\hat{b}}{3} \\
x_3 &= \frac{1}{2}\left\{(A - B)^{\frac{1}{3}} + (B + A)^{\frac{1}{3}}\right\} - \frac{i\sqrt{3}}{2}\left\{(B + A)^{\frac{1}{3}} - (A - B)^{\frac{1}{3}}\right\} - \frac{\hat{b}}{3}
\end{aligned} \qquad (72)$$

where

$$A = \frac{\hat{q}}{2}\ ; \qquad B^2 = \frac{\hat{q}^2}{4} + \frac{\hat{p}^3}{27}$$

$$\hat{q} = \hat{d} - \frac{\hat{b}\hat{c}}{3} + \frac{2\hat{b}^3}{27}\ ; \qquad \hat{p} = \hat{c} - \frac{\hat{b}^2}{3}\ ; \qquad \hat{a} = 1$$

$$\hat{b} = -\left[\sum_{i=1}^{4} p_i - \sum_{i=1}^{4} c_i\right]$$

$$\hat{c} = \frac{c_1}{p_1}(p_2p_3+p_2p_4+p_3p_4) + \frac{c_2}{p_2}(p_1p_3+p_1p_4+p_3p_4) + \frac{c_3}{p_3}(p_1p_2+p_1p_4+p_2p_4) + \frac{c_4}{p_4}(p_1p_2+p_1p_3+p_2p_3)$$

$$\hat{d} = -\left(\frac{c_1}{p_1}\, p_2p_3p_4 + \frac{c_2}{p_2}\, p_1p_3p_4 + \frac{c_3}{p_3}\, p_1p_2p_4 + \frac{c_4}{p_4}\, p_1p_2p_3\right) .$$

So the eigen values of (67)i are

$$\lambda = 1,\ 0,\ x_1,\ x_2,\ x_3\ .$$

Now (68) in terms of λ becomes:

$$X_i^{(r)} = \frac{1-p_i}{\lambda_r - p_i}\ ; \qquad Y_i^{(r)} = \frac{c_i}{\lambda_r - p_i}$$

$$X_5 = 1 \qquad\qquad Y_5 = 1$$

$$i = 1,\ 4\ ; \qquad r = 1,\ 5\ .$$

Let $M^n = (p_{jk}^{(n)})$. Then

$$p_{jk}^{(n)} = \sum_{r=1}^{5} t_r X_j^{(r)} Y_k^{(r)} \lambda_r^n \tag{73}$$

where

$$1 = t_r \sum_{v=1}^{5} X_v^{(r)} Y_v^{(r)}$$

or

$$t_r = \left[1 + \sum_{i=1}^{4} \frac{c_i(1-p_i)}{(\lambda_r - p_i)^2} \right]^{-1} .$$

If we write

$$t_o = t_r\Big|_{\lambda_r=1} \quad ;$$

and for $r = 1, 3,\ \ t_r = t_{\lambda_r};\ \ \lambda_r = x_1, x_2, x_3$ then

$$t_o = \left[\sum_{i=1}^{4} \frac{c_i}{p_i(1-p_i)}\right]^{-1} \quad (= P_1)$$

$$t = t_r\Big|_{\lambda_r=0} = \left[\sum_{i=1}^{4} \frac{c_i}{p_i^2}\right]^{-1}$$

and

$$p_{jk}^{(n)} = \frac{c_k}{1-p_k}\, t_o + c_k(1 - p_j) \sum_{r=1}^{3} \frac{t_r\lambda_r^n}{(\lambda_r-p_j)(\lambda_r-p_k)}$$

$$j,k = 1, 4$$

$$p_{j5}^{n} = t_o + (1 - p_j) \sum_r \frac{t_r\lambda_r^n}{(\lambda_r-p_j)}$$

$$p_{5k}^{(n)} = \frac{c_k}{1-p_k}\, t_o + c_k \sum_r \frac{t_r\lambda_r^n}{(\lambda_r-p_k)}$$

$$p_{55}^{(n)} = t_o + \sum_r t_r\lambda_r^n$$

Appendix III

Proposition

The capacity, C, of the burst-noise channel is given by

$$C = 1 - H \tag{74}$$

where

$$H = - \lim_{n\to\infty} \sum_{z_i = 0 \text{ or } 1} P(z_1, \ldots, z_{n+1}) \log P(z_{n+1} | z_1, \ldots, z_n)$$

Proof: The proof is by classical information-theoretic arguments. The mutual information of the n-extension of the channel is:

$$I(X^n, Y^n) = H(Y^n) - H(Y^n | X^n) \tag{75}$$

where $X(Y)$ is the input (output) and $H(\cdot)$ is the entropy function. The transmission rate is then

$$R = \lim_{n\to\infty} \frac{I(X^n, Y^n)}{n} \tag{76}$$

and if $p(x)$ is an input distribution, the capacity C is given by

$$C = \max_{p(x)} R$$

$$= \max_{p(x)} \lim_{n\to\infty} \left\{ \frac{H(Y^n)}{n} - \frac{H(Y^n|X^n)}{n} \right\} \tag{77}$$

Now for additive noise, i.e. $Y = X + Z$, Z the noise sequence, we can show easily that

$$H(Y^n|X^n) = H(z_1, \ldots, z_n) \tag{78}$$

independent of X^n and that for $p(x_1, \ldots, x_n) = 2^{-n}$, $\frac{H(Y^n)}{n}$ achieves its maximum equal to 1. But

$$\lim_{n\to\infty} \frac{H(z_1,\ldots,z_n)}{n} = \lim_{n\to\infty} H(z_n|z_1, \ldots, z_{n-1}) \, .$$

So by (77), (78) we have

$$C = 1 - \lim_{n\to\infty} H(z_n|z_1, \ldots, z_{n-1}) \tag{79}$$

$$= 1 - H$$

where H denotes $\lim_{n\to\infty} H(z_n|z_1, \ldots, z_{n-1})$ and

$$H(z_n|z_1,\ldots,z_{n-1}) = - \sum_{z_i=0 \text{ or } 1} P(z_1,\ldots,z_n) \log P(z_n|z_1,\ldots,z_{n-1}) \, .$$

Now let us write H as

$$H = \lim_{n\to\infty} \sum_{\{z_1,\ldots,z_n\}} P(z_1,\ldots,z_n)h(z_1,\ldots,z_n)$$

with

$$h(z_1,\ldots,z_n) = - \sum_{z_{n+1}=0 \text{ or } 1} P(z_{n+1}|z_1,\ldots,z_n)\log P(z_{n+1}|z_1,\ldots,z_n). \tag{80}$$

If we assume that our model has only one error state B then we can show, see Gilbert [4], that $h(z_1, \ldots, z_n)$ can assume only $(n+1)$ values

$$h(0^n),\ h(10^{n-1}),\ h(10^{n-2}),\ \ldots,\ h(10),\ h(1) \tag{81}$$

where $\{10^k,\ k = 0, \ldots, n\}$ is the event that an error is followed by k error-free bits. Using (80) and (81) we can write

$$H = \sum_{k=0}^{\infty} P(10^k)h(10^k). \tag{82}$$

In terms of $U(k)$ and $V(k)$, $P(10^k)$ is given by

$$P(10^k) = P_1U(k)$$

and hence

$$P(0|10^k) = \frac{U(k+1)}{U(k)} \tag{83}$$

so

$$h(10^k) = -\frac{U(k+1)}{U(k)} \log \frac{U(k+1)}{U(k)} - \left(1 - \frac{U(k+1)}{U(k)}\right) \log\left(1 - \frac{U(k+1)}{U(k)}\right)$$

and

$$H = -P_1 \sum_{k=0}^{\infty} U(k)\left\{\frac{U(k+1)}{U(k)} \log \frac{U(k+1)}{U(k)} + \left(1 - \frac{U(k+1)}{U(k)}\right) \log\left(1 - \frac{U(k+1)}{U(k)}\right)\right\} \tag{84}$$

Note that H can also be written in terms of $V(k)$ viz:

$$V(k) = U(k) - U(k+1)$$

$$H = -P_1 \sum_{k=0}^{\infty} V(k)\log V(k) \tag{85}$$

although we shall not use this form in our calculations.

Block-Bit Statistics

Expressions for the following statistics will be given in this section:

(i) Proportion of blocks in error.

(ii) $P(k,n)$: probability of k errors in a block of length n.

(iii) $P_t(k,n)$: probability of k errors in n information digits on our channel sampled at every t^{th} step.

(iv) $P(k$ bits between extreme errors in n-block$|\geq 2)$.

We shall get (i) as a special solution to (ii).

(ii)
$$P(k,\ n) = P\left\{0^{\ell_1}10^{\ell_2}1 \cdots 0^{\ell_{L-1}}10^{\ell_L}\right\}; \qquad 0 \leq \ell_1 \leq n - k$$
$$= \sum_{\ell_1=0}^{n-k} P(0^{\ell_1}1)P\left\{0^{\ell_2}10^{\ell_3}1 \cdots 0^{\ell_{L-1}}10^{\ell_L}|1\right\}; \qquad \sum_j \ell_j = n-k \tag{86}$$

Now

$$P\Big\{\underbrace{0^{\ell_2}10^{\ell_3}1 \cdots 0^{\ell_{L-1}}10^{\ell_L}}_{n-\ell_1-1}|1\Big\} = \sum_{\ell_2=0}^{n-\ell_1-k} P\left(0^{\ell_2}1|1\right)P\Big\{\underbrace{0^{\ell_3}1 \cdots 0^{\ell_{L-1}}10^{\ell_L}}_{N-\ell_2-1}|1\Big\} \tag{87}$$

where $N = n-\ell_1-1$. Denoting the LHS of (87) by $\overline{P}(k-1,\ n-\ell_1-1)$, we have:

$$\overline{P}(k-1,\ N) = \sum_{\ell_2=0}^{n-\ell_1-k} V(\ell_2)\overline{P}(k-2,\ N-\ell_2-1) \tag{88}$$

So from (86):

$$P(k, n) = P_1 \sum_{\ell_1=0}^{n-k} U(\ell_1)\overline{P}(k-1, n-\ell_1-1) \tag{89}$$

$$= P_1 \sum_{\ell_1=0}^{n-k} \sum_{\ell_2=0}^{n-\ell_1-k} U(\ell_1)V(\ell_2)\overline{P}(k-2, n-\ell_1-\ell_2-2)$$

where

$$P_1 = \left[1 + \sum_{i=1}^{4} \frac{c_i}{1-p_i}\right]^{-1} .$$

Method of calculating $P(k,n)$ from the recursion (89):

Note that (89) and (88) imply that it suffices to know, for every $k \geq 2$,

$$\{\overline{P}(k-2, n-j), \quad j = 2, \ldots, n-k+2; \quad k \leq n\}. \tag{90}$$

That is, it is enough to know

$$\{\overline{P}(0, n-j); \quad j = 2, \ldots, n\} . \tag{91}$$

But

$$\overline{P}(0, n) = U(n)$$

$$= \sum_{i=1}^{4} c_i p_i^{n-1} . \tag{92}$$

To find $P(k, n)$ for $k = 0, 1$ observe that

$$P(0, n) = \sum_{j=1}^{4} u_j p_j^{n-1} ; \qquad u_j = \frac{c_j}{1-p_j} u_5$$

$$= P_1 \sum_{j=1}^{4} \frac{c_j}{1-p_j} p_j^{n-1} ; \qquad u_5 = P_1 . \tag{93}$$

Hence P(error block) = 1 - P(0, n). For k = 1 (86) becomes:

$$P(1, n) = P_1 \sum_{\ell_1=0}^{n-1} U(\ell_1)\overline{P}(0, n - \ell_1 - 1)$$

and using (92) we have

$$P(1, n) = P_1 \sum_{\ell_1=0}^{n-1} U(\ell_1)U(n - \ell_1 - 1) \tag{94}$$

(iii) $\underline{P_t(k, n)}$

Given the matrix M^t – the t-step transition matrix, the problem reduces to (ii) with M replaced by M^t. Let

$$U_t(k) = P_t(0^k|1)$$
$$V_t(k) = P_t(0^k 1|1) \tag{95}$$

be the gap statistics w.r.t. M^t. We want to find expressions in terms of M^t for $U_t(k)$ and $V_t(k)$. Write

$$G_{t,i}(k) = P_t(0^k, s_k = G_i | 1) ; \qquad i = 1, 4 .$$

Then

$$U_t(k) = \sum_{i=1}^{4} G_{t,i}(k) \tag{96}$$

Let us write

$$M^t = (P_{ij})_{5x5}$$

and $\overline{M}$ as the 4 x 4 matrix obtained by deleting the last row and column of M^t. Thus $\overline{M}$ are the t-step transitions between the good states only. Then

$$(G_{t,1}(k), G_{t,2}(k), G_{t,3}(k), G_{t,4}(k)) = (p_{51}, p_{52}, p_{53}, p_{54})\overline{M}^{k-1} \tag{97}$$

It is appropriate to note here that (97) is true in general whenever we are interested only in a sequence of error-free transmissions starting with an error. The only part of the original transition matrix to use is that denoting transitions only between the good states. Thus in calculating $U(k)$ and $V(k)$ w.r.t. the basic transition matrix M, we used the transitions between the good states

$$\overline{P} = \begin{pmatrix} P_1 & 0 & 0 & 0 \\ 0 & P_2 & 0 & 0 \\ 0 & 0 & P_3 & 0 \\ 0 & 0 & 0 & P_4 \end{pmatrix}$$

giving

$$\bar{P}^{k-1} = \begin{pmatrix} P_1^{k-1} & 0 & 0 & 0 \\ 0 & P_2^{k-1} & 0 & 0 \\ 0 & 0 & P_3^{k-1} & 0 \\ 0 & 0 & 0 & P_4^{k-1} \end{pmatrix}.$$

If then

$$G_i(k) = P(0^k, s_k = G_i | 1)$$

$$U(k) = \sum_{i=1}^{4} G_i(k)$$

we have

$$(G_1(k), G_2(k), G_3(k), G_4(k)) = (c_1, c_2, c_3, c_4)\bar{P}^{k-1}$$

i.e.

$$U(k) = \sum_{i=1}^{4} c_i p_i^{k-1} .$$

Now

$$P_t(k, n) = P_t\left\{0^{\ell_1}10^{\ell_2}1 \cdots 0^{\ell_{L-1}}10^{\ell_L}\right\}$$

$$= \sum_{\ell_1=0}^{n-k} P_t(0^{\ell_1}1)P_t\left\{\underbrace{0^{\ell_2}1 \cdots 0^{\ell_{L-1}}10^{\ell_L}}_{\longleftarrow n-\ell_1-1 \longrightarrow}|1\right\} \tag{98}$$

$$L \le k + 1$$

$$0 \le \ell_1 \le n - k$$

$$\sum \ell_j = n - k$$

As for $\overline{P}(k, n)$ we find that

$$\overline{P}_t(k-1, n-\ell_1-1) = P_t\{\underbrace{0^{\ell_2}10^{\ell_3}1 \cdots 0^{\ell_{L-1}}10^{\ell_L}}_{n-\ell_1-1}|1\}$$

$$= \sum_{\ell_2=0}^{n-\ell_1-k} V_t(\ell_2)\overline{P}_t(k-2, n-\ell_1-\ell_2-2)$$

and

$$P_t(k, n) = P_t(1) \sum_{\ell_1=0}^{n-k} U_t(\ell_1)\overline{P}_t(k-1, n-\ell_1-1)$$

$$= P_t(1) \sum_{\ell_1=0}^{n-k} \sum_{\ell_2=0}^{n-k-\ell_1} U_t(\ell_1)V_t(\ell_2)\overline{P}_t(k-2, n-\ell_1-\ell_2-2) \qquad (99)$$

where

$$V_t(\ell) = U_t(\ell) - U_t(\ell+1)$$

and

$$P_t(1) = P_t(B) = P_1 \quad . \qquad (100)$$

To prove (100) note that since

$$\Pi M^t = \Pi M \cdot M^{t-1} = \Pi M^{t-1} = \cdots = \Pi$$

where $\Pi = (u_1, u_2, u_3, u_4, u_5)$ is the steady state distribution for M, Π is also the steady state distribution for M^t $t \geq 1$. Thus

$$P_t(B) = u_5$$

$$= \left[1 + \sum_{i=1}^{4} \frac{c_i}{1-p_i}\right]^{-1}$$

$$= p_1 \quad .$$

To evaluate (99) it is enough to know

$$\overline{P}_t(0, N) \quad \text{for} \quad N = 0, 1, \ldots, n \quad .$$

But note at once that

$$\overline{P}_t(0, N) = U_t(N) \quad .$$

(iv) P(k bits between extreme errors in n-block | ≥ 2 errors):

Denote this probability by P_k. Then by definition:

$$P_k = \frac{P(\text{k bits between extreme errors and} \geq 2 \text{ errors in the block})}{P(\geq 2 \text{ errors in the block})}$$

The numerator is equal to:

$$P\{0^x 1 \xleftrightarrow{k} 1 0^y\} = \sum_{x=0}^{n-k-2} P(0^x 1 \xleftrightarrow{k} 1 0^y); \qquad y = n-k-2-x$$

$$= \sum_{x=0}^{n-k-2} P(0^x 1) P(z_{k+1}=1 | z_o=1) P(0^y | 0^x 1 \xleftrightarrow{k} 1)$$

$$= P_1 r(k+1) \sum_{x=0}^{n-k-2} P(0^x | 1) P(0^y | 1)$$

$$= P_1 r(k+1) \sum_{x=0}^{n-k-2} U(x) U(n-k-2-x)$$

where

$$r(k+1) = P(z_{k+1}=1 | z_o=1)$$

and

$$U(k) = P(0^k | 1) \ .$$

Also

$$P(\geq 2 \text{ errors in a block}) = 1 - P(0, n) - P(1, n).$$

That is

$$P_k = \frac{P_1 r(k+1) \sum_{x=0}^{n-k-2} U(x) U(n-k-2-x)}{1-P(0,n)-P(1,n)} \ . \qquad (101)$$

Appendix V

Block (Symbol) Error Distribution

It is desired in this section to find for symbols in a block all statistics we have found for bits making up each symbol. In this section we shall used symbol to mean a fixed number of bits and a block to be made up of a fixed number of symbols. Specifically we shall find expressions, in terms of channel parameters, for

(i) $P^s(0^k|1)$: the probability of k error-free symbols following a given error symbol.

(ii) $P^s(0^k1|1)$: the symbol gap distribution.

(iii) $P^s(k, n)$: the distribution of error symbols in n-symbol word.

(iv) $P^s(0^n)$: probability of error-free n-symbol word.

(v) $r^s(k) = P(\text{symbol } k \text{ in error}|\text{initial symbol error})$.

Throughout this paper we shall assume a symbol to be in error if one or more of its bits are in error. It is convenient to start with

$$P^s(0) = P(\text{no symbol error})$$

where we assume each symbol is of length N bits.

$$P^s(0) = P(0, N)$$

$$= P_1 \sum_{k\geq N} U(k)$$

$$= P_1 \sum_i \frac{c_i p_i^{N-1}}{1-p_i}$$

Therefore

$$P_1^s = P(\text{symbol error})$$

$$= 1 - P_1 \sum_i \frac{c_i p_i^{N-1}}{1-p_i}$$

using

$$P_1\left[\sum \frac{c_i}{1-p_i} + 1\right] = 1$$

and

$$\sum \frac{c_i}{p_i} = 1$$

we have:

$$P_1^s = P_1\left[\sum_i \frac{c_i}{1-p_i} + 1\right] - P_1 \sum_i \frac{c_i p_i^{N-1}}{1-p_i}$$

$$= P_1\left\{\sum_{i=1}^{4} \frac{c_i}{1-p_i} + \sum_{i=1}^{4} \frac{c_i}{p_i} - \sum_i \frac{c_i p_i^{N-1}}{1-p_i}\right\}$$

$$P_1^s = P_1 \sum_i \frac{c_i(1-p_i^N)}{p_i(1-p_i)} \tag{102}$$

Next let us find

$$P^s(10^k) = P(1^{\underline{st}} \text{ symbol in error followed by } k \text{ error-free symbols})$$

$$= \sum_{j=N}^{1} P(\underset{\longleftarrow j \longrightarrow}{\vdash\!\!\longrightarrow 1}) P(0^{Nk+N-j}|1)$$

where the indicated j bits are the number of bits up to the last error in the error symbol. Thus

$$P^s(10^k) = \sum_{j=N}^{1} P_1 \sum_i c_i p_i^{Nk+N-j-1}$$

$$= P_1 \sum_i c_i p_i^{Nk+N-1} \frac{(1-p_i^N)}{p_i^N(1-p_i)}$$

or

$$P^s(10^k) = P_1 \sum_i c_i p_i^{Nk-1} \frac{(1-p_i^N)}{1-p_i} . \tag{103}$$

Putting $k = 0$ in (103) gives (101). Further

(i)

$$P^s(0^k|1) = \frac{P^s(10^k)}{P_1^s} .$$

$$= \frac{\sum c_i p_i^{Nk-1} \frac{(1-p_i^N)}{1-p_i}}{\sum \frac{c_i(1-p_i^N)}{p_i(1-p_i)}} \tag{104}$$

Hence

(ii)

$$P^s(0^k1|1) = P^s(0^k|1) - P^s(0^{k+1}|1)$$

$$P^s(0^k1|1) = \frac{\sum c_i p_i^{Nk-1} \frac{(1-p_i^N)^2}{1-p_i}}{\sum \frac{c_i(1-p_i^N)}{p_i(1-p_i)}} \tag{105}$$

$$\text{(iii)} \qquad P^s(k,n) = P^s\left\{0^{\ell_1}10^{\ell_2}1 \cdots 0^{\ell_{L-1}}10^{\ell_L}\right\}$$

$$= \sum_{\ell_1=0}^{n-k} P^s(0^{\ell_1}1)P^s\left\{\underbrace{0^{\ell_2}1 \cdots 0^{\ell_{L-1}}10^{\ell_L}}_{n-\ell_1-1}\Big|1\right\} \tag{106}$$

$$0 \le \ell_1 \le n-k$$

$$\sum \ell_j = n-k$$

$$L \le k+1 \quad .$$

First we show that

$$P^s(0^{\ell}1) = P^s(10^{\ell}) \,.$$

We can write

$$P^s(0^{\ell}1) = \sum_{j=0}^{N-1} P(0^{N\ell+j}1 \underbrace{\longmapsto}_{N-j-1})$$

where the indicated $(N-j-1)$ bits are the remaining bits after the first error of the error symbol. Thus

$$P^s(0^{\ell}1) = \sum_{j=0}^{N-1} P(0^{N\ell+j}1)$$

$$= P_1 \sum_{j=0}^{N-1} c_i p_i^{N\ell+j-1}$$

$$= P_1 \sum_i c_i p_i^{N\ell-1} \frac{(1-p_i^N)}{1-p_i} = P^s(10^{\ell}) \tag{107}$$

by (103).

Now

$$P^s\Big\{\underbrace{0^{\ell_2}1\cdots0^{\ell_{L-1}}10^{\ell_1}|1}_{n-\ell_1-1}\Big\} = \sum_{p_2=0}^{n-k-\ell_1} P^s(0^{\ell_2}1|1)P^s\Big\{\underbrace{0^{\ell_3}1\cdots0^{\ell_{L-1}}10^{\ell_1}|1}_{n-\ell_1-\ell_2-2}\Big\}$$

Denoting the LHS by $\overline{P}^s(k-1,\ n-\ell_1-1)$ we have

$$\overline{P}^s(k-1,\ n-\ell_1-1) = \sum_{\ell_2=0}^{n-k-\ell_1} P^s(0^{\ell_2}1|1)\overline{P}^s(k-2,\ n-\ell_1-\ell_2-2) \tag{108}$$

Using this in (106) and because of (107) we can write:

$$P^s(k,n) = P_1^s \sum_{\ell_1=0}^{n-k} P^s(0^{\ell_1}|1)\overline{P}^s(k-1,\ n-\ell_1-1) \tag{109}$$

That is, to evaluate $P^s(k,n)$, we need only know

$$\{\overline{P}^s(k-2,\ n-j);\ j = 2,\ \cdots,\ n-k+2;\ k \geq 2\}\ .$$

For $k = 2$ and any L

$$\overline{P}^s(0,\ L) = P^s(0,\ L|1)$$

$$= P^s(0^L|1)$$

$$\overline{P}^s(0,\ L) = \frac{\sum_i c_i p_i^{NL-1} \dfrac{(1-p_i^N)}{1-p_i}}{\sum_i c_i \dfrac{(1-p_i^N)}{p_i(1-p_i)}} \tag{110}$$

For $k > 2$, use (108) to find $\overline{P}^s(k-1, n-\ell_1-1)$. The case $k = 0$:

$$P^s(0,n) = P^s(0^n)$$

$$= \sum_{k \geq n} P^s(10^k)$$

$$= \sum_{k \geq n} P_1 \sum_i c_i p_i^{Nk-1} \frac{(1-p_i^N)}{1-p_i} \qquad \text{by (103)} .$$

i.e.

(iv) $$P^s(0^n) = P_1 \sum_i \frac{c_i p_i^{Nn-1}}{(1-p_i)} \tag{111}$$

The case $k = 1$ is included in (109).

We now find $r^s(k)$: the correlation of error symbols.

It is convenient to find

$$P(\text{symbol } k \text{ error-free} | \text{initial error symbol}) = P^s(0_k | 1_o)$$

and then use the fact that

$$r^s(k) = P^s(1_k | 1_o)$$

$$= 1 - P^s(0_k | 1_o) . \tag{112}$$

Now let

$c = (c_1, c_2, c_3, c_4)$.

M_{5x5} = transition matrix.

R_{4x4} = matrix of transitions between the good states only obtained by deleting the last row and column of M.

$S = M -$ (last column of M).

U = column vector (4x1) of 1's.

Then

$$P^S(1_o, 0_k) = P^S\{1_o \leftarrow k\text{-}1 \rightarrow 0_k\}$$

where the indicated (k-1) symbols are arbitrary N(k-1) bits. Hence

$$\begin{aligned} P^S(1_o, 0_k) = \sum_{i=1}^{N-1} P(\underbrace{\longmapsto 1}_{\leftarrow N-i \rightarrow})(CR^{i-1}, 0)M^{N(k-1)}SR^{N-1}U \\ + P(\underbrace{\longmapsto 1}_{\leftarrow N \rightarrow})(0, 0, 0, 0, 1)M^{N(k-1)}SR^{N-1}U \end{aligned} \tag{113}$$

where the indicated (N-i) bits are the number of bits up to and including last error position in the error-symbol, i = 1, N-1. The case i = 0, when the last error in the symbol is in the last bit position is shown in the second term on the RHS. So

$$P^S(1_o, 0_k) = \sum_{i=1}^{N-1} P_1\left((c_1, c_2, c_3, c_4)\begin{pmatrix} p_1^{i-1} & & & 0 \\ & p_2^{i-1} & & \\ & & p_3^{i-1} & \\ 0 & & & p_4^{i-1} \end{pmatrix}, 0\right) M^{N(k-1)}SR^{N-1}U$$

$$+ P_1(0, 0, 0, 0, 1)M^{N(k-1)}SR^{N-1}U$$

$$= P_1(CQ, 1)M^{N(k-1)}SR^{N-1}U \tag{114}$$

where

$$Q = \begin{pmatrix} \frac{1-p_1^{N-1}}{1-p_1} & & & 0 \\ & \frac{1-p_2^{N-1}}{1-p_2} & & \\ & & \frac{1-p_3^{N-1}}{1-p_3} & \\ 0 & & & \frac{1-p_4^{N-1}}{1-p_4} \end{pmatrix}.$$

Thus since

$$P^s(0_k | 1_o) = \frac{P^s(1_o, 0_k)}{P_1^s}$$

we have by (112) and (114):

$$r^s(k) = 1 - \frac{P_1}{P_1^s} (CQ, 1) M^{N(k-1)} SR^{N-1} U$$

$$= 1 - \frac{(CQ,1) M^{N(k-1)} SR^{N-1} U}{\sum c_i \frac{(1-p_i^N)}{p_i(1-p_i)}} . \qquad (115)$$

Let us check (114) by putting $N = 1$ and $k = 1$, in which case the LHS

$P^s(1_o, 0_k)$ reduces to $P(10)$.

If $N = k = 1$, then $Q \equiv 0$; $M^{N(k-1)} = I_{5x5}$; $R^{N-1} = I_{4x4}$; and so the RHS becomes: $P_1(0, 0, 0, 0, 1) ISIU = P_1 \sum_i c_i = P(10)$.

Burst Statistics

a. Distribution and mean of burst lengths

b. P(k errors in a given burst of length n) and its mean.

c. Block burst distribution.

(a) Let $L(n) = P(\text{burst of length } n)$; $n \geq 1$. Then by definition

$$L(n) = P\left\{\underbrace{0^{\ell_1}10^{\ell_2}1 \cdots 10^{\ell_L}1}_{n-1}0^{t}1 \,\middle|\, 0^{r}1\right\} \tag{116}$$

where

$r, t \geq G$ the guard space,

$0 \leq \ell_j \leq G-1$

$\sum \ell_j \leq n-2$.

Proposition 1

$$L(n) = \begin{cases} 0 & \text{for} \quad n \leq 0 \\ U(G)\bar{L}(n); & n \geq 1 \end{cases} \tag{117}$$

where

$$\bar{L}(n) = \sum_{\ell=0}^{\min(G-1,n-2)} V(\ell)\bar{L}(n-\ell-1); \qquad n \geq 2$$

$$\bar{L}(1) = 1 \quad .$$

Proof: It is clear by definition that $L(0) = 0$. By (116)

$$L(n) = P\{0^{\ell_1}10^{\ell_2}1 \cdots 10^{\ell_L}10^t1|1\}$$

$$= \sum_{t\geq G} P\{\underbrace{0^{\ell_1}10^{\ell_2}1 \cdots 10^{\ell_L}1}_{n-1}|1\}P(0^t1|1) \ . \tag{118}$$

$$\sum_{t\geq G} P(0^t1|1) = \sum_i c_i p_i^{G-1} = U(G) \tag{119}$$

and if we denote

$$\bar{L}(n) = P\{\underbrace{0^{\ell_1}10^{\ell_2}1 \cdots 10^{\ell_L}1}_{n-1}|1\}$$

$$= \sum_{\ell_1=0}^{\min(G-1,n-2)} P(0^{\ell_1}1|1)P\{\underbrace{0^{\ell_2}1 \cdots 10^{\ell_L}1}_{n-\ell_1-2}|1\}$$

$$= \sum_{\ell=0}^{\min(G-1,n-2)} V(\ell)\bar{L}(n-\ell-1) \tag{120}$$

so that $\bar{L}(1) = 1$, then substituting (119) and (120) in (118) gives (117).

The expected (average) burst length is given by

$$\sum_{n\geq 1} nL(n) = U(G) \sum_{n\geq 1} n\bar{L}(n) \tag{121}$$

We shall find $\sum_{n\geq 1} n\bar{L}(n)$ by the method of moment generating functions (MGF)*.

Let us find the MGF of $\bar{L}(n)$. Define

$$V_G(\ell) = \begin{cases} V(\ell) & \text{if} \quad \ell \leq G-1 \\ 0 & \text{otherwise} \end{cases} . \tag{122}$$

Since

$\bar{L}(k) = 0$ for $k \leq 0$ we have

$$\bar{L}(n)X^n = X \sum_{\ell=0}^{\min(G-1,n-2)} V(\ell)X^{\ell}\bar{L}(n-\ell-1)X^{n-\ell-1}$$

$$= X \sum_{\ell=0}^{\infty} V_G(\ell)X^{\ell}\bar{L}(n-\ell-1)X^{n-\ell-1} \; ; \quad n \geq 2 \tag{123}$$

and if we define

$$R(X) \equiv \sum_{n=1}^{\infty} \bar{L}(n)X^n \; ; \quad V_G(X) \equiv \sum_{\ell=0}^{\infty} V_G(\ell)X^{\ell} \tag{124}$$

$R(X)$ and $V_G(X)$ exist because $\bar{L}(n)$ are probabilities and $V_G(\ell)$ is non-zero only for finitely many ℓ's. Then summing over n in (123) we get

---

*We are grateful to Howard Rumsey for pointing out this elementary but powerful way of generating the function $\bar{L}(n)$ from its recursion expression (120).

$$R(X) - X = X \sum_{\ell=0}^{\infty} V_G(\ell)X^{\ell} \sum_{n=\ell+2}^{\infty} \overline{L}(n-\ell-1)X^{n-\ell-1}$$

$$= XV_G(X)R(X) \tag{125}$$

or

$$R(X) = \frac{X}{1-XV_G(X)}$$

$$= \sum_{j=0}^{\infty} X(XV_G(x))^j \ . \tag{126}$$

Thus

$$\sum_{n\geq 1} nL(n) = U(G)R'(X)\Big|_{X=1} \tag{127}$$

and

$$R'(X)\Big|_{X=1} = \frac{1+X^2V_G'(X)}{(1-XV_G(X))^2}\Bigg|_{X=1} = \frac{1+V_G'(1)}{[1-V_G(1)]^2} \ . \tag{128}$$

We shall need

$$R''(X) = \frac{(1-XV_G(X))^2[2XV_G'(X)+X^2V_G''(X)]+2(1+X^2V_G'(X))(1-XV_G(X))(V_G(X)+XV_G'(X))}{(1-XV_G(X))^4}$$

giving

$$R''(X)\Big|_{X=1} = \frac{(1-V_G(1))(2V_G'(1)+V_G''(1))+2(1+V_G'(1))(V_G(1)+V_G'(1))}{(1-V_G(1))^3} \ , \tag{129}$$

in finding the variance of $L(n)$. Thus, since

$$R''(1) = \sum_{n\geq 1} n(n-1)\overline{L}(n)$$

the variance of $L(n)$ is given by:

$$U(G) \sum_{n\geq 1} n^2\overline{L}(n) - \left[U(G) \sum_{n\geq 1} n\overline{L}(n)\right]^2$$

$$= U(G)R''(1) + U(G)R'(1) - (U(G)R'(1))^2 \;; \qquad (130)$$

where

$$V_G(X) = \sum_{\ell=0}^{G-1} V(\ell)X^{\ell}$$

$$= \sum_{i=1}^{4} \frac{c_i(1-p_i)}{p_i} \sum_{\ell=0}^{G-1} p_i^{\ell}X^{\ell}$$

$$= \sum_i \frac{c_i(1-p_i)(1-(p_iX)^G)}{p_i(1-p_iX)}$$

so

$$V_G(1) = \sum_i \frac{c_i(1-p_i^G)}{p_i} \;. \qquad (131)$$

$$V_G'(X) = \sum_i \frac{c_i(1-p_i)\{1-(p_iX)^G)-G(1-p_iX)(p_iX)^{G-1}\}}{(1-p_iX)^2}$$

so

$$V_G'(1) = \sum_i \frac{c_i}{1-p_i} \{1 - p_i^G - G(1 - p_i)p_i^{G-1}\} \tag{132}$$

$$V_G''(X) = \sum_i \frac{c_i(1-p_i)p_i\{2(1-(p_iX)^G)-G^2(1-p_iX)^2(p_iX)^{G-2}-2G(1-p_iX)(p_iX)^{G-1}\}}{(1-p_iX)^3}$$

so

$$V_G''(1) = \sum_i \frac{c_ip_i\{2(1-p_i^G)-G^2(1-p_i)^2p_i^{G-2}-2G(1-p_i)p_i^{G-1}\}}{(1-p_i)^2} \quad . \tag{133}$$

(b) As in Ref. [2], we write

$$P(k \text{ errors in a given burst of length } n) = \frac{\overline{Q}(k,n)}{\overline{L}(n)} \tag{134}$$

and state expression for $\overline{Q}(k,n)$ in the following.

<u>Proposition 2:</u>

$$\overline{Q}(k,n) = \begin{cases} 0 & \text{if} \quad k = 0,\ n \geq 0; \quad \text{or} \quad k > n \\ 1 & \text{if} \quad n = k = 1 \\ \sum\limits_{\ell=0}^{\min(G-1,n-k)} V(\ell)\overline{Q}(k-1,n-\ell-1) \ ; & n \geq k \geq 2 \end{cases} \tag{135}$$

<u>Proof</u>:

$P(k$ errors in a given burst of length $n)$

$$= \frac{P(\text{burst of length } n \text{ with } k \text{ errors})}{P(\text{burst of length } n)} \tag{136}$$

Denote the numerator by $Q(k,n)$. Then by definition

$$Q(k,n) = P\{0^{\ell_1}10^{\ell_2}1 \cdots 10^{\ell_L}10^{t}1 | 0^{r}1\}$$

where

$$r, t \geq G; \quad L \leq k-1; \quad \sum \ell_j = n-k$$

$$0 \leq \ell_j \leq \min(G-1, n-k-\sum_{i=1}^{j-1} \ell_i) \quad .$$

Thus by (118),

$$Q(k,n) = U(G)P\{\underbrace{0^{\ell_1}10^{\ell_2}1 \cdots 10^{\ell_L}1}_{n-1}|1\} \quad . \tag{137}$$

Let

$$\bar{Q}(k,n) = P\{\underbrace{0^{1}10^{2}1 \cdots 10^{L}1}_{n-1}|1\} \; . \tag{138}$$

Since the (n-1) bits indicated must contain (k-1) errors we have:

$$\bar{Q}(1,n) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\bar{Q}(0,n) \equiv 0 \qquad \text{for} \quad n \geq 0$$

and it is clear that

$$\bar{Q}(k,n) = 0 \qquad \text{for} \quad k > n \quad .$$

Hence for $2 \le k \le n$ we can write from (138)

$$\bar{Q}(k,n) = \sum_{\ell_1=0}^{\min(G-1,n-k)} V(\ell_1) P\{\underbrace{0^{\ell_2} 1 \cdots 1 0^{\ell_L} 1}_{n-\ell_1-2} | 1\}$$

$$= \sum_{\ell=0}^{\min(G-1,n-k)} V(\ell)\bar{Q}(k-1,\ n-\ell-1) \qquad (139)$$

since the $(n-\ell-2)$ bits indicated must contain $(k-2)$ errors. Thus by (137)

$$Q(k,n) = U(G)\bar{Q}(k,n) \quad . \qquad (140)$$

Combining this with (117) and (136) we obtain

$$P(k \text{ errors in a given burst of length } n) = \frac{Q(k,n)}{L(n)}$$

$$= \frac{\bar{Q}(k,n)}{\bar{L}(n)} \; ; \quad n \ge 1 \; .$$

Here also let us use the method of moment generating functions to find the mean number of errors in a burst of given length. Denote this mean by $\bar{K}_n$; n is the burst length. So

$$\bar{K}_n = \frac{\sum_{k=2}^{n} k\bar{Q}(k,n)}{\bar{L}(n)} \; ; \qquad n \ge k \ge 2 \qquad (141)$$

We know that

$$\bar{K}_1 = 1$$

$$\bar{K}_2 = 2 \quad .$$

By definition of $\bar{Q}(k,n)$ we can write

$$\overline{Q}(k,n) = \sum_{\ell=0} V_G(\ell)\overline{Q}(k-1,\ n-\ell-1) \qquad (142)$$

where $V_G(\ell)$ is as defined in (122). Note that (142) holds for all values of k and n except for $k = n = 1$. So if we denote

$$Q(y,x) \equiv \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \overline{Q}(k,n)y^k x^n$$

$$= \sum_{k\geq 2} \sum_{n\geq 2} \overline{Q}(k,n)y^k x^n + yx \qquad (143)$$

$Q(y,x)$ is well defined since $\overline{Q}(k,n)$ are probabilities. We can write, for all k and n

$$Q(y,x) - yx = xy \sum_{k=2}^{\infty} \sum_{\ell=0}^{\infty} V_G(\ell)x^{\ell} \sum_{n=2}^{\infty} \overline{Q}(k-1,\ n-\ell-1)x^{n-\ell-1}y^{k-1}$$

$$= xyV_G(x) \sum_{k=2}^{\infty} \sum_{n=\ell+2}^{\infty} \overline{Q}(k-1,\ n-\ell-1)x^{n-\ell-1}y^{k-1}$$

$$= xyV_G(x)Q(y,\ x) \qquad (144)$$

or

$$Q(y,x) = \frac{xy}{1-xyV_G(x)} \ . \qquad (145)$$

Denote the partial derivative of $Q(y,x)$ w.r.t. y by $Q_y(y,x)$. Then

$$Q_y(y,x) = \frac{x}{(1-xyV_G(x))^2}$$

so

$$Q_y(1,x) = \frac{1}{x} R^2(x) \tag{146}$$

$$= \frac{1}{x}\left[\sum_{n=1}^{\infty} \overline{L}(n)x^n\right]^2 \qquad \text{by (124)}$$

By (143), for a fixed n

$\sum_{k\geq 1} k\overline{Q}(k,n)$ is equal to the coefficient of the $n^{\underline{th}}$ term of $Q_y(1,x)$

i.e. the coefficient of the $n^{\underline{th}}$ term of $\frac{1}{x}\left[\sum_{n=1}^{\infty} \overline{L}(n)x^n\right]^2$ which is equal to

$$\sum_{k=1}^{n} \overline{L}(k)\overline{L}(n-k+1) \tag{147}$$

Therefore by (141)

$$\overline{K}_n = \frac{\sum_{k=1}^{n} \overline{L}(k)\overline{L}(n-k+1)}{\overline{L}(n)} \tag{148}$$

The variance of $\frac{\overline{Q}(k,n)}{\overline{L}(n)}$ is given by the coefficient of the $n^{\underline{th}}$ term of

$$\frac{Q_{yy}(1,x)}{\overline{L}(n)} + \overline{K}_n - \overline{K}_n^2 \tag{149}$$

where

$Q_{yy}(y,x)$ is the second partial derivative of $Q(y,x)$

w.r.t. y i.e.

$$Q_{yy}(y,x) = \frac{2X^2 V_G(x)}{(1-xyV_G(x))^3}$$

and

$$Q_{yy}(1,x) = \frac{2X^2 V_G(x)}{(1-xV_G(x))^3}$$

$$= 2V_G(x)\,\frac{1}{x}\,R^3(x) \qquad (150)$$

(c) Let d = block guard space and $L^s(n)$ = probability of a block burst of length n (each block is s bits long).

$$L^s(n) = P^s\{0^{\ell_1}10^{\ell_2}1 \cdots 10^{\ell_L}10^t1|0^r1\}$$

$$r,t \geq d\ ; \quad 0 \leq \ell_j \leq d-1; \quad \sum \ell_j \leq n-2 \qquad (151)$$

Thus

$$L^s(n) = \sum_{t\geq d} P^s\{\underbrace{0^{1}1 \cdots 10^{1}1}_{n-1}|1\}P^s(0^t1|1)$$

with

$$\sum_{t\geq d} P^s(0^t1|1) = \frac{\sum_i c_i p_i^{sd} \frac{(1-p_i^s)}{p_i(1-p_i)}}{\sum_i c_i \frac{(1-p_i^s)}{p_i(1-p_i)}} \qquad (152)$$

$$= U^s(d)\ .$$

Put

$$\overline{L}^s(n) = P^s\{\underbrace{0^{\ell_1}1 \cdots 10^{\ell_L}1}_{\longleftarrow\ n-1\ \longrightarrow}|1\}$$

$$= \sum_{\ell=0}^{\min(d-1,n-2)} V^s(\ell)\overline{L}^s(n-\ell-1) \tag{153}$$

so that $\overline{L}^s(1) = 1$; where

$$V^s(\ell) = P^s(0^\ell 1|1\ )$$

$$= \frac{\sum c_i p_i^{s\ell} \frac{(1-p_i^s)^2}{p_i(1-p_i)}}{\sum c_i \frac{(1-p_i^s)}{p_i(1-p_i)}} .$$

Then we can state $L^s(n)$ in the following

Proposition 3:

$$L^s(n) = \begin{cases} 0 & \text{for} \quad n \le 0 \\ U^s(d)\overline{L}^s(n)\ ; & n \ge 1 \end{cases} .$$