ESD-TR-67-379

MTP-55

ERROR CONTROL TECHNIQUES USING
BINARY SYMBOL BURST CODES

SEPTEMBER 1967

K. Brayer

AD0660582

ERROR CONTROL TECHNIQUES USING
BINARY SYMBOL BURST CODES

SEPTEMBER 1967

K. Brayer

FOREWORD

This report was prepared by the Range Communications Planning and Technology Subdepartment of The MITRE Corporation, Bedford, Massachusetts, under Contract AF 19(628)-5165. The work was directed by the Development Engineering Division under the Aerospace Instrumentation Program Office, Air Force Systems Command, Electronics Systems Division, Laurence G. Hanscom Field, Bedford, Massachusetts. Captain J. J. Centofanti served as the Air Force Project Monitor for this program, identifiable as ESD (ESSI) Project 5932, Range Digital Data Transmission Improvement.

REVIEW AND APPROVAL

Publication of this technical report does not constitute Air Force approval of the report's findings or conclusions. It is published only for the exchange and stimulation of ideas.

OTIS R. HILL, Colonel, USAF
Director of Aerospace Instrumentation
    Program Office

ABSTRACT

Much has been written on the theoretical description of error correcting codes but, due to a lack of actual channel error patterns, little has been said of practical performance. In this paper the performance of three types of error control is evaluated for the case of independent random errors and for an actual channel exhibiting dense bursts. The selected codes are burst codes with high probabilities of error detection and correction.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Concluded)

SECTION I

INTRODUCTION

In recent years, work in the areas of coding through block, retransmission, and hybrid techniques was reported in the literature. Due to a lack of real channel data [1, 2], the investigation of how these techniques perform in real channel environments is still unknown. However, the MITRE Corporation has collected extensive error pattern data at high speeds in real channel environments such as HF and troposcatter (Appendix 1) [3, 4, 5, 6]. It is the purpose of this paper to describe and compare the performance of techniques for hybrid error control, forward error control, and retransmission error control in a real channel environment and to contrast this performance with that which may be expected in a channel exhibiting independent errors.

Error control methods are needed in the transmission of digital data if the error rate without control is unacceptable. An example of such a situation is computer-to-computer communications. Not only is nearly-perfect transmission required in the transmission of data, but error-free transmission is mandatory when computer programs are transmitted. The method of error control described in this paper will enable the user to select the probability with which errors in the received message are acceptable and then to transmit the information with the highest possible efficiency available for the combination of modulation technique, data rate, and environment in which the transmission will take place.

1

SECTION II
HYBRID ERROR CONTROL SYSTEM

GENERAL DESCRIPTION

Hybrid error control is the combination of a forward error correcting code (e.g., block code) with the use of retransmission error control. As a combination of two coding procedures, hybrid error control exhibits the advantages of both and the disadvantages of neither. Specifically, in a random error environment, the forward error correcting code will provide error correction within the code block and retransmission will eliminate any residual errors. In a high density burst* error environment, the retransmission will be used for error correction while the forward error control eliminates the errors in the intervals between bursts. If the burst error density increases to an extremely high level and the bursts get excessively long, or if random errors occur with a bit error rate higher than a threshold (percentage of correctable error bits) derived from the code, the system enters continuous retransmission and will fail. In failure, the code corrects an insufficient number of errors and burst length exceeds the block length, causing continuous retransmission. However, since the codes will be chosen from a family, the code power can be increased to ward off failure.

The hybrid error control system described herein is a combination of retransmission and random burst symbol error correcting codes, operated as a continuous system. The system will be described initially in terms of retransmission with a code, and the specific properties of the selected code will be identified later. Finally, the hybrid system will be specialized to forward error correction alone and retransmission alone.

---

*See Appendix I for definition of error pattern parameters.

The Hybrid Retransmission Technique

The hybrid error control system operates according to the following transmission procedure. Data is collected at some data center in large amounts. This data is encoded with an error detecting code and then encoded with an error correcting code. The information is now transmitted in blocks to a receiver site where the process of error correction is activated. After error correction, a final detection is made for residual errors. If there are no errors, the data is transferred to the data sink. If there are errors, the block identification number is recorded for future use by the retransmission system and the block is rejected. During the finite time that it takes to decode one block of information, the next block is being transmitted. If transmission were interrupted for retransmission of rejected blocks, the finite delay would cause a system slowdown. Thus, retransmission requests transmitted back to the data source initiate block regeneration at the conclusion of the initial transmission. The only case in which there can be delay time between blocks is if the last block must be retransmitted.

The above technique can be used for transmission of non-real-time information such as computer programs or large amounts of data to be used as pre-flight and post-flight information in space vehicle launches. If near-real-time transfer is required, a request for retransmission can be initiated immediately and the block (rejected) in question will be regenerated and retransmitted following the block in transmission when the request is made. Actual real-time operation cannot be achieved with this system.

The performance of such a hybrid system is described by the following transmission efficiency equation:

$$\text{Transmission efficiency (E)} = \frac{\text{Information to be transmitted (sec)}}{\text{Total time of operation (sec)}}$$

$$\text{Transmission efficiency (E)} = \frac{I}{I(1 + R_d + R_c) + I_R(1 + R_d + R_c) + D_T} \tag{1}$$

3

where I = information to be transmitted (seconds) = Data (bits)/Data Rate (Bits/Sec)

R_d = error detection redundancy = ratio of detection parity to information bits

R_c = error correction redundancy = ratio of correction parity to detection parity and information bits.

I_R = information rejected (Sec)

D_T = delay time (Sec) = error control delay (1 + messages transmitted) (Sec) for interrupted systems

If the system operates as a continuous system (as will be the case in this paper),

$$D_T = \text{error control delay.}$$

If there are no retransmissions,

$$E = \frac{I}{I\,(1 + R_d + R_c) + D_T} \tag{2}$$

since $I_R = 0$.

In the case where D_T is negligible compared to I and there are no retransmissions,

$$E = \frac{1}{1 + R_d + R_c} = \text{rate of the hybrid code.} \tag{3}$$

If D_T is zero or if $D_T << I$, the efficiency is independent of I and of the data rate. Thus, it is possible to consider the efficiency as the effective rate $\frac{1}{1 + R}$ of the hybrid system. The effective rate is always less than the code rate. The nearness of E to the code rate acts as a system performance criterion.

Block Numbering Techniques

Although it was previously stated that the block number is recorded, it is not necessary, in the case of block rejection, to transmit block numbers if a powerful synchronization technique is used. Since every block has the same

4

number of bits and the receiving computer knows the sequence in advance (i.e., sequence is 1, 2, ..., n), the computer can assign these numbers. Similarly in retransmission, since all retransmission blocks are placed together at the end of the message, the computer can take advantage of its prior knowledge. The only exception is if there is a delay between blocks, in which case a delay message must be sent to prevent confusion of the receiving computer by the random channel output, or if the near-real-time mode is used. In the latter mode, the retransmitted block may occur immediately after the block in the channel or one block later if the regeneration time is not sufficient.

Need for Retransmission Data Link

It might appear that the need for a retransmission link would reduce the efficiency of the system. This is only partially true. If a link must be set up, a narrow bandwidth will be required since there is little information to transmit. It is much more likely that the return link already exists due to a need for communication in two directions. The retransmission requests can be time-division multiplexed with the messages of other users on the return link and, with careful allocation of communication availability, there would be no hardship on the regular users. This paper will consider efficiency on the basis of the data link only assuming return link availability. Furthermore the retransmission requests can be low rate coded to protect against errors.

Specialization to Retransmission Error Control

In a system where there is retransmission error control only, two modifications need to be made to the hybrid error control system. The obvious one is that the error correction redundancy is removed. Thus in the transmission efficiency equation, $R_c = 0$.

The other modification to both the system and the transmission efficiency equation is with respect to delay. In the initial hybrid system, delay is due to

5

the use of error correction and spacing between blocks. With the error correcting code eliminated ($R_c = 0$), the part of the delay due to the error correction is likewise eliminated.

CODE DESCRIPTION

The codes which will be used are from a class of p^m symbol error correcting codes.[7] These codes have been selected because of their high error correcting capability for little redundancy and because of their high residual error detecting capability. These capabilities are such that the codes will be used for both R_c and R_d simultaneously, thus simplifying the efficiency equation. The codes are referred to as (q, n, e) codes, where q is the number of symbols in the alphabet expressible as p^m, n is the number of symbols in a code block, and e is the number of symbol errors that can be corrected in a code block. The number p may be any prime power, (p = 2, 3, 5, 7 ..., m = 1, 2, 3, 4, ...), and n is any positive integer not divisible by p. For the binary case, p = 2 and $q = 2^m$. Thus the alphabet used by the code consists of the set of binary m-tuples. It is convenient in this binary subfamily of codes to select $n = 2^m - 1$. The correction of e symbol errors then requires that 2e (2e< n) check symbols be included in the block.

The codes are random symbol error correcting codes. Within a block of n symbols, e symbols are correctable if there are at most e symbol errors in the block. With respect to bit errors, the codes are burst codes since if there is one bit error in a symbol, it is corrected, but the code would be more efficient if there were m bit errors in a symbol. As an example if m = 3, n = 7, and e = 2, then any two random bit errors in the 21-bit block are corrected by the code. But the code will also correct two solid bursts of three errors in the block or one solid burst of six errors (which exists in two adjacent symbols).

The parameter values which will be given special interest in this paper are:

m	(bits per symbol)	=	8
n	(symbols per block)	=	255

m·n	(bits per block)	=	2040
e	(correctable symbols)	=	0 to 32
2e	(check symbols per block)	=	0 to 64
n-2e	(information symbols per block)	=	255 to 191
$\dfrac{n-2e}{n}$	(code rate)	=	1.0 to 0.75

In the encoding and decoding process, the m bit symbols of a message are treated as elements of a finite field. When dealing with binary m bit symbols, the finite field may be considered to consist of all possible m-tuples or m bit numbers. A special m bit arithmetic is used such that all arithmetic operations between elements of the field yield results which are also m bit numbers and therefore contained in the field. Also, it may be stated that all non-zero members of the field may be represented by a special number, α, raised to power i, where i can vary from zero to 2^m-2. The number α is referred to as a primitive element of the field. Addition is performed modulo-two with no carry. Multiplication in finite field arithmetic is analogous to conventional multiplication in that numbers raised to a power may be multiplied by adding their exponents. Thus, for example, α^2 times α^6 equals α^8. The addition of exponents when multiplying is performed modulo 2^m-1. Thus, in the field with m=4, $\alpha^{14} \times \alpha^3$ would equal α^2 since 17 modulo 15 is equal to 2.

The equation below defines a relationship which is true for all valid messages or code blocks.

$$\sum_{i=0}^{2^m-2} B_i \, \alpha^{ij} = 0, \qquad j = 0, 1, 2 \ldots 2e - 1 \qquad (4)$$

where

B_i	=	the ith message symbol
m	=	bits per message symbol
α	=	primitive elements of finite field
e	=	number of symbols which can be corrected.

7

According to Equation (4), the summation of the products of all message symbols multiplied by α^{ij} is equal to zero where i specifies the order of a symbol in the message and j takes on all integer values from 0 to 2e-1.

To describe the encoding and decoding operations, a simple example in which m = 3, n = 7, and e = 2 is considered. For these parameters, a message or code block consists of seven 3-bit symbols. Since e = 2, there will be four check symbols in each block, and the remaining three symbols are information symbols. The problem in encoding is: given three information symbols, B_0 to B_2, compute the four check symbols, B_3 to B_6. From Equation (4), four simultaneous equations corresponding to the four values of j can be written:

$$
\begin{aligned}
j &= 0; & B_0 &+ B_1 + B_2 & \ldots &+ B_6 &= 0 \\
j &= 1; & B_0 &+ B_1\alpha + B_2\alpha^2 & \ldots &+ B_6\alpha^6 &= 0 \\
j &= 2; & B_0 &+ B_1\alpha^2 + B_2\alpha^4 & \ldots &+ B_6\alpha^{12} &= 0 \\
j &= 3; & B_0 &+ B_1\alpha^3 + B_2\alpha^6 & \ldots &+ B_6\alpha^{18} &= 0
\end{aligned}
\tag{4a}
$$

Solution of the four simultaneous equations above gives the required check symbols, B_3 to B_6.

Now consider the problem of decoding [7, 8, 9, 10, 11, 12, 13]. The received message symbols are called B' where each B_i', consists of the original message symbol B_i added (modulo two) to an error symbol V_i which may or may not be zero. The first step of the decoding procedure is to compute 2e numbers called syndromes, S_j, in accordance with Equation (5).

$$
S_j = \sum_{i=0}^{2^m - 2} B_i' \alpha^{ij}, \qquad j = 0, 1, \ldots, 2e - 1 \tag{5}
$$

Substituting $B_i' = B_i + V_i$, we obtain

$$S_j = \sum_{i=0}^{2^m-2} B_i \alpha^{ij} + \sum_{i=0}^{2^m-2} V_i \alpha^{ij} . \qquad (5a)$$

Since, from Equation (4)

$$\sum_{i=0}^{2^m-2} B_i \alpha^{ij} = 0,$$

$$S_j = \sum_{i=0}^{2^m-2} V_i \alpha^{ij} . \qquad (6)$$

If no errors have occurred, the V_i are all zero and the syndromes will also be zero. Hence, if the syndromes are found to be all zeros, the message is assumed to have been received error-free. If one or more non-zero syndromes are obtained however, errors are known to have occurred and the message must be corrected. If we assume that two errors had occurred in our sample message, they could be determined from the following equations

$$
\begin{aligned}
S_0 &= V_1 + V_2 \\
S_1 &= V_1 \alpha^{i_1} + V_2 \alpha^{i_2} \\
S_2 &= V_1 \alpha^{2i_1} + V_2 \alpha^{2i_2} \qquad (6a) \\
S_3 &= V_1 \alpha^{3i_1} + V_2 \alpha^{3i_2}
\end{aligned}
$$

The four unknowns in these equations are V_1, V_2, α^{i_1}, and α^{i_2}, where V_1 and V_2 are the two error symbols, and i_1 and i_2 correspond to their locations in the message. After solving for the error values and their locations, the syndromes are adjusted for the errors tentatively assumed to have occurred. If the syndromes are now equal to zero, it is decided that the assumed errors did

9

in fact occur and the message is corrected accordingly. If any of the adjusted syndromes are non-zero, however, the message is considered to be uncorrectable due to excessive errors.

It was stated that if the syndromes are zero the code assumes the message to be error-free or corrected. There is, of course, a possibility that the code has failed to detect an error. In this case, no retransmission will be requested in a hybrid system and the errors will be passed on. In Figures 1 and 2 the probability of an undetected error is presented for various codes. The curves are based on the derivation resulting in Equation (10). For purpose of this derivation, the following terms are defined.

m = number of bits/symbol

n = number of symbols/block

k = number of information symbols/block

e = maximum number of errors that can be corrected.

This class of codes fails to detect a block in error if, and only if, the received code vector is no greater than distance e away from some code vector other than the one sent. Thus, assuming all n-symbol error vectors are equally likely, and that an error has occurred, the probability of an erroneous block going undetected (and thus also being erroneously "corrected") is

$$\frac{\text{Number of error vectors of distance} \leq e \text{ from a non-zero code vector}}{\text{Number of error vectors not of distance} \leq e \text{ from the zero vector}} \tag{7}$$

where the correct vector is used as a zero reference.

Number of ways a symbol can be non-zero $= 2^m - 1 = n$

Number of vectors with i non-zero symbols $= \binom{n}{i} n^i$

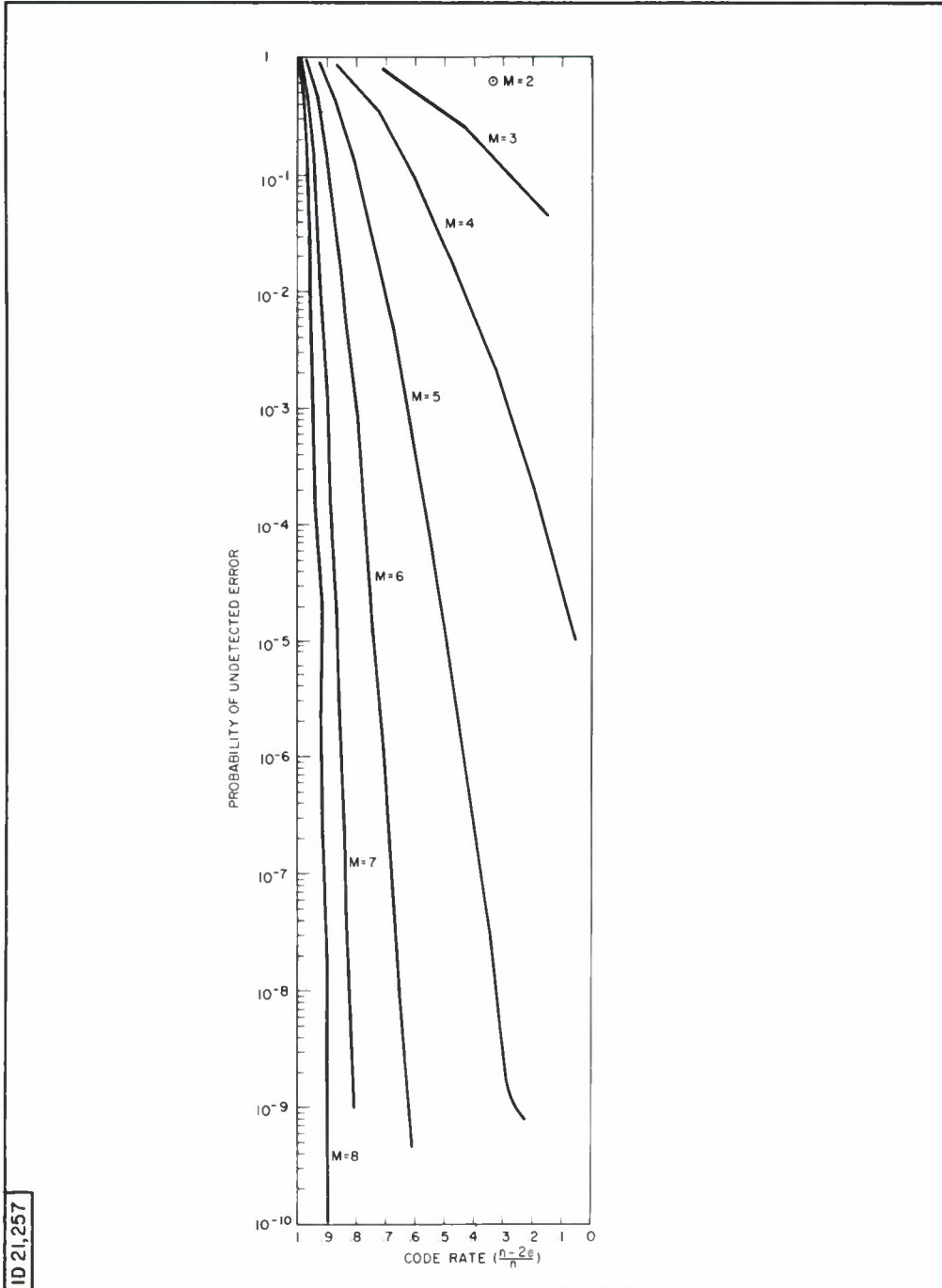Number of non-zero code vectors $= 2^{km} - 1$

10

Figure I. Probability of Undetected Error versus Code Rate for P^M Symbol Codes
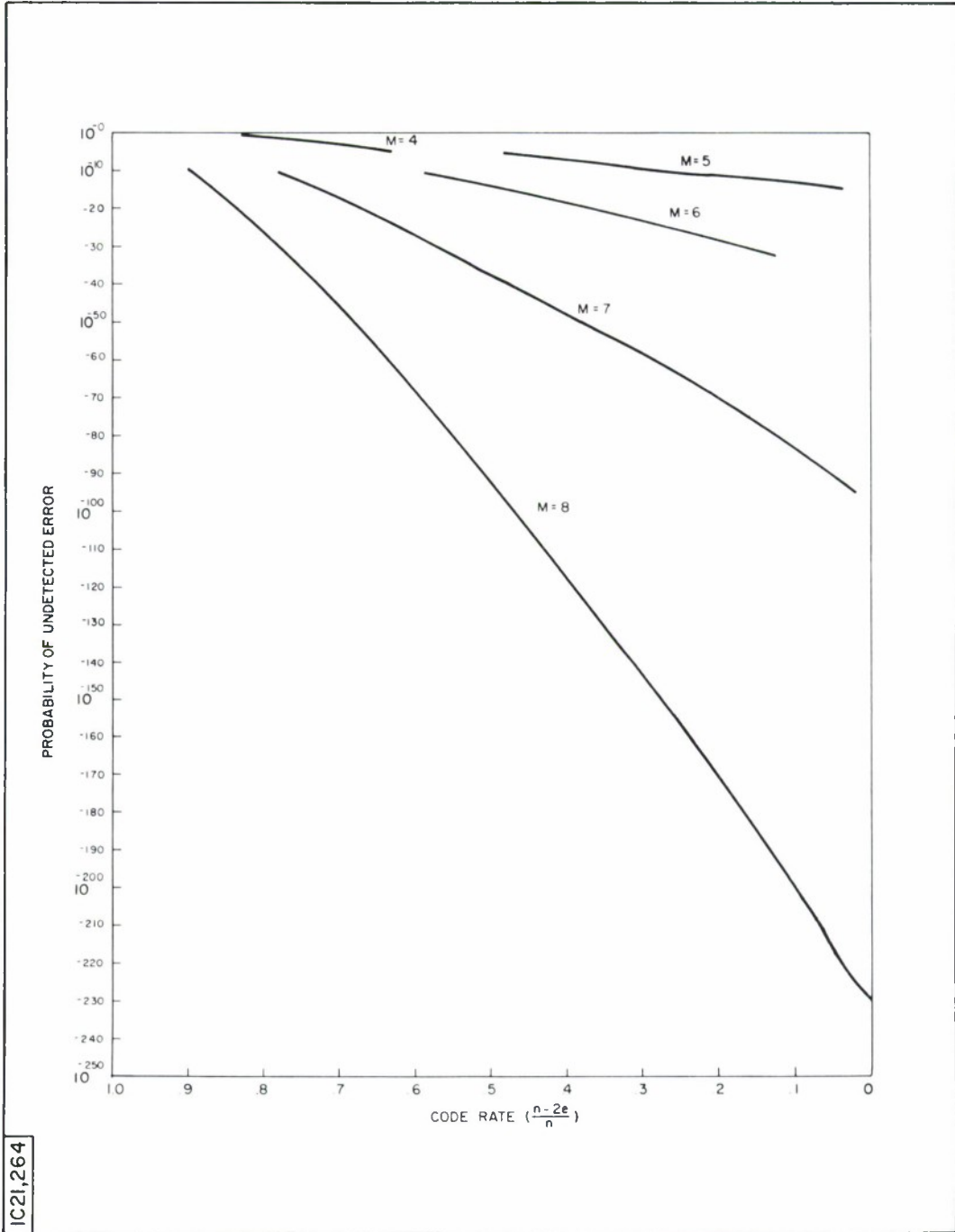
11

Figure 2. Probability of Undetected Error versus Code Rate for P^M Symbol Codes

12

Number of vectors of distance $\le e$ from a non-zero code vector = (8)

$$(2^{km}-1) \sum_{i=0}^{e} \binom{n}{i} n^i$$

Number of vectors of distance $> e$ from the zero vector $= 2^{nm} - \sum_{i=0}^{e} \binom{n}{i} n^i$ (9)

$$P\{\text{Undetected error}\} = \frac{(2^{km}-1) \sum_{i=0}^{e} \binom{n}{i} n^{i*}}{2^{nm} - \sum_{i=0}^{e} \binom{n}{i} n^i}$$ (10)

The next section contains a description of a case with an actual error distribution which shall be taken into account in lowering the bound on the probability of an undetected error.

The curves demonstrate the performance of the code in failing to detect errors as a function of code rate for $2 \le m \le 8$. The case $m = 1$ can be considered as the trival code where there is one bit/symbol and one symbol/block and the failure probability is obviously 1 since there is no redundancy.

Unlike many other codes, a bursty channel will not reduce the performance of these codes. Since the code corrects only symbols, it does not matter whether or not one or all bits in the symbol are in error. The bit error correction capability is, in fact, better if all the bits in a symbol are in error. As an example, if $m = 8$ and the code rate is 3/4, 32 symbols can be corrected. If these error symbols are consecutive, the code can correct 256 consecutive errors. If the bursts are excessively long, m must be increased.

When used in conjunction with the efficiency curves in later sections, Figures 1 and 2 will serve to give the maximum efficiency available for a

---

*This gives a worst case bound on P (undetected block error given a symbol error). An alternative approach to this calculation is presented by Mitchell.[14]

combination of environment, code, data rate, length of information message, and probability of acceptable undetected errors.

Delay Time of Codes

One of the penalties of error correction is the delay introduced into the system by parity check coding. To evaluate the delay, the three places in which delay can occur (i.e., encoding, transmission, and decoding) are considered.

In encoding, as the information bits are made available by the source, they are used in starting the calculation of check symbols and "outputed" to the modem. When the last information bit is "outputed", all check bits are now immediately available. Thus, there is no delay in encoding as long as the coder can process bits at a rate higher than the channel data rate.

In decoding, calculations cannot start until all symbols are available. Thus, there is a delay of one block in transmission. For available hardware, calculations in decoding require one block time at most. Thus, the overall delay due to coding for continuous data transmission is twice the block length. If the blocks were sent in a spaced discontinuous system, the delay due to coding would be twice the block length per transmitted block. If such a system were used, this delay could dominate the denominator of the efficiency expression, Equation (1), thus drastically reducing the efficiency. Since an actual system could have various delays up to twice the block length, we will assume a worst case in our performance analysis and use the maximum value of delay in all cases.

Specialization to Retransmission and Forward Error Control

In the case of retransmission error control, the same codes will be used for error detection only. Since there is no correction, more blocks will be retransmitted and the system efficiency will be poorer. While this result is obvious, it is still instructive to see the difference.

14

In contrast to the hybrid and retransmission approaches where, for a given probability of undetected error, the maximum efficiency is achieved, in forward error correction the efficiency expression reduces to

$$E = \frac{I}{I(1 + R)} = \frac{1}{1 + R} \tag{11}$$

which is a constant. The penalty caused by this apparent constant efficiency is that some percentage of the errors will be neither detected nor corrected. If some errors are acceptable, this trade-off of errors versus code rate may also be acceptable.

SECTION III

PERFORMANCE IN A MEMORY-LESS BINARY SYMMETRIC CHANNEL (BSC)

In this section, the efficiency relation previously presented [Equation (1)] will be evaluated in the binary symmetric memory-less channel.

As a brief review, a binary symmetric memory-less channel is a two-state channel in which the probability of a transition from one state to the other is given as p_1, and the probability of remaining in the same state is $1-p_1$, independent of past results. Thus, the probability of an error is p_1, independent of the nature of the input to the channel. As an example, if the message 101011 was transmitted, the probability is p_1 that any bit may be inverted independent of what happened to any other bit.

Treating the hybrid system first and starting with the efficiency expression

$$E = \frac{I}{I(1 + R) + I_R(1 + R) + D_T},$$

it is desired to find E versus I. If I, D_T, and R are given, it is first necessary to find $I_R(1 + R)$ which equals the rejected messages. The probability of rejecting a message is equal to the number of rejected messages divided by the total number of messages transmitted. A message will only be rejected and thus re-transmitted if a symbol error is detected in it. Error detection operations are performed after all possible corrections. Thus, the probability of rejecting a message is numerically identical to the probability of detecting a symbol error (if no error is detected, the message is either properly corrected or erroneously corrected). If the same assumptions are made as were made in the calculation of P {undetected block error} for the code, it is found that

$$P\{\text{detection}\} = P\left\{\begin{array}{l}\text{the block is not erroneously corrected and} \\ \text{an un-correctable error pattern exists}\end{array}\right\} \quad (12)$$

The calculation is best initiated by finding

P {at least e + 1 symbol errors in a block}.

$$P\left\{\text{at least } e+1 \text{ symbol errors in a block}\right\} = \sum_{i=e+1}^{n} P\left\{i \text{ symbol} \quad (13) \right.$$

$$\left. \text{errors in a block}\right\}$$

$$\sum_{i=e+1}^{n} P\left\{i \text{ symbol errors in a block}\right\} = 1 - \sum_{i=0}^{e} P\left\{i \text{ symbol errors} \quad (14) \right.$$

$$\left. \text{in a block}\right\}$$

Given n symbols, the probability that i symbols are in error is $p^i(1-p)^{n-i}$ where p is the probability of a symbol error. These i symbols can be positioned in any of $\binom{n}{i}$ ways.

Thus $P\left\{i \text{ symbol errors in a block}\right\} = \binom{n}{i} p^i (1-p)^{n-i}$ (15)

The probability that a symbol is in error is the probability of at least one bit error in the symbol. Thus

$$p = 1 - (1 - p_1)^{L} \tag{16}$$

Where p_1 is the probability that a bit is in error and L is the length of the symbol in bits.

For the codes selected

$$L = m$$
$$n = 2^m - 1$$

Thus, (following Mitchell [15])

$P\left\{\text{correction failure}\right\}$

$$= 1 - \sum_{i=0}^{n} F(i) \binom{2^m - 1}{i} \left[(1 - (1 - p_1)^m\right]^i \left[(1 - p_1)^m\right]^{(2^m - 1) - i} \tag{17}$$

17

$F(i)$ = Fraction of i symbol error patterns in 2^m-1 code symbols which are correctable.

$$F(i) = 1 \qquad\qquad 0 < i \leq e$$
$$F(i) < 1 \qquad\qquad e < i \leq n$$

For $e < i \leq n$

$$F(i) = \frac{N(i)}{\binom{n}{i}}$$

where $\binom{n}{i}$ is the total number of i symbol error patterns in n symbols. $N(i)$ is the number of such patterns that are correctable. The portion of $\sum_{i=0}^{n} N(i)$ for $i \leq e$ is known (i.e., all patterns). However, only some patterns are correctable for $i > e$ and these patterns are a function of the code, channel, and the decoding scheme [15]. Thus, we shall be conservative and calculate the probability of detection by setting $F(i) = 1$ all $0 < i \leq e$ and $F(i) = 0$ for $e < i \leq n$ (the guaranteed correction). This is a reasonable approach, since the error is small [14, 15].
The first portion of Equation (12) is now available.

$$P\{\text{Detection}\} = 1 - \sum_{i=0}^{e} \binom{2^m-1}{i} [1 - (1-p_1)^m]^i [(1-p_1)^m]^{(2^m-1)-i}$$

$$- P\{\text{undetected error}\} \tag{18}$$

To find the probability of an undetected block error in a channel exhibiting independent errors, we follow the method used by Nesenbergs [2]. Errors will be undetected if some code vector is transformed by the channel into a vector which is less than i symbols from some other code vector. The probability of combinations falling into this class is

$$P = \sum_{i=e+1}^{n} p_n(i) \tag{19}$$

18

where $p_n(i)$ is the probability of i symbol errors in an n symbol sequence.

For independent errors this is

$$P = \sum_{i=e+1}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

$$= \sum_{i=e+1}^{2^m-1} \binom{2^m-1}{i} [1 - (1-p_1)^m]^i \, [(1-p_1)^m]^{(2^m-1)-i} \qquad (20)$$

Using the previously derived expression where an error is always assumed to occur, we find

$$P \{ \text{undetected block error} \} = \frac{(2^{km}-1) \sum_{i=0}^{e} \binom{n}{i} n^i P}{2^{nm} - \sum_{i=0}^{e} \binom{n}{i} n^i} \qquad (21)$$

This value is less than the worst-case bound since, from the binomial theorem, $P < 1$. Thus in a binary symmetric channel, the probability of undetected block errors is less than as given previously.

$$P \{ \text{detection} \} = P \{ \text{retransmission} \} = \frac{\text{messages retransmitted}}{\text{total messages}} \qquad (22)$$

$$P \{ \text{retransmission} \} = \frac{I_R (1 + R)}{I(1 + R) + I_R(1 + R)} = \frac{I_R}{I + I_R} \qquad (23)$$

$$I_R = \frac{I \cdot P \{ \text{retransmission} \}}{I - P \{ \text{retransmission} \}} = \frac{I \cdot P \{ \text{detection} \}}{I - P \{ \text{detection} \}} \qquad (24)$$

19

Using this formula, it is now possible to find E versus I for various values of D_T, p_1, and R (where R implies specific code). This operation has been performed and is presented in Figure 3 for m = 8 and e = 32 for a range of values p_i. D_T is taken as 3.5 sec which corresponds to two blocks at 1200 bits/sec. The performance of the system is a monotonically increasing function of I, except where the probability of correction is zero, such that P {retransmission} is one and the system breaks down in both correction and retransmission. On the curves shown, the efficiency reaches 0.7 out of a maximum value of 0.75 for a 35 sec message where the probability of a bit error is 1×10^{-2}. There is significant degradation for a bit error probability of 1.6×10^{-2}, but this simply means that a lower effective rate is being achieved. This could be improved by going to a code rate less than 0.75 but greater than 0.4, or increasing m.

RETRANSMISSION ERROR CONTROL

The expression for efficiency in retransmission error control is

$$E = \frac{I}{I(1 + R_d) + I_R(1 + R_d) + D_T}$$

If it is assumed that all errors are detected and retransmissions occur for every detection, it is again found that

$$P \{retransmission\} = \frac{message \ retransmitted}{total \ messages}$$

and

$$I_R = \frac{I \cdot P \{retransmission\}}{I - P \{retransmission\}} = \frac{I \cdot P \{detection\}}{I - P \{detection\}}$$

Since all errors are detected

$$P \{detection\} = P \{occurrence \ of \ a \ block \ error\} = 1 - (1 - p_1)^L \quad (25)$$

Thus

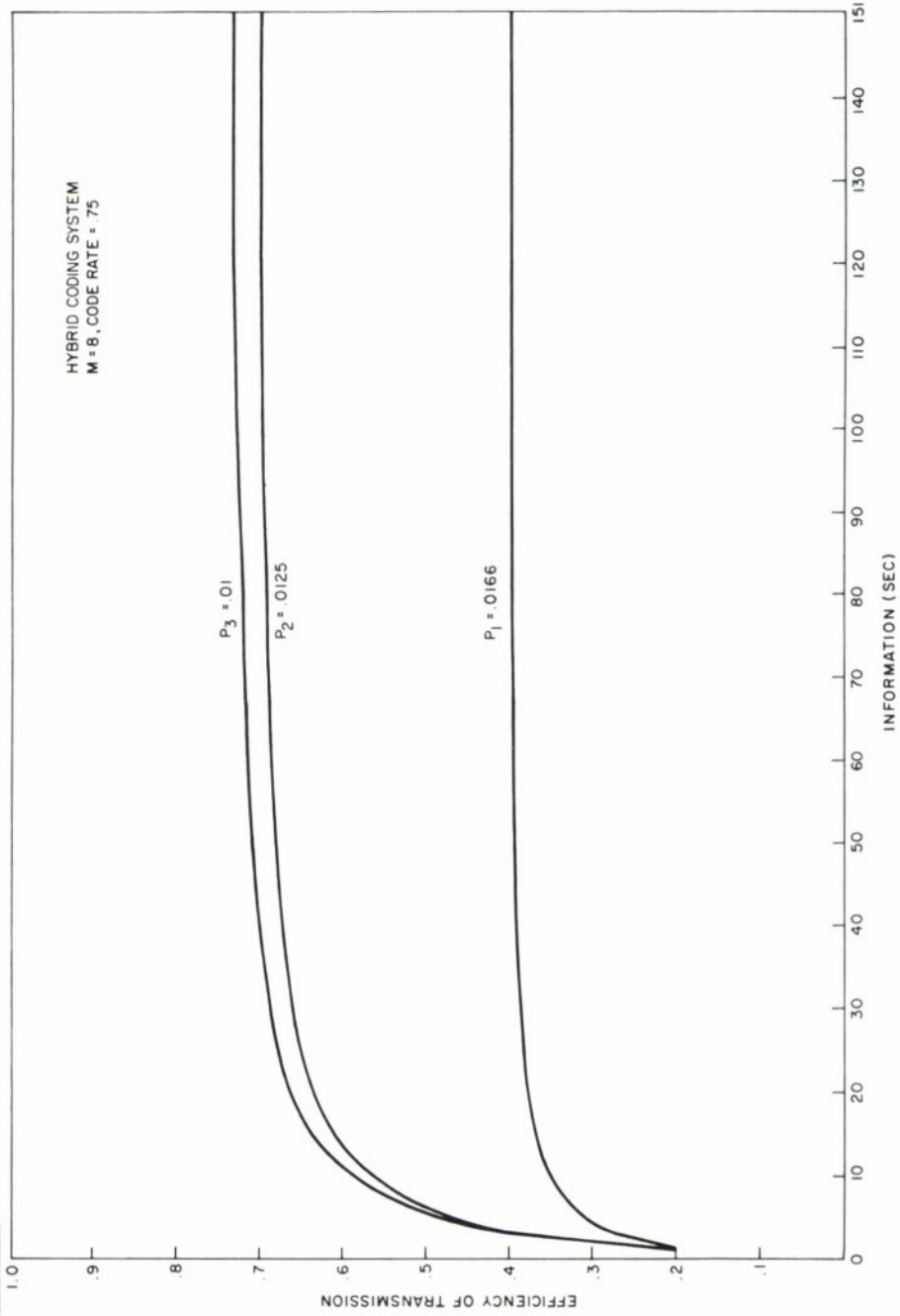$$I_R = \frac{[1 - (1 - p_1)^L] \ I}{(1 - p_1)^L} \quad (26)$$

20

Figure 3. Efficiency versus Information Transmitted for a Hybrid Coding System as a Function of Bit Error Probability

The same set of codes will be used and, after finding the efficiency, the answer should have appended to it the probability of an undetected block error as derived for the BSC. In this case, L is the block length.

$$L = m \ (2^m - 1)$$

D_T was taken as two blocks in hybrid coding but, as previously explained, D_T is only one block in retransmission. From the efficiency expression and the expressions for I_R , it is found that when D_T is negligible with respect to I,

$$E = \frac{1}{(1 + R)\left(1 + \left[1 - \left(1 - p_1\right)^L\right] / \left(1 - p_1\right)^L\right)} = \frac{\left(1 - p_1\right)^L}{1 + R} \tag{27}$$

From this expression it is found that, for a given value of L, as p_1 approaches zero the efficiency approaches the code rate. Similarly for a given p_1 , the efficiency goes to zero with increasing L. These results are presented in Figure 4 along with the results for 104 ms. delay.

FORWARD ERROR CORRECTION

Using forward error correction only in the independent error environment the following relationship is found (assuming guaranteed correction only).

$$P \ \{correction\} \ = \ P \ \{not \ more \ than \ e \ symbol \ errors \ in \ a \ block\}$$

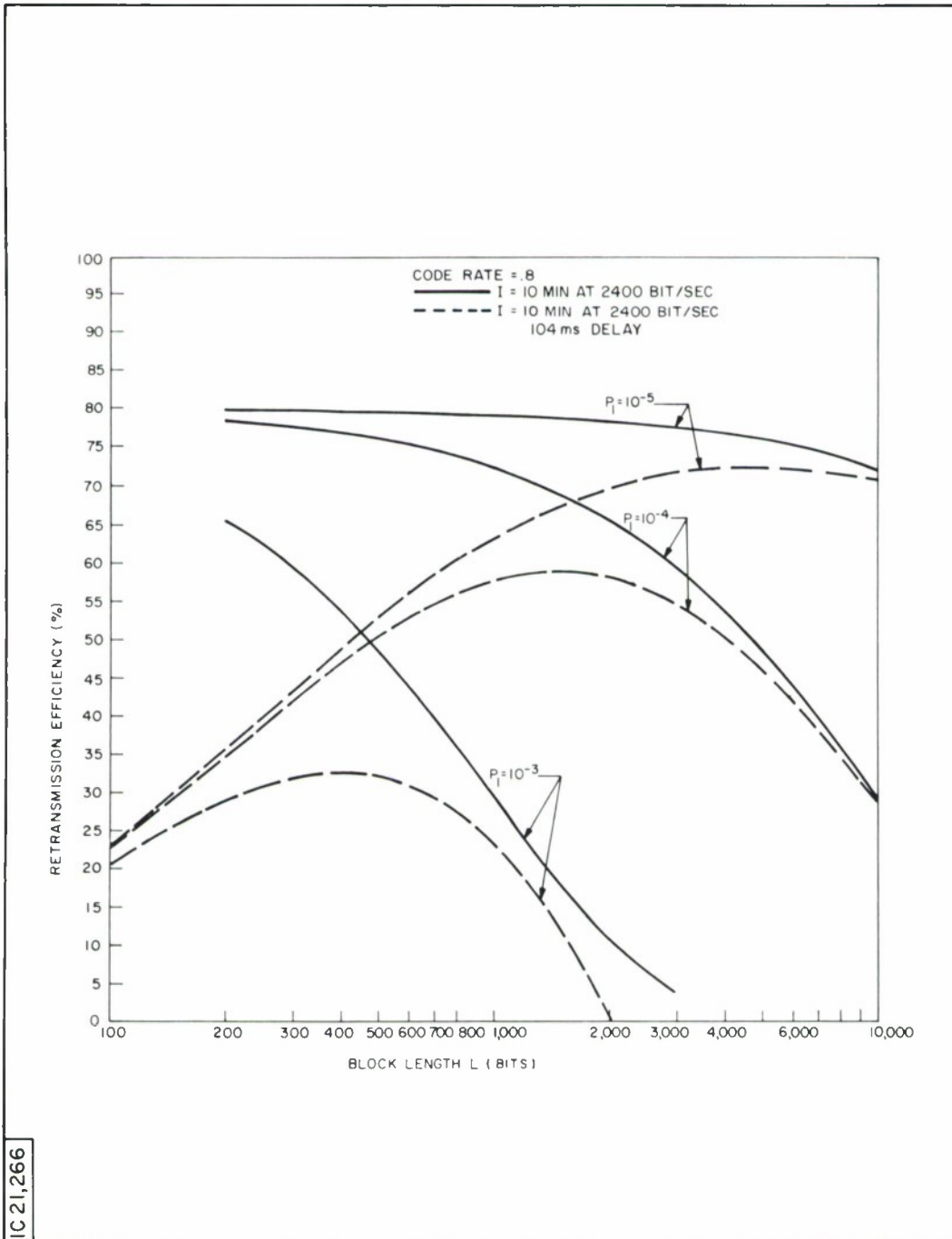$$= \ \sum_{i=0}^{e} \binom{2^m - 1}{i} [1 - (1 - p_1)^m]^i \ [(1 - p_1)^m]^{(2^m - 1) - i} \tag{28}$$

22

Figure 4. Efficiency versus Block Length as a Function of BER in a BSC

23

Using m = 3, 4, 5, 6, 7, 8, and code rates (2e < n), a set of curves can be developed for the percent of errors corrected as a function of channel bit error rate. However, the curves will lie on the vertical axis between 99 and 99.9999 percent of the errors corrected for all values of $p_1 < 10^{-3}$. Therefore, the results will be presented as improvement factor (ratio of input errors to output errors) as a function of symbol error probability. A block is in error after decoding if there are at least e + 1 symbol errors in the block and a false correction is not made. Since this adjustment is not made, the results should be weighted in the light of the undetected error probabilities.

The results in Figure 5 give the improvement factor as a function of symbol error probability.

The curves relating symbol error probability to bit error probability are given in Figure 6. To convert the symbol improvement factors which result from Figure 5 to bit improvement factors, the following relations should be used.

$$\text{symbol improvement factor} = \text{expected number of symbol errors input/number of symbol errors output}$$

$$= \frac{p}{p - \frac{1}{n} \sum_{i=0}^{e} i \binom{n}{i} p^i (1-p)^{n-i}} \tag{29}$$

and

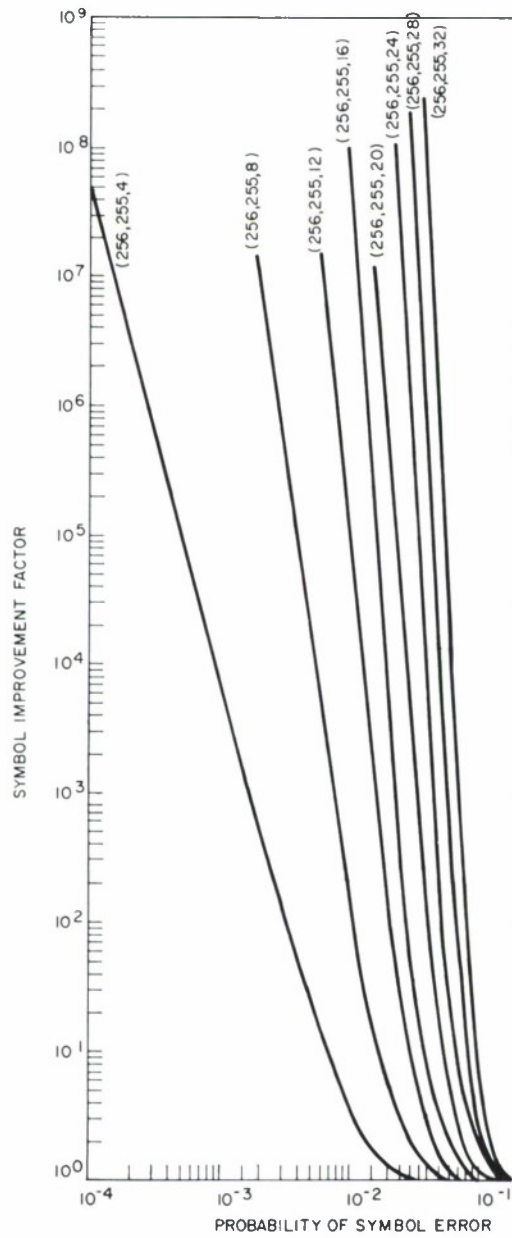$$p = p \text{ symbol} = 1 - (1 - p \text{ bit})^m . \tag{30}$$

24

Figure 5. Symbol Improvement Factor versus Probability of Symbol Error
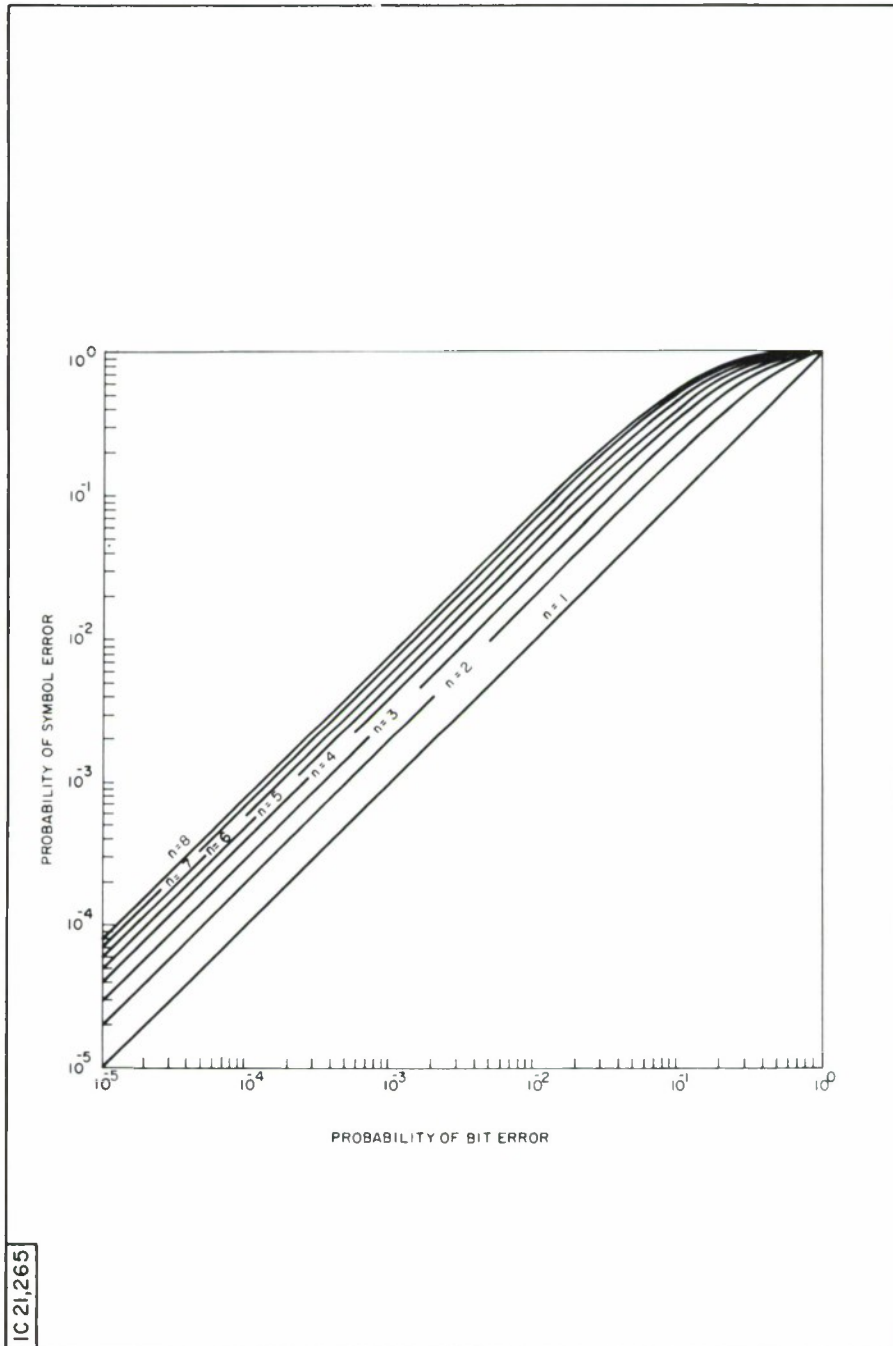
Figure 6. Symbol Error Probability versus Bit Error Probability

SECTION IV

PERFORMANCE IN REAL CHANNELS

In this section, the performance of the three coding systems in a troposcatter environment is examined. The overall distributions of the errors measured in the environment is described in Appendix I along with a statement of the modulation technique. The error data was collected in 90-minute test samples. A system which performs error control functions was simulated on an IBM 7030 computer and was operated for each 90-minute data sample. The computer program used in this analysis performs all the functions of hybrid error control in terms of actual message structure and the decoding and correction of errors. However, since the locations of errors are known in advance, if the program determines that there is an excessive number of symbol errors in a block, no attempt to correct is made and detection is assumed. Thus, undetected errors are not allowed for, and the results must be used in conjunction with the probability of an undetected error to get a true picture.

FORWARD ERROR CONTROL

The function of forward error correction was performed by taking all data at a given data rate and correcting as a whole. Thus the results presented in Figures 7 and 8 indicate the percent of errors corrected as a function of code rate $\frac{(n-2e)}{n}$ for all $(2e<n)$ acceptable code rates for the total mass of data at the given transmission rate. The results presented are for $m = 2, 3, 4,$ 5, 6, 7, and 8 for both the bit error correction and block error correction. The power of the code is seen to increase as a function of m. Thus it is possible to reduce the redundancy necessary for a given amount of error correction by increasing m. The burst structure of the data indicated in the data description is even more evident in terms of the code correction curves. If for the
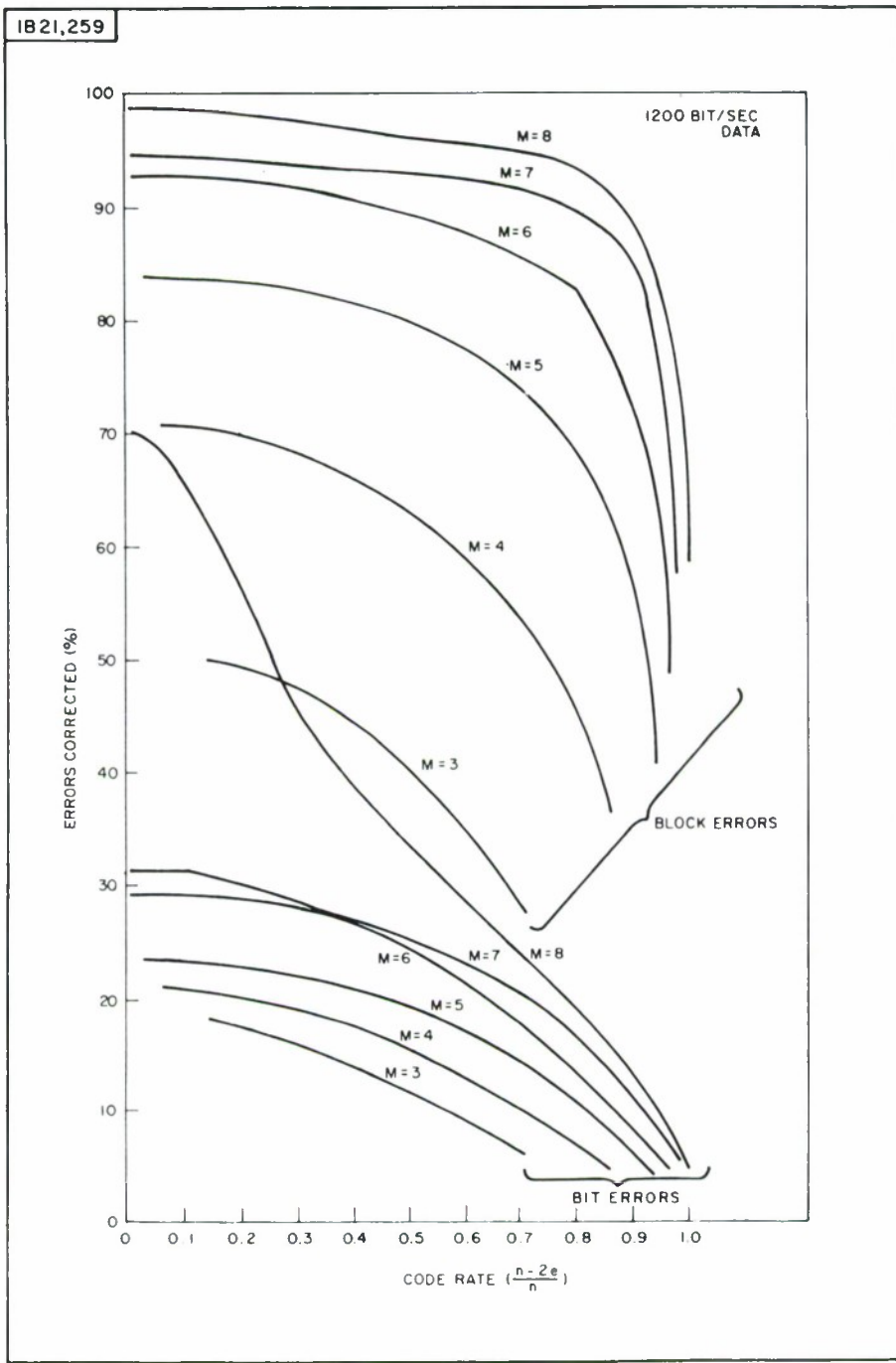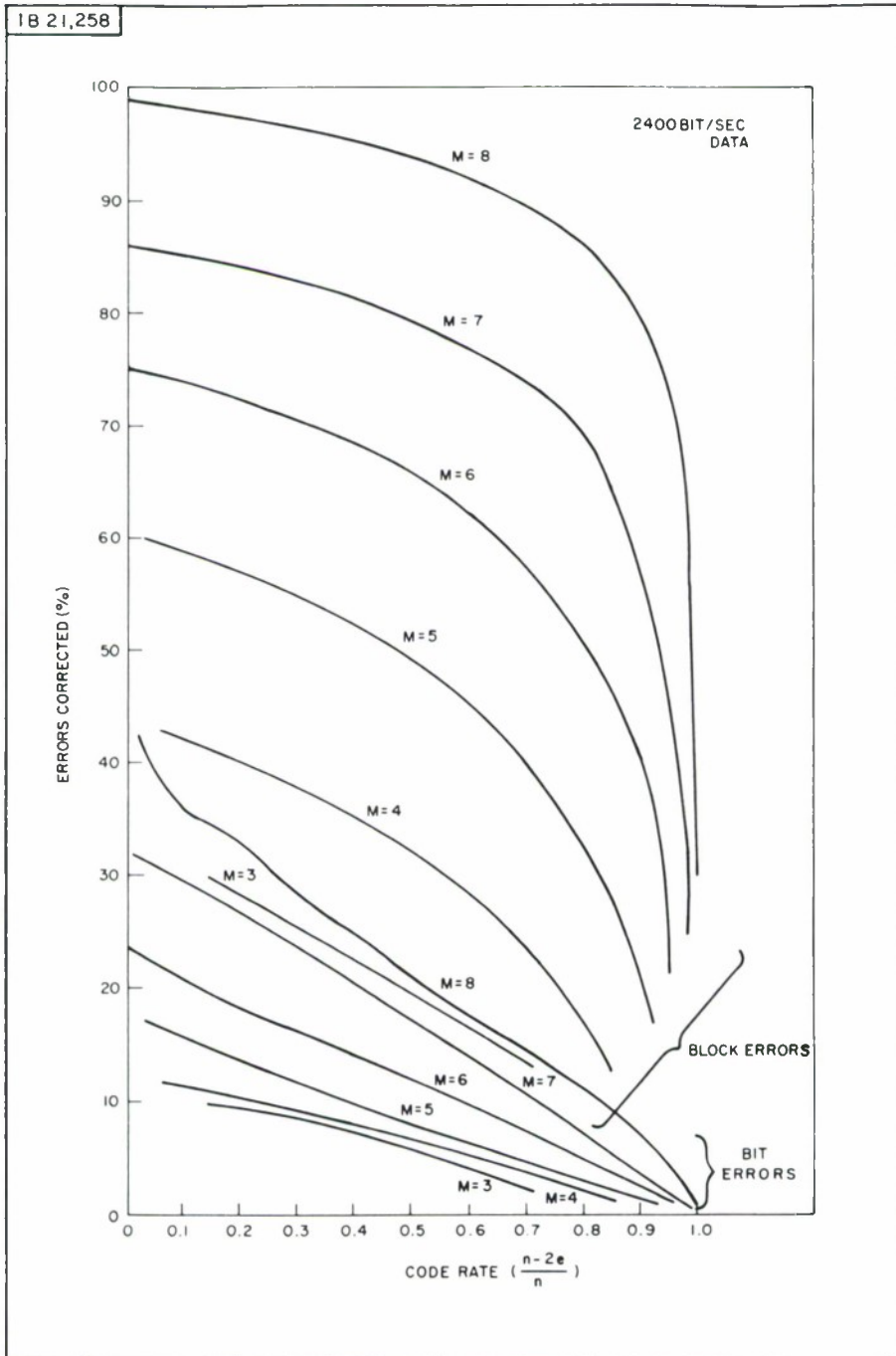
Figure 7. Percentage of Errors Corrected versus Code Rate

28

Figure 8.   Percentage of Errors Corrected versus Code Rate

29

2400 bit/sec data we consider the curve for m = 4 at a code rate of 0.2, 70 percent of the blocks in error are corrected but only 20 percent of the bit errors are corrected. Thus at this block size of 60 bits, 30 percent of the blocks with errors contain 80 percent of the errors. As m changes in value, the block structure changes, and the fact that the code for m = 7 did poorer than that for m = 6 in a portion at the range of codes for the 1200 bit/sec data should not be surprising. At both data rates, the pattern is broken for m = 8. Here the code block becomes substantially longer than the bursts and with respect to the code the data appears as random bursts, which is the mode of greatest efficiency. The curves indicate that the codes do poorly as forward bit error correcting codes. However, the codes perform excellently as message error correcting codes and can be used in the hybrid system since error correction significantly reduces the number of retransmissions necessary in a hybrid system. The 2400 bit/second data has bursts which are of approximately the same length as the 1200 bit/sec data but are more than twice as dense (see Appendix I). For the small values of m, this difference in density shows up in error correction performance; for large values of m the differences is not so great. However, it should be noted that for a 2040 bit block (m=8), 57 percent of the errors are contained in one percent of the blocks. Based on these results and the excellent undetected error performance for m=8, this value has been chosen for use in error detection for the retransmission system and correction and detection in the hybrid system.
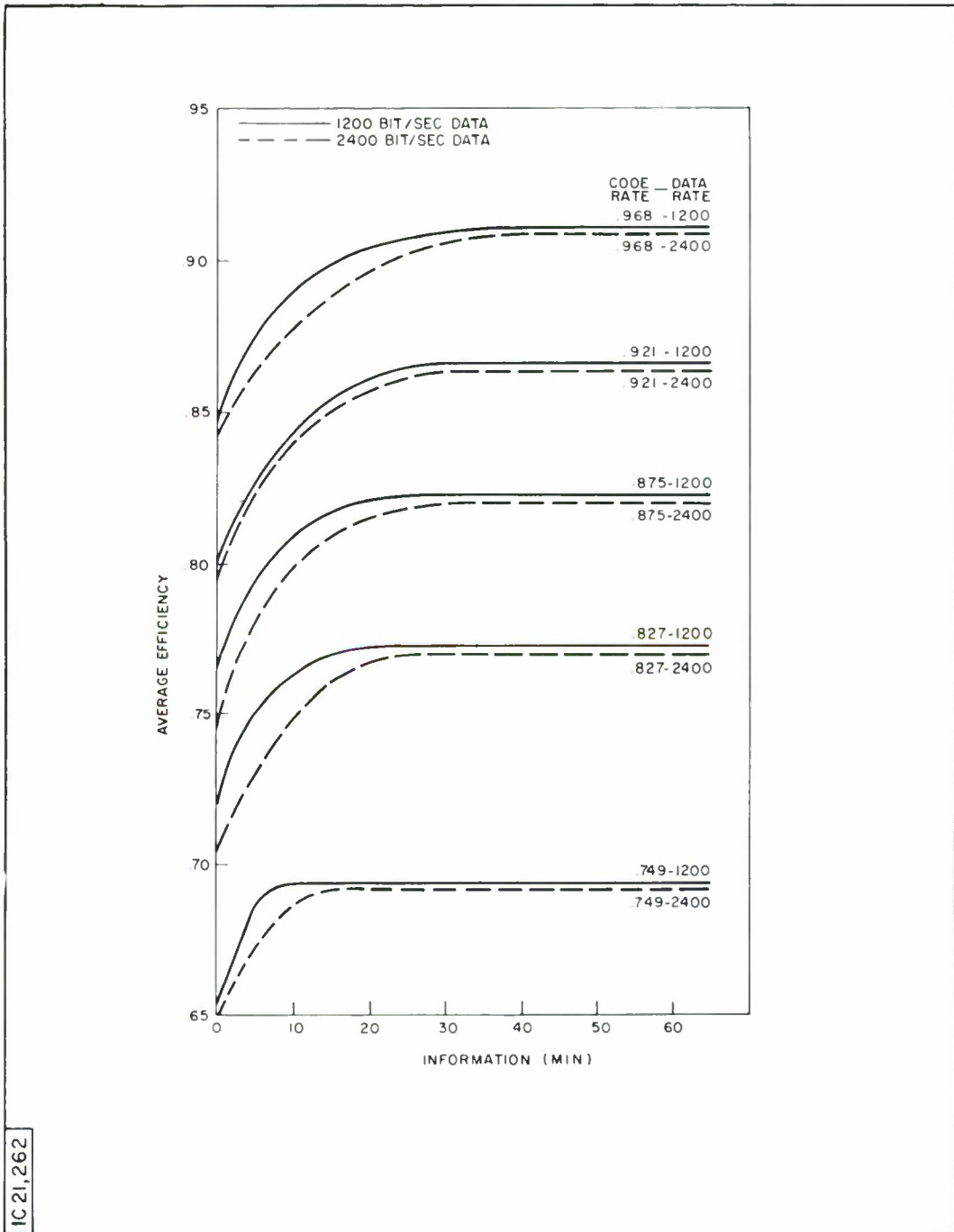
RETRANSMISSION ERROR CONTROL

The performance of retransmission error control in the real channel is presented in two ways. The first is efficiency versus information transferred, and the second is efficiency versus block length.

The results for efficiency versus information transferred were obtained by attempting to transfer the given amount of information through each 90-minute test sample and calculating the efficiency for each test sample. The

results are the average across all the samples to obtain the average efficiency, and are used to predict the expected efficiency of transfer of large masses of data by retransmission error control only. The results are presented in Figures 9 and 10. The block lengths of 2040 bits and the redundancy levels have been chosen so that there can be easy comparison with the hybrid system which will use the most powerful correction (m=8). In retransmission control, the symbol codes are used for detection only, as previously described, and also the results must be used in combination with the undetected error probabilities of the code. The return link was coded with m = 10 using one information symbol.

For comparison with the efficiency versus block length results presented for the BSC, a similar set of results have been obtained for the real channel for 10 minutes of information (Figures 11 and 12).

Figure 9. Average Efficiency of Retransmission System on Troposcatter Data

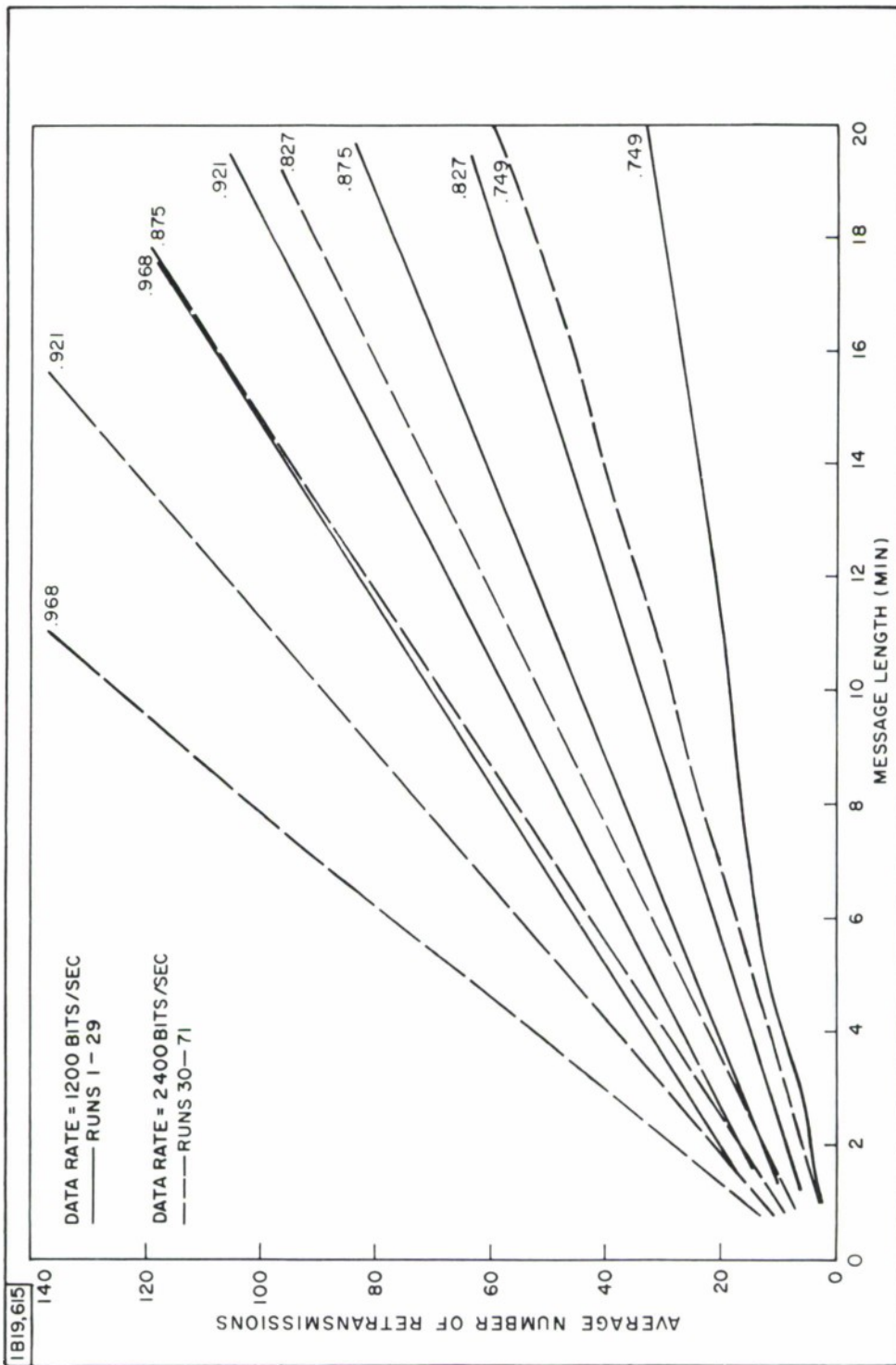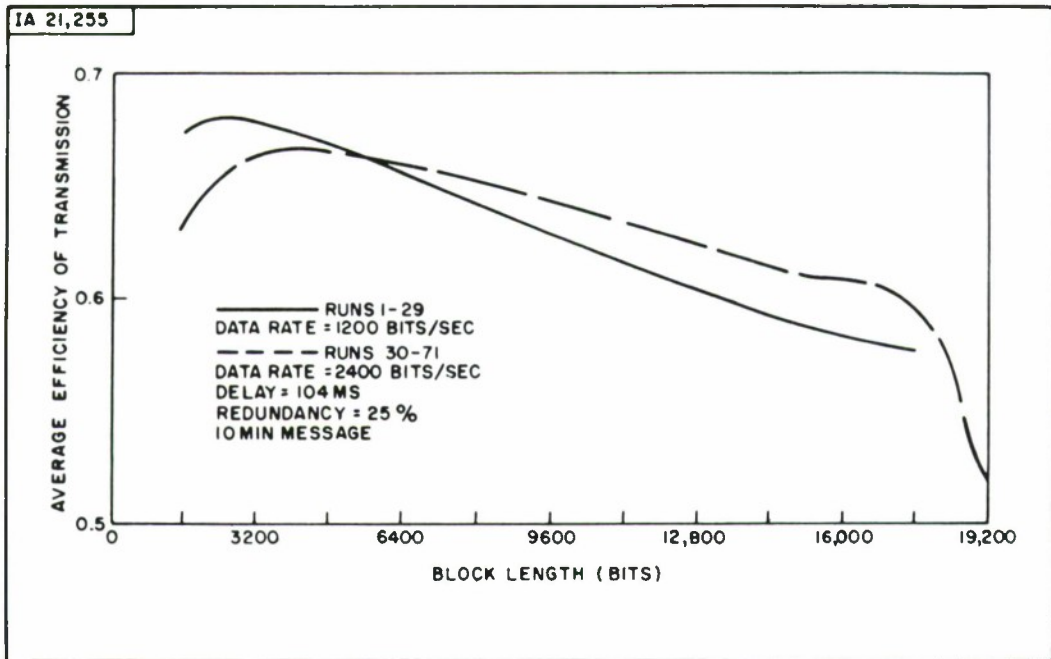Figure 10. Number of Retransmissions Required
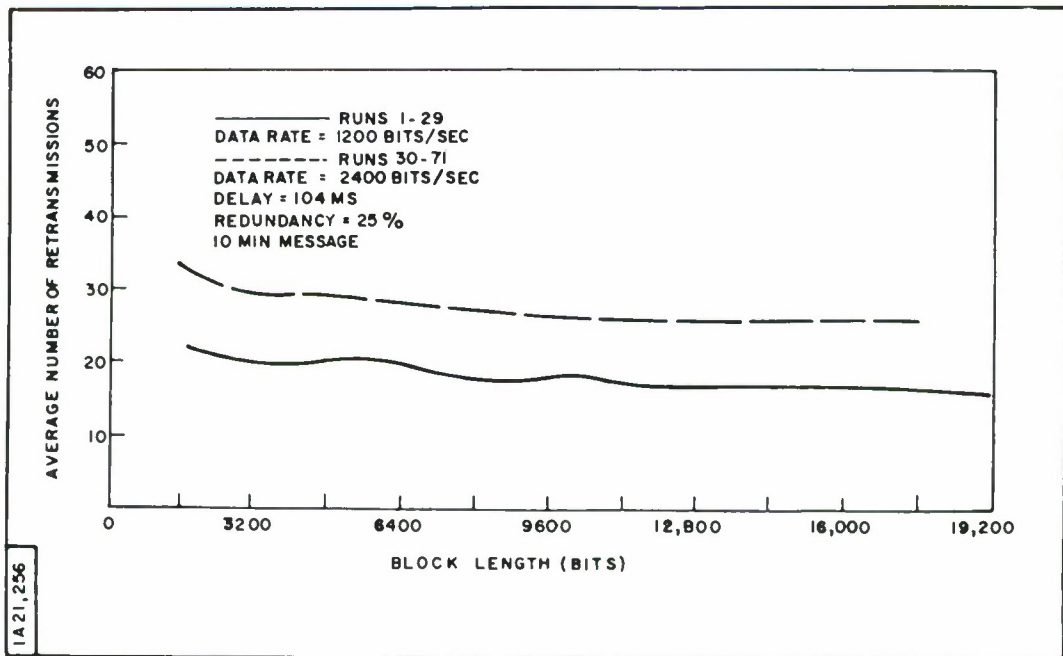
Figure II. Fixed Block Efficiency

Figure 12. Fixed Block Retransmission Performance

34

HYBRID ERROR CONTROL

Using a combination of retransmission error control and error correction (m=8), the efficiency in the hybrid mode as a function of information transmitted has been obtained (Figures 13, 14, and 15). Figure 13 indicates the average efficiency as a function of I for various code rates and data speeds in the tropo-scatter channel. The curves indicate both the maximum effective rate which can be achieved for a value of I and the data rate as a function of the acceptable probability of undetected error. As before, the curves must be read in conjunction with Figures 1 and 2. For code rates other than those used, interpolation can be performed. Figure 14 shows the percent of blocks in error which are corrected by the error correcting code. The dip starting at five minutes in the 2400 bit/sec curves is due to an abnormal burst of three million bits duration in one of the data runs. At first it was considered that this burst be eliminated from the data sample. However, on the basis of performance of the system, this burst was left in. The code failed to correct the errors but the retransmission part of the system compensated, with the result that the efficiency curves show no degradation for the range I=5 to I=20. The average number of retransmissions is presented in Figure 15. The previously mentioned low rate coding with m = 10 was used on the return link. The important conclusion from this data is that in the actual channel (as in the BSC) it is possible to operate at an effective rate arbitrarily close to the code rate and transfer large masses of data for a given, acceptable probability of error.
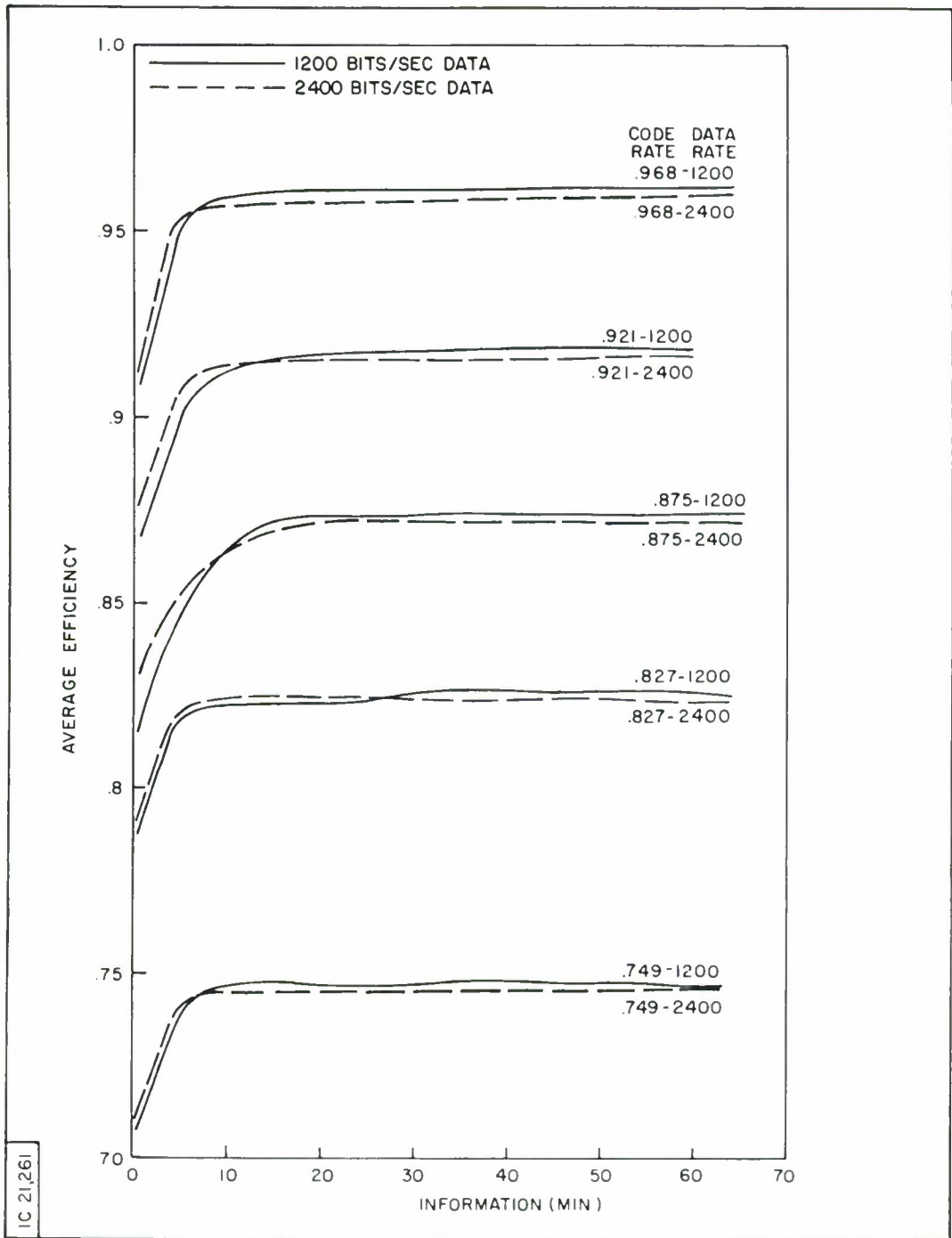
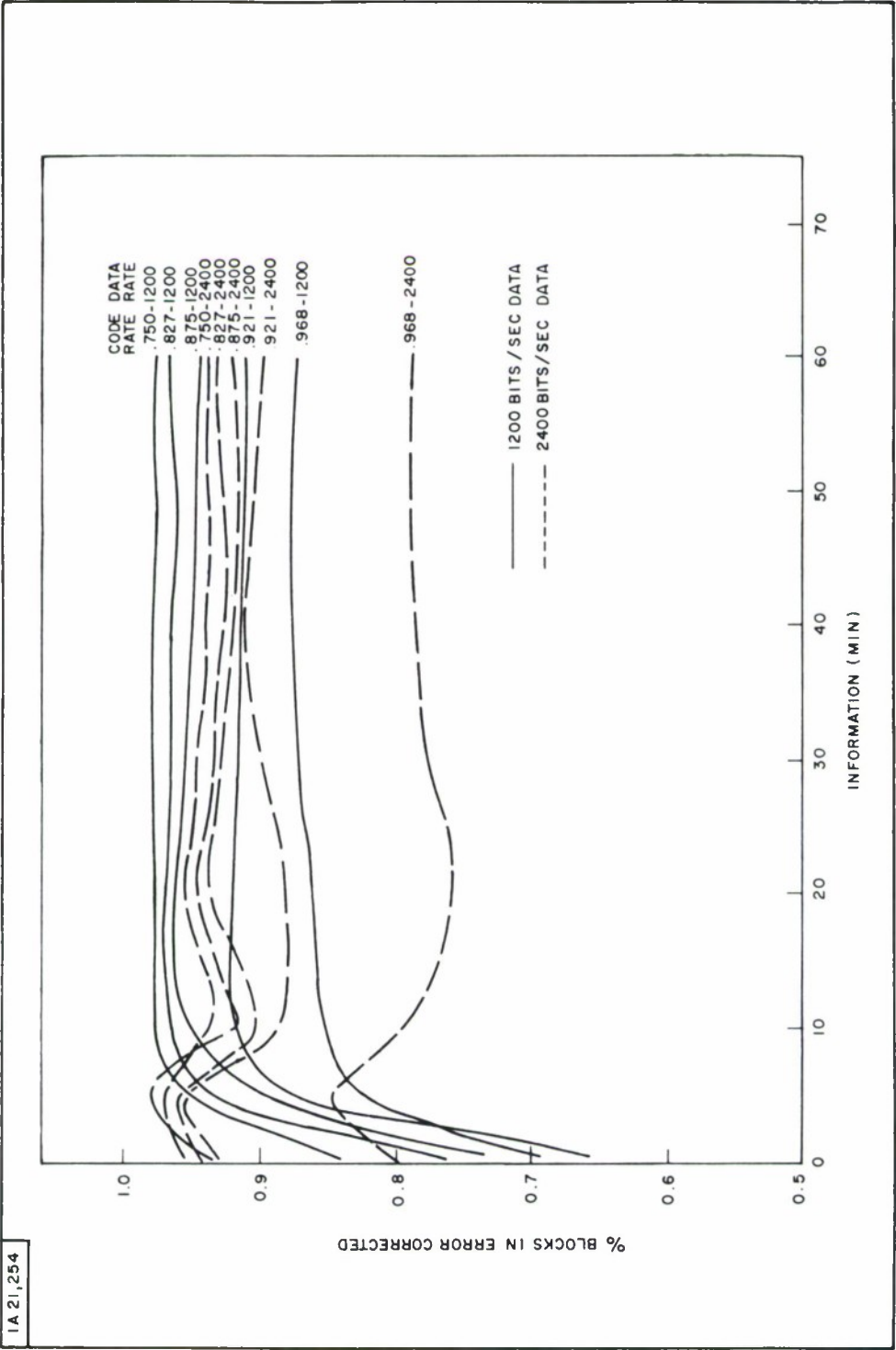Figure 13. Average Efficiency of Hybrid System on Troposcatter Data

Figure 14. Percentage of Blocks Corrected by P^M Code in Hybrid System on Troposcatter Data
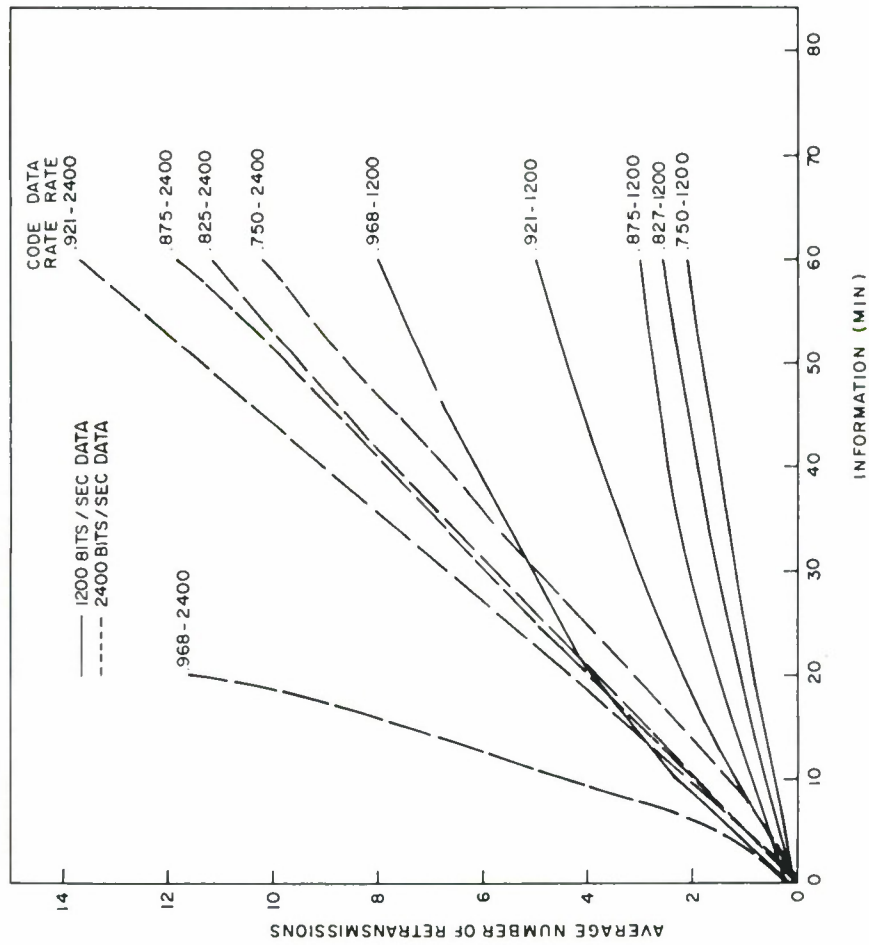
Figure 15. Average Number of Retransmissions by Hybrid System on Troposcatter Data

SECTION V
CONCLUSIONS

A method has been developed to provide combined error correction and
retransmission in such a manner that the highest possible rate of data transfer
can be achieved for a given set of conditions. The performance of the system
in a binary symmetric memory-less channel and a real channel has been de-
scribed. It is demonstrated that, for a sufficiently long message, it is possible
to operate at an efficiency that differs only slightly from the code rate for some
value of acceptable error probability. Unlike many other systems, this
system can be realized with present-state-of-the-art hardware and should
prove useful in providing reliable digital communication between any two
remote geographical locations.

APPENDIX I

In the spring of 1966, data transmission tests were conducted by The MITRE Corporation on a U. S. Air Force data circuit (see Figure 16). The link is characterized as having three types of transmission media: troposcatter, microwave, and wireline. The troposcatter is multiple-hop, and the wireline consists of the interconnection of numerous leased telephone wirelines. The dominant sub-path is a troposcatter-hop of over 500 miles. During the tests, Rixon Sebit 24B vestigial sideband AM modems were used at data rates of 1200 and 2400 bit/sec in a 3 kHz channel which was FM multiplexed with other channels in transmission. A total of 38 hours of 1200 bit/sec and 61 hours of 2400 bit/sec data was collected for a total of 1.67×10^8 and 5.20×10^8 bits respectively, in 90-minute samples.

The error pattern data was obtained by one-way transmission of a 52-bit digital data test pattern which was compared with a locally-generated test pattern at the receiving end. When there was agreement between the two, a logical zeros were declared and when the bits were different, a logical one was declared. The zeros and ones were recorded on magnetic tape. These magnetic tapes were then processed through a tape-to-tape converter facility to produce an IBM-compatible tape for processing in an IBM 7030 computer in order to determine the statistics of these bit-by-bit error patterns.

The data for the two transmission rates was processed for four types of statistics: the occurrence of consecutive errors, the occurence of intervals between errors, the occurrence of bursts, and word error rate information. The extent of the data analyzed was as follows:

Table 1-1

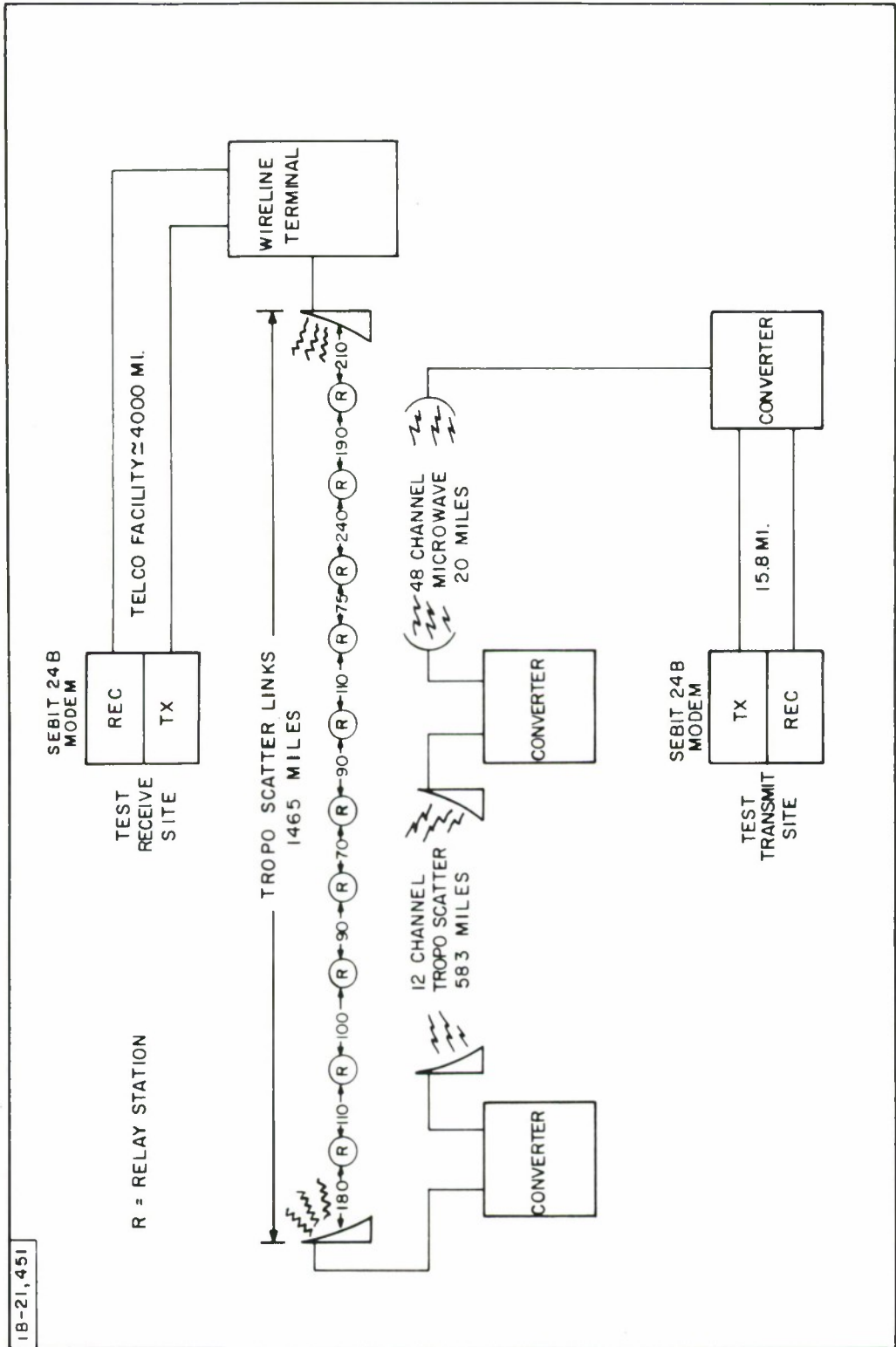| Data Rate, bits/sec	Modem	Total Bits	Average Error Rate (p_e)
1200	Sebit 24B	1.67×10^8	1.39×10^{-4}
2400	Sebit 24B	5.2×10^8	3.95×10^{-4}

41

Figure 16. Full Duplex Communication Circuit Configuration

The relative frequency of consecutive errors is presented in Figure 17 for the two data rates. The Figure indicates that the relative occurrence of multiple errors for the Troposcatter data is independent of the data rate for these two rates. Further, the relative occurrence is certainly not of an independent random nature since, if it were, over 98 percent of the errors would occur as single errors while in fact only 70 percent occur as single errors.

This lack of independence is further illustrated in Figure 18 where the cummulative frequency at occurrence of error-free gaps is presented.

All these curves confirm that this error pattern data is widely different from independent random errors at the same error rate. The cummulative distribution of error-free gaps for independent random curves is (in theory) described by

$$P\left\{ {}_c^n e \middle| e \right\} = \sum_{k=0}^{n} p(1 - p)^k$$

where

      c represents a correct bit

      e represents an error bit

      n represents number of consecutive bits

The message error rate (defined as the occurrence of at least one error in a message) as a function of message size is presented in Figure 19 for these two classes of data.

It is interesting to note that the value of p_m for n = 2000 in the actual data is independent of error rate and signaling speed. This again indicates that the errors are occurring in bursts and, while for the overall data p_e at 2400 bits/sec is 2.85 times are large as p_e at 1200 bits/sec, the additional errors which are occurring at the higher signaling speed occur as increased bursts density. Thus it is evident that bursts occur as fades in time unrelated to the data rate (for the rates measured). This conclusion is confirmed by a visual observation of the error patterns.
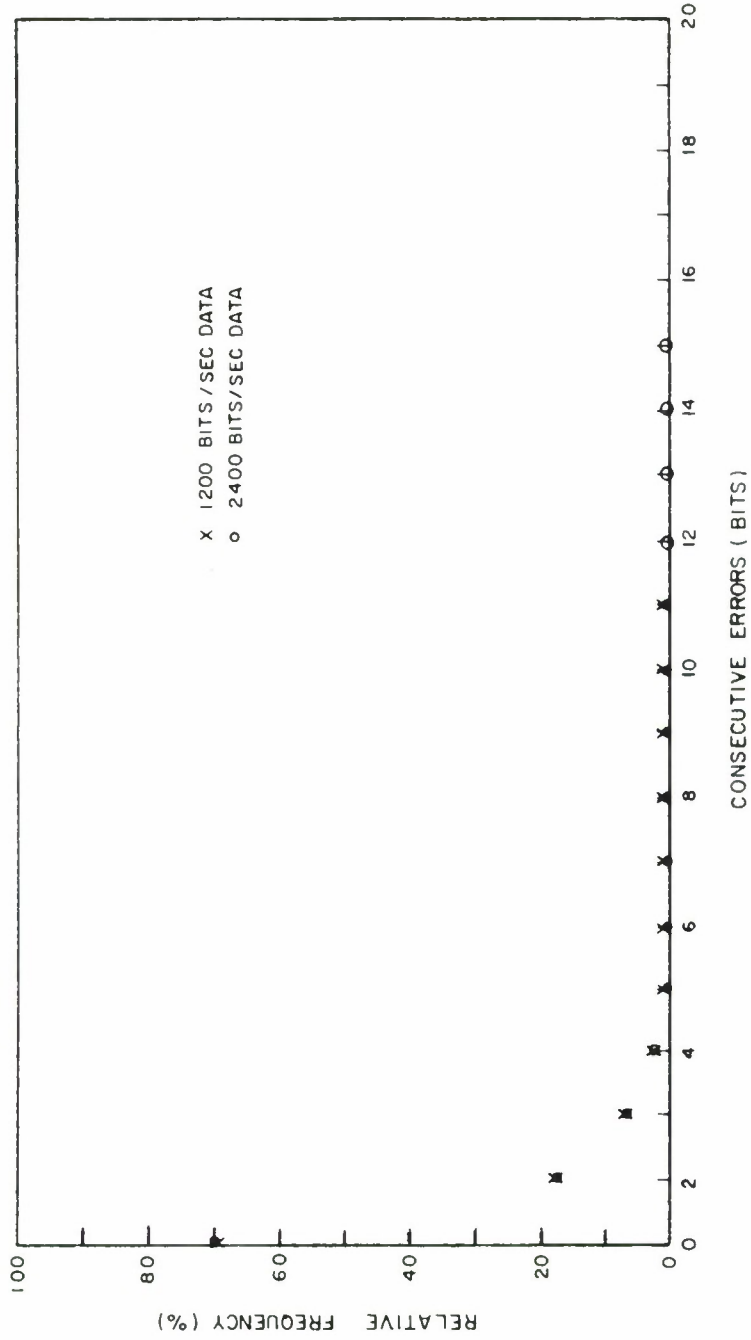
43

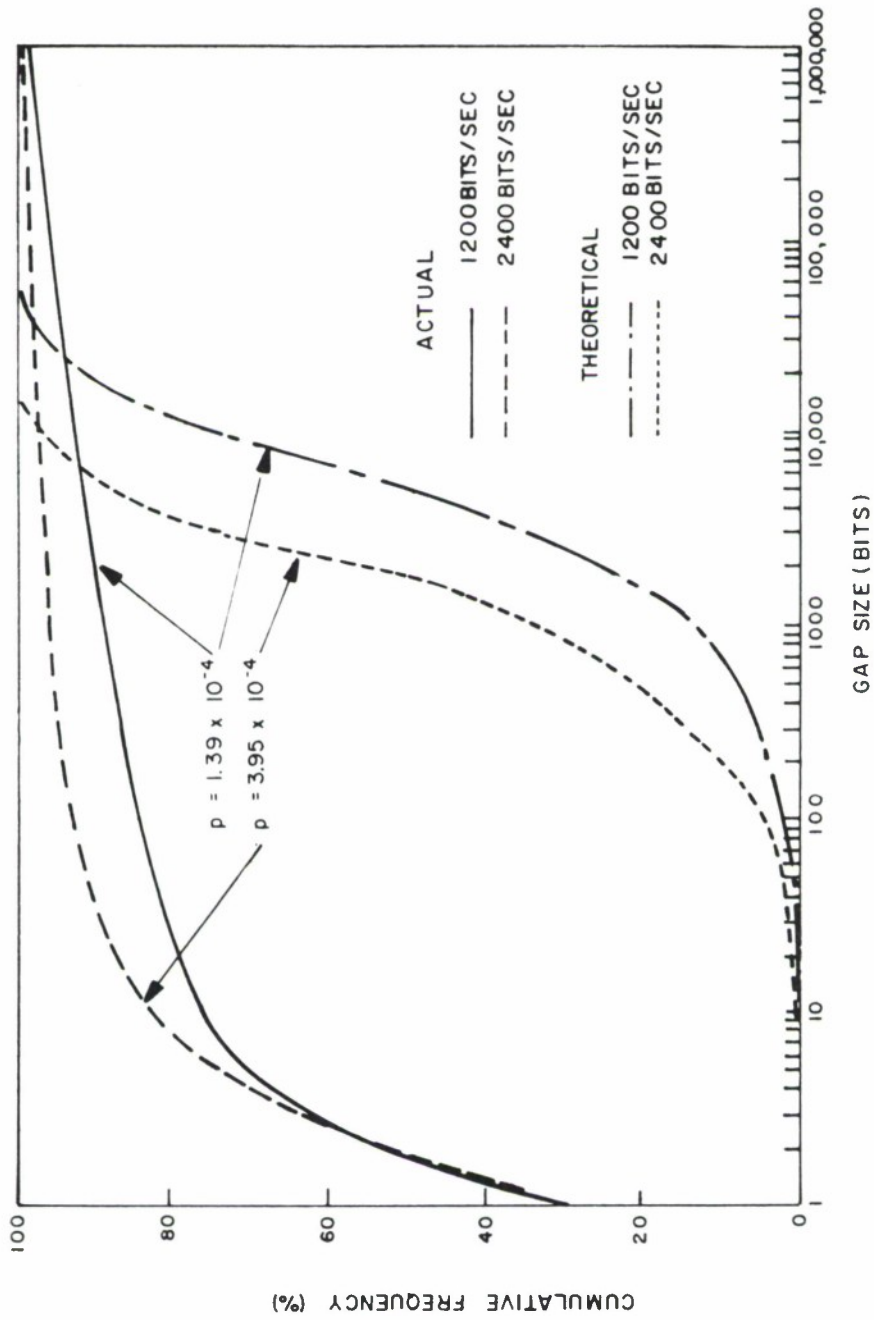Figure 17. Frequency of Consecutive Error Occurrence

44

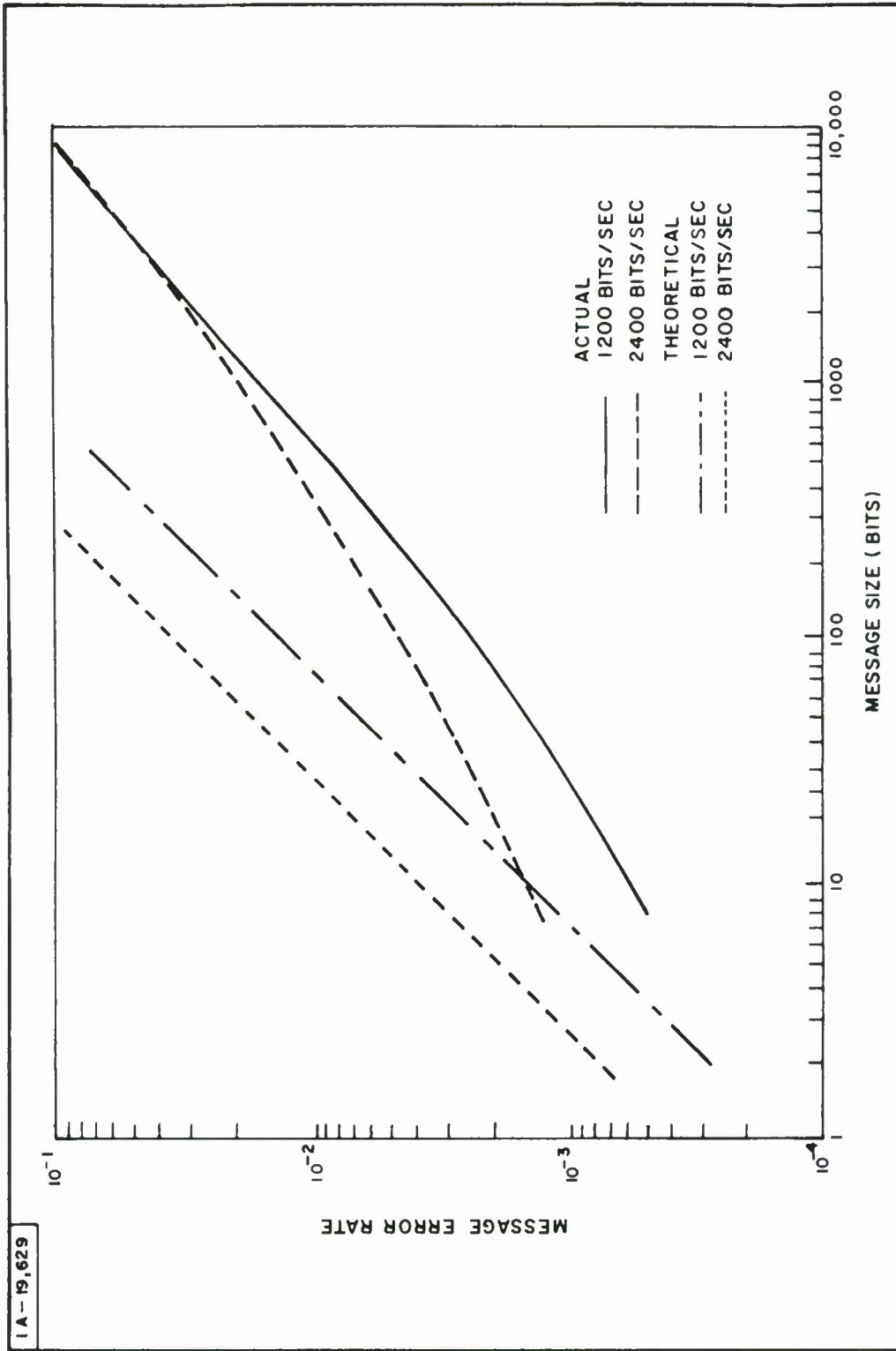Figure 18. Distribution of Gaps Between Errors

IA-20,185

45

Figure 19. Message Error Rate versus Message Size

IA — 19,629

46

The description of the error statistics presented thus far have made it clear that the errors fall in a non-random clump distribution. To describe these error bursts and the intervals between bursts (guard space) an additional statistic shall be used.

DEFINITION

A burst in defined as a region of the serial data stream where the following properties hold. A minimum number of errors, M_e, are contained in the region and the minimum density of errors in the region is Δ. Both of these conditions must be satisfied for the chosen values of M_e and Δ for the region to be defined as a burst. The density of errors is defined as the ratio of bits in error to the total number of bits in the region.

The following properties hold for the bursts. The burst always begins with a bit in error and ends with a bit in error. A burst may contain correct bits. Each burst is immediately preceded and followed by an interval in which the density of errors is less than Δ.

The burst probability density function is defined as the probability of occurrence of a burst of size N, where N is any positive integer. The burst size is measured in terms of the total number of bits in the burst. A separate burst probability density function may be determined for each pair of values of Δ and M_e.

The minimum number of errors in a burst has been chosen to be 2 for all the data included here.

The interval is defined as the region of the serial data stream where the following properties hold. The minimum density of errors in less than Δ, and the region begins and ends in a correct bit. An interval may contain errors. An interval is always immediately preceded and followed by a burst. Thus, each and every bit in the data stream must lie in either a burst region or an interval region.

47

The interval probability density function is defined as the probability of occurrence of an interval of length L, where L is any positive integer. The interval probability density is a joint function of both Δ and M_e.

In Figure 20, the distribution of observed burst lengths is presented for the data. The burst lengths range from 2 (M_e) to over 1,000,000. While 90 percent are less than 1000 bits long, the remaining bursts contain most of the errors. The densities of these bursts are displayed in Figure 21. This Figure confirms the previous conclusions that at 2400 bits/sec the additional errors serve to increase the density of the bursts. Further, while the minimum criteria (Δ) is chosen at 0.01, over 50 percent of the bursts have densities greater than 0.1. A correlation was performed on the burst lengths and their associated densities and it was found that while bursts less than 10 bits in length are generally 100 percent dense (consecutive errors) and bursts over 1,000,000 bits in length are near the minimum criteria, the other bursts range in density between the minimum and 75 percent, indicating that there is no correlation between the length of bursts and densities. The value of Δ was chosen such that values of burst lengths and density are independent of Δ.

The distribution seen on these curves indicates that the errors fall into adjacent clusters of errors with long error-free intervals between clusters. There are no intermediate grouping of errors. In a theoretical, random error distribution, the intermediate groupings would predominate. Figures 22 and 23 show the distribution of interval lengths and associated densities. The important facts to note here are that over 60 percent of the intervals are error free and that the remaining intervals have no more than one or two random errors within them.
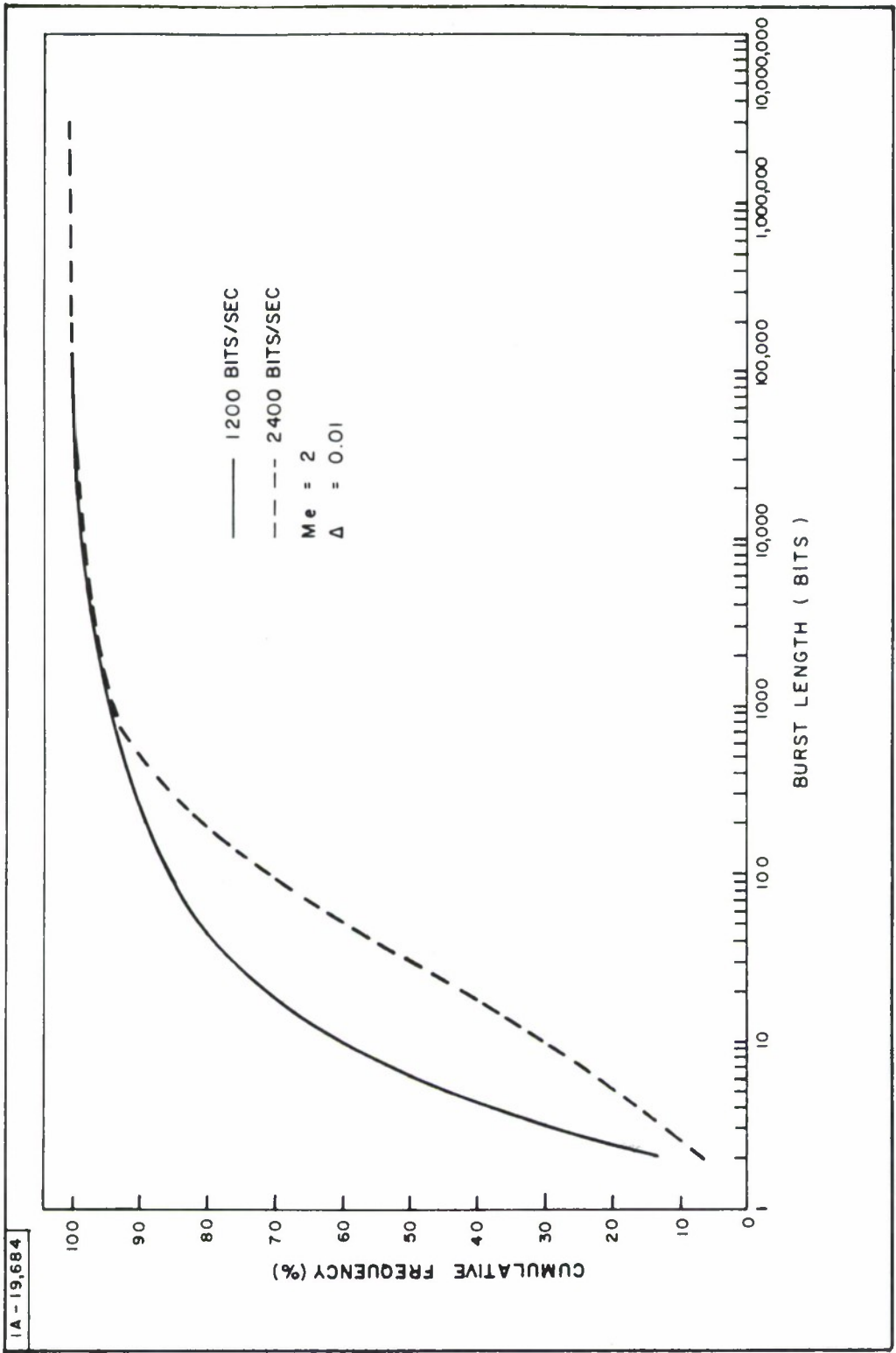
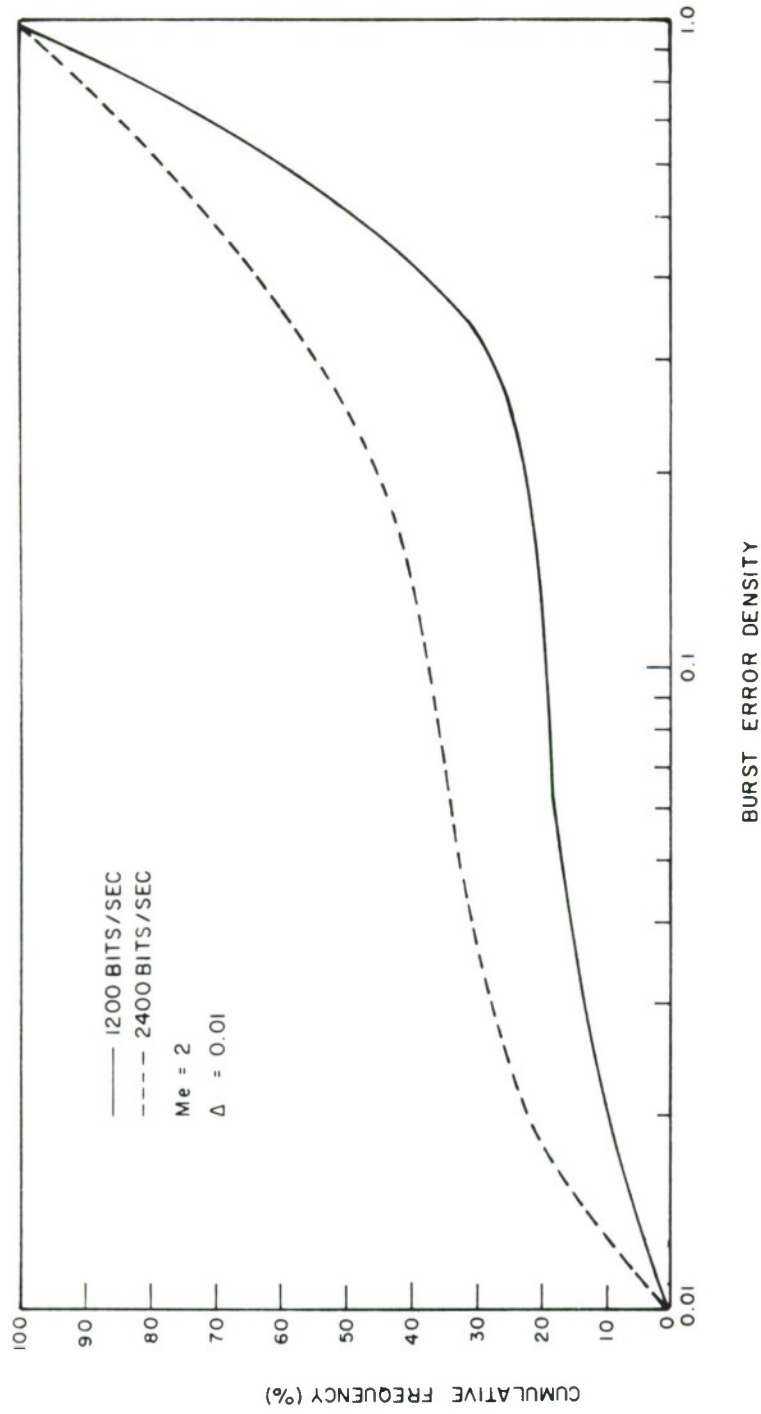Figure 20. Distribution of Lengths of Bursts
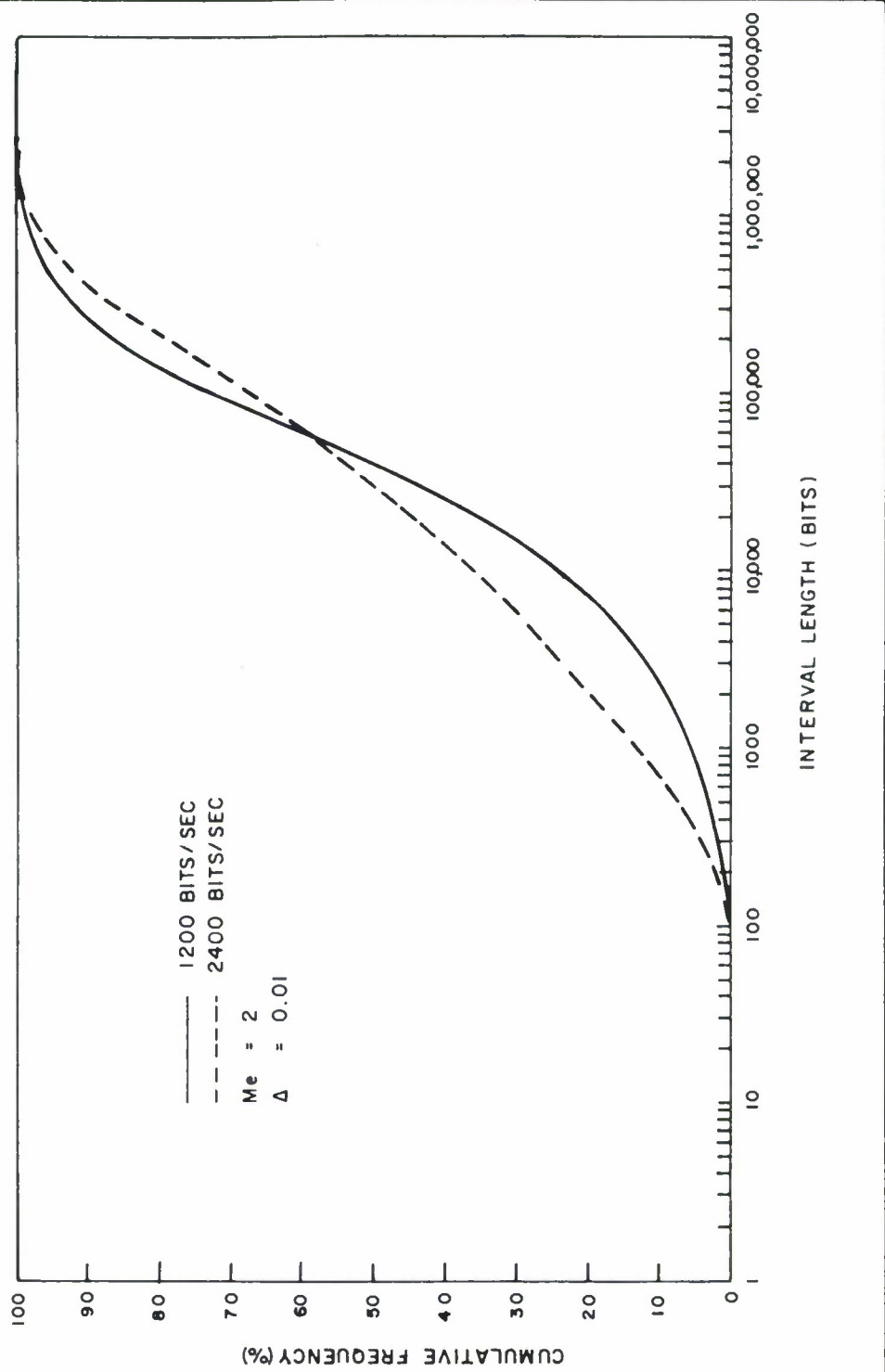
Figure 2l. Distribution of Burst Error Densities

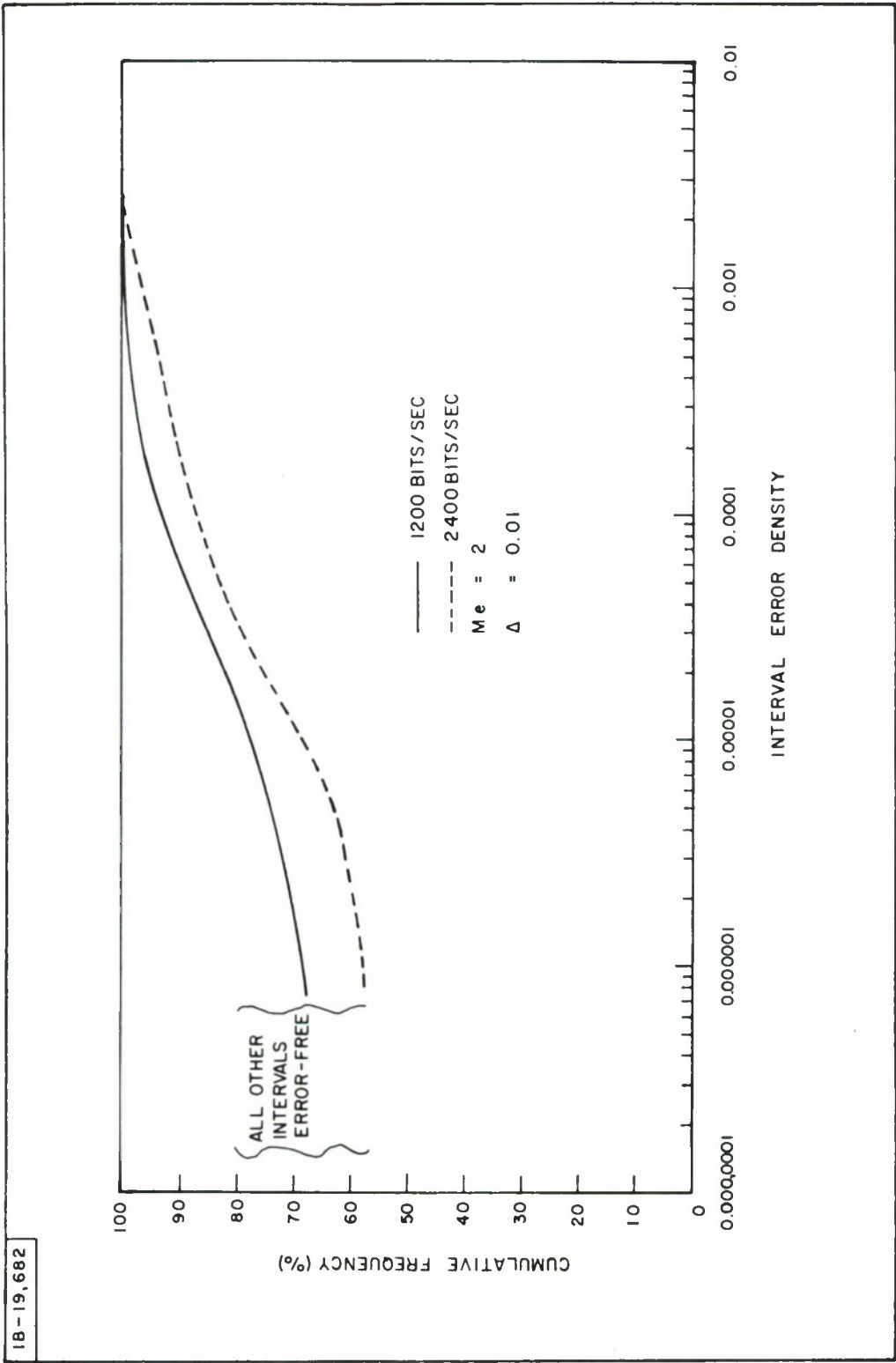Figure 22. Distribution of Lengths of Intervals

Figure 23. Distribution of Interval Error Densities

REFERENCES

1. T. Kuhn, Retransmission Error Control, <u>IEEE Trans. on Comm. Systems,</u> June 1963.

2. M. Nesenbergs, Comparison of the 3-out-of-7 ARQ with Bose-Chaudhuri-Hocquenghem Coding Systems, <u>IEEE Trans. on Comm. Systems,</u> June 1964.

3. K. Brayer, Error Statistics of a Multiple Link Data Transmission Circuit, The MITRE Corporation, MTP-40, Contract AF 19(628)-5165, Bedford, Mass., Nov. 1966; also available as ESD-TR-66-337.

4. K. Brayer, Error Patterns Measured on Transequatorial HF Communication Links, The MITRE Corporation, MTP-46, Contract AF 19(628)-5165, Bedford, Mass., Jan. 1965; also available as ESD-TR-67-99.

5. K. Brayer and O. Cardinale, HF Channel Data Error Statistics, Vol. I, The MITRE Corporation, MTR-171, ESD-TR-66-290, Contract AF19(628)-5165, Bedford, Mass., April 1966.

6. K. Brayer and O. Cardinale, HF Channel Data Error Statistics, Vol. II, The MITRE Corporation, MTR-172, ESD-TR-66-107, Contract AF 19(628)-5165, Bedford, Mass., April 1966.

7. D. Gorenstein and N. Zierler, A Class of Error-Correcting Codes in p^m Symbols, <u>J. Siam,</u> <u>9,</u> June 1961.

8. E. Berlekamp, On Decoding Binary Bose - Chaudhuri - Hocquenghem Codes, <u>IEEE Trans. on Info. Theory,</u> October 1965.

9. G. D. Forney, On Decoding BCH Codes, <u>IEEE Trans. on Info. Theory,</u> Oct. 1965.

10. G. D. Forney, On Decoding BCH Codes, MIT Research Laboratory of Electronics, TR 440, Dec. 1965.

11. G. D. Forney, Generalized Minimum Distance Decoding, <u>IEEE Trans. on Info. Theory</u>, April 1966.

12. J. Massey, Step-by-Step Decoding of the Bose-Chaudhuri-Hocquenghem Codes, <u>IEEE Trans. on Comm. Systems</u>, Oct. 1965.

13. Private Communication with J. Terzian.

14. M. Mitchell, Inhibited Error-Correction Decoder Performance, <u>IEEE Trans. on Comm. Systems</u>, Dec. 1962.

15. M. Mitchell, Performance of Error Correcting Codes, <u>IEEE Trans. on Comm. Systems</u>, Mar. 1962.

BIBLIOGRAPHY

W. Bennett, F. E. Froehlich, Some Results on the Effectiveness of Error-Control Procedures in Digital Data Transmission, IEEE Trans. on Comm. Systems, Mar. 1961.

E. Berlekamp, Block Coding with Noiseless Feedback, MIT PHD Thesis, September 1964.

K. Brayer, O. Cardinale, The Application of HF Channel Data to the Development of Block Error Correctors, The MITRE Corporation MTP-32, Contract AF 19(628)-5165, Bedford, Mass., June 1966; also available as ESD-TR-66-317.

K. Brayer and O. Cardinale, Evaluation of Error Correction Block Encoding for High Speed HF Data, IEEE Trans. on Comm. Tech., June 1967.

A. E. Fein et al, Time-Spread Coding and its Effectiveness in Transmission over Real Channels, IEEE Conference on Communication, June 1965.

M. Horestein, Sequential Transmission of Digital Information with Feedback, MIT Research Laboratory of Electronics, TR 375, Sept. 1960.

M. Horstein, Sequential Transmission Using Noiseless Feedback, IEEE Trans. on Info. Theory, July 1963.

F. Jelinek, Coding for and Decomposition of Two-Way Channels, IEEE Trans. on Info. Theory, Jan. 1964.

J. J. Metzner and K. C. Morgan, Coded Binary Decision-Feedback Communication Systems, IEEE Trans. on Comm. Systems, June 1960.

E. Parzan, Modem Probability Theory and its Applications, John Wiley and Sons, 1965.

W. W. Peterson, Error Correcting Codes, John Wiley and Sons, Inc., 1961.

L. S. Schwartz, Some Recent Developments in Digital Feedback Communication Systems, IEEE Trans. on Comm. Systems, June 1961.

L. S. Schwartz, Feedback for Error Control and Two-Way Communication, IEEE Trans. on Comm. Systems, March 1963.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
The MITRE Corporation Bedford, Massachusetts	Unclassified
	2b. GROUP

3. REPORT TITLE

Error Control Techniques Using Binary Symbol Burst Codes

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

N/A

5. AUTHOR(S) (First name, middle initial, last name)

Brayer, Kenneth

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1967	61	28

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
AF 19(628)-5165	ESD-TR-67-379
b. PROJECT NO. 705B	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	MTP-55

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Deputy for Surveillance and Control Systems, Aerospace Instrumentation Program Office; Electronic Systems Division, L. G. Hanscom Field, Bedford, Mass.

13. ABSTRACT

Much has been written on the theoretical description of error correcting codes but, due to a lack of actual channel error patterns, little has been said of practical performance. In this paper the performance of three types of error control is evaluated for the case of independent random errors and for an actual channel exhibiting dense bursts. The selected codes are burst codes with high probabilities of error detection and correction.

DD FORM 1 NOV 65 **1473**

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
SYSTEMS AND MECHANISMS						
Data Transmission Systems						
Multichannel Radio Systems						
Voice Communication (Troposcatter, Microwave, Wireline) Systems						
INFORMATION THEORY						
Coding						
MATHEMATICS						
Statistical Analysis (Troposcatter, Microwave, Wireline Error Locations)						
Statistical Distributions (Troposcatter, Microwave, Wireline Error Locations)						
Statistical Data (Troposcatter, Microwave, Wireline Error Locations)						