

October 15, 1957

Probability of Undetected Errors in Various
Data Transmission Systems

TABLE OF CONTENTS *

I	Introduction	10-2
II	System Requirements	10-2
	a. Probability of Undetected Character Error of 10^8	
	b. Number of Undetected Errors per Year	
III	Assumed Error Rates	
IV	Definitions	
	a. Symbols	
	b. Approximations	
	c. Conditional Probability	
	d. Dependent and Independent Error	
	e. Two Types of Bit Errors	
V	Case One -- Four-Out-of-Eight Code	10-8
VI	Case Two -- Six-Bit Code with Longitudinal Block Check	10-9
VII	Case Three -- Conditional Probability of Undetected Errors with Three Longitudinal Check Numbers.	10-18
VIII	Case Four -- Multiplying Factor due to Addition of Vertical Checking	10-21.1
IX	Case Five -- 7-Bit Code Without Longitudinal Check	
	a. Independent Errors	10-21.4
	b. Dependent Errors	10-21.6
X	Case Six -- 7-Bit Code With Longitudinal Block Check	10-22*
	a. Independent Errors	10-22.1
	b. Dependent Errors	10-22.2
XI	Note on Error Correcting Code for Reducing Number of Message Repeats on Human Input	10-23

Probability of Undetected Errors in Various Data Transmission Systems

I. Introduction

Combinations of horizontal and vertical erroring have been discussed by R. M. Gryb¹.

A more general mathematical model for determining the probability of undetected errors in magnetic tape systems has been developed by Schatzoff and Harding². The above types of models are now extended to several types of punched card data transmission systems. The mathematical models developed here are used to determine what type of error checking is sufficient to meet specified limits on the number of undetected errors that can be tolerated. Unless otherwise stated the examples used in this report are based on the transmission of standard IBM 80-column cards over telephone voice lines by means of transceivers or card readers and punches utilizing the 700 bit-per-second data line service to be offered by the Bell System.

II. System Requirements

a. Probability of Undetected Character Error of 10^{-8}

A specification of one undetected error per 10^8 characters has been suggested by J. A. McLaughlin³ for data transmission from a storage media. This class includes transmission of data from punched cards, paper tape, and magnetic tape.

-
1. R. M. Gryb, Bell Telephone Laboratories, "Error Checking with Particular Reference to Telegraph Systems" AIEE Summer and Pacific General Meeting: Conference Paper 56-844, June 26, 1956.
 2. M. Schatzoff and W. B. Harding, "A Mathematical Model for Determining the Probability of Undetected Errors in Magnetic Tape Systems" IBM Journal 1, 177-180 (April 1957)

3. J. A. McLaughlin, "Allowable Error Frequencies in Data Transmission",
Memorandum of 8/27/57. Section 3.

b. Number of Undetected Errors Per Year

Another criterion that has been suggested is to allow one undetected error per year. For a seven bit code transmitted at 700 bits per second, one error per year is equivalent to:

$$\frac{1 \text{ (yr/error)} \cdot 365 \text{ (cal day/yr)} \cdot 8 \text{ (hrs/wk day)} \cdot 3600 \text{ (sec/hr)} \cdot 700 \text{ (bits/sec)}}{7/5 \text{ (cal day/wk day)} \cdot 7 \text{ (bits/char)}}$$

$$= 7.5 \times 10^8 \text{ (char/year)} \times 1 \text{ (yr/error)} = 7.5 \times 10^8 \text{ (char/error)}$$

c. Methods of Obtaining Required Accuracy

Different methods of decreasing the probability of undetected errors are considered including redundant character codes such as the 7-bit code and the four-out-of-eight code. Also six-bit codes have no redundancy by character, but having a block parity check for each block of n characters. Combinations of both vertical (redundant parity bit in each character) and longitudinal block checking are considered.

III. Assumed Error Rates

Since data is not yet available on the expected error rate on the Bell System 700 bit per second data transmission service, the following error rate will be used for the sample calculations.

Take the error rate of one per 100,000 bits which the Bell System guarantees on specially engineered systems operating at 1600 bits per second as a limiting value. Multiply by ten to obtain a conservative estimate of the error rate at 700 bits/sec on commercial dial-up lines without any special engineering. This gives us an assumed probability of a bit error of:

$$P_b = 0.0001$$

The probabilities of a single error in a character of a bits is then:

$$P_c = \frac{a!}{1! (a-1)!} (P_b)^1 (1-P_b)^{a-1} \approx a(P_b) \quad (1)$$

TABLE I

Assumed Bit and Character Error Probabilities

$P_b^a \rightarrow$ ↓	P_c		
	6	7	8
.0001	.0006	.0007	.0008
.00001	.00006	.00007	.00008

The two sets of error probabilities are used in sample calculations in this report to give a range of potential values of the probabilities undetected errors. In some examples curves are plotted over a larger range of probability of bit errors.

IV. Definitions

a. Symbols

P_b = probability that a bit is in error.

$P_b(0-1)$ = probability that a "0" bit has been changed to a "1" bit.

$P_b(1-0)$ = probability that a "1" bit has been changed to a "0" bit.

P_c = probability that a character is in error.

P_r = probability that a record (card) is in error

$P_r(u)$ = probability that there is an undetected error in a record (card)

$P_r(u;n)$ = probability that there is an undetected error in a record (card) caused by n bits in error

$P_r(u/lj)$ = probability that there is an undetected error in a record (card) consisting of i (1 - 0) errors and j (0 - 1) errors.

$P_r(m:i, j)$ = probability that there is a multiple error in a record consisting of m total errors where there are

$P_r(i, j)$ i (1 - 0) errors and j (0 - 1) errors, such that:
 $i + j = m$.

$P(A)$ = probability of event A

$P(B/A)$ = probability of event B conditions upon the event A,
(conditional probability)

$P(A, B)$ = probability of event A and B (joint probability).

$P_r(b_k, b_{k+1})$ = probability of an error in bit k and in bit $k+1$ of a record.

b. Approximations

The probability of undetected errors where an error checking system is used, is dependent upon multiple errors occurring in a pattern that can slip by the checking system. The probability that m errors occur in a block of n events where P is the probability of a single independent event is:

$$P_n(m) = \frac{n!}{m! (n-m)!} P^m (1-P)^{n-m} \quad (2)$$

A short exposition of the application of the Poisson distribution where n is large, P small, making nP small is given by Goode & Machol⁴. Using the Poisson distribution as described in ref's 4 we have the probability of m bit errors in a card or row of n bits.

4. H. H. Goode and R. E. Machol, System Engineering, N.Y. McGraw-Hill (1957) Ch. 5. "Dist. of Discrete Variables".

$$\begin{array}{cccccccc}
 m \rightarrow & & 0 & 1 & 2 & 3 & & x \\
 \sum_m P_n(m) = & e^{-nP} & (1 + nP + \frac{(nP)^2}{2!} + \frac{(nP)^3}{3!} + \dots + \frac{(nP)^x}{x!} + \dots) & (3)
 \end{array}$$

c. Conditional Probability

The understanding of conditional probability is important in two areas:

(1) cases where a distinction between types of errors (i.e., 1 - 0 and 0-1) makes a difference and (2) where multiple errors are dependent such as the situation where the probability of an error is greater in the bit following a bit with an error.

$$\text{Axiom I: } P(A \text{ or } B \text{ or both}) = P(A) + P(B) - P(A, B). \quad (4)$$

$$\text{Axiom II: } P(A, B) = P(A/B) P(B) = P(B/A) P(A) \quad (5)$$

d. Dependent and Independent Errors

For a record of n bits having "a" total bit errors and "b" double errors (i.e., adjacent bits in error) the probabilities of error are as follows:

$$P_b(b_k) = \frac{a}{n} \quad (6) \text{ single errors}$$

$$P_b(b_k, b_{k+1}) = \frac{b}{n} \quad (7) \text{ double errors}$$

For example, when: $n = 10^6$, and

$$a = 100, b = 2,$$

$$P_b(b_k) = 100/10^6 = 10^{-4}$$

$$P_b(b_k, b_{k+1}) = 2/10^6 = 2 \times 10^{-6}$$

To convert experimental data on non-independent errors to a standard form, use;

$$P(b_k, b_{k+1}) = P(b_{k+1}/b_k) P(b_k) \quad (8A)$$

$$P(b_{k+1}/b_k) = \frac{P(b_k, b_{k+1})}{P(b_k)} = \frac{b/n}{a/n} = \frac{b}{a}$$

For the above example the conditional probability of a double error is:

$$P(b_{k+1}/b_k) = \frac{b}{a} = \frac{2}{100} = .02$$

If the errors are independent, the probability of a double error in n bits is:

$$P(b_k, b_{k+1}) = C_2^n P(b_k)^2 [1 - P(b_k)]^{n-2} \quad (9)$$

where: $C_2^n = \frac{n!}{2! (n-2)!}$

For dependent double errors:

$$P_n(b_k, b_{k+1}) = (n-1) P(b_k) P(b_{k+1}/b_k) \quad (10)$$

Table II

Two examples of dependent errors:

a	b	n	$P_b(b_k) = \frac{a}{n}$	$P_b(b_k, b_{k+1}) = \frac{b}{n}$	$P_b(b_{k+1}/b_k)$
100	2	10^6	10^{-4}	2×10^{-6}	.02
10	1	10^6	10^{-5}	1×10^{-6}	.10

Two sample calculations of the probability of a double error in adjacent bits in an 8-bit character for independent and dependent errors are tabulated below.

Table III

		$P_c(b_k, b_{k+1})$	
		Independent $P_n(b_k) = 10^{-4}$	Dependent $P_n(b_k) = 10^{-4}$
n		a/b	
8	7×10^{-8}	.02	1.4×10^{-5}
8	7×10^{-8}	.10	7×10^{-5}

The above case are for errors in two adjacent bits in a record of n bits from equation (10).

For simple double errors in any two bits within the n bits use equation (9).

c. Two Types of Bit Errors

In a binary system, two types of errors can occur:

<u>Type</u>	<u>Symbol</u>
0 → 1	$P_b(0-1)$
1 → 0	$P_b(1-0)$

The total bit error probability of error is:

$$P_b = P_b(0-1/1) P_b(1) + P_b(1-0/0) P_b(0) \quad (11)$$

Where the distribution of 0's and 1's are equal,

$$P_b(1) = P_b(0) = \frac{1}{2} \quad (12)$$

$$\text{Then, } P_b = P_b(0-1/1) = P_b(1-0/0) \quad (13)$$

V. Case One - Four-Out-of-Eight-Code

a. Independent Errors

The probability of an undetected error in the 4-out-of-8 code is for independent errors:

$$P_c(u) = \left\{ \frac{m!}{1! (m-1)!} P_b(0-1) [1 - P_b(0-1)]^{m-1} \right\} \\ \times \left\{ \frac{n!}{1! (n-1)!} P_b(1-0) [1 - P_b(1-0)]^{n-1} \right\}. \quad (14)$$

For $m = n = 4$ and $P_b(0-1) = P_b(1-0)$, this reduces to:

$$P_c(u) \approx 16 P_b^2 \quad (15)$$

for independent errors.

For dependent errors, where a 01 is changed to a 10 or a 10 is changed to a 01, which would not be detected by the 4-out-of-8 code Eq. (10)⁽⁸⁾ can be used with typical data

Sample calculations for independent errors are tabulated below:

Table IV

Independent Errors Four-Out-Of-Eight Code				
P_b	Undetected Errors		Detected Errors	
	$P_c(u)$	Characters per Und. Error	Records per Repeat Order	
			$n = 80$	$n = 1000$
10^{-4}	1.6×10^{-7}	6×10^6	16	
10^{-5}	1.6×10^{-9}	6×10^8	160	12
7×10^{-6}	8×10^{-10}	1.2×10^9	220	18
2×10^{-6}	6.4×10^{-11}	1.56×10^{10}	770	63
1.1×10^{-7}	1.8×10^{-13}	5.5×10^{12}	13900	1130

A curve of records per repeat order for the 4-out-of-8 code for $n = 20$ is plotted on page 10-25.

Table V-A
Four-out-of-Eight Code
Dependent Errors

Error Probability Conditions	Undetected Errors		Detected Errors Records per Repeat	
	$P_c(u)$	Char. per Und. Error	n = 80	m = 1000
$P_b = 1.1 \times 10^{-7}$ a = 108, b = 7 $n^1 = 6.3 \times 10^7$	5×10^{-8}	2×10^7	13,900	1130
$P(b_{k+1}/b_k) = .065$	2.5×10^8	4×10^7		
$P_b = 2 \times 10^{-6}$ $\frac{b}{a} = .0014$	1×10^{-8}	10^8		
$P_b = 10^{-4}$ b/a = .01	3.5×10^{-6}	2.9×10^5		

*Eq. (10) gives the probability of two errors occurring in a character such that the two errors are in adjacent bits. In the four-out-of-eight code only half of these are undetected, namely double errors affecting 01 and 10 sequences. In this case Eq. (10) should be replaced by:

$$P_n(u/2) = P(u/b_k, b_{k+1}) P_n(b_k, b_{k+1}) \quad (15B)$$

$$\text{when: } P(u/b_k, b_{k+1}) = \frac{n^1}{n-1} \quad (15C)$$

Both by the fraction of possible sequences which are 01 or 10 and by counting the sequences used in the four-out-of-eight code tabulated on pages 2-2 and 2-2.1 we have:

$$\frac{n^1}{n-1} = \frac{3.5}{7} = \frac{1}{2} \quad (15D)$$

See also curves of Fig. 10.0d

b. Approximate Analysis of Dependent Errors

The effect of dependent errors is a function of the frequency spectrum of the noise. This in turn is a function of the detection system and the point in the transmission system where the noise enters the systems. The noise can have a spectrum ranging through the following.

1. Infinite Spectrum, like a pure impulse entering system at receiving exchange.
2. Noise spectrum peaked at frequency equal to bit rate (i.e., twice the fundamental transmission frequency).
3. Noise spectrum peaked at fundamental transmission frequency (i.e., half of bit rate) corresponding to noise entering system near sending end so that noise is filtered by transmission network. In this case noise and signal have similar spectrum at receiver.

Sample noise and signal waveform are plotted in Fig. 10.0

The use of $P(C/B)$ from Tables V-B and V-C is as follows:

1. Determine $P(B)$, where B is the event of a double error (in adjacent bits) from Eq. (9) if independent, or Eq. (10) if dependent.
2. Determine the type of noise signals as illustrated in Fig. 10.0, Cases (1, 2, 3).
3. From Table V-B determine $P(C/B)$ as a function of code and noise case.
4. Multiply $P(C/B)$ and $P(B)$ to get $P(u)$.
5. A further connection is required: That is to analyse the detection system being used.

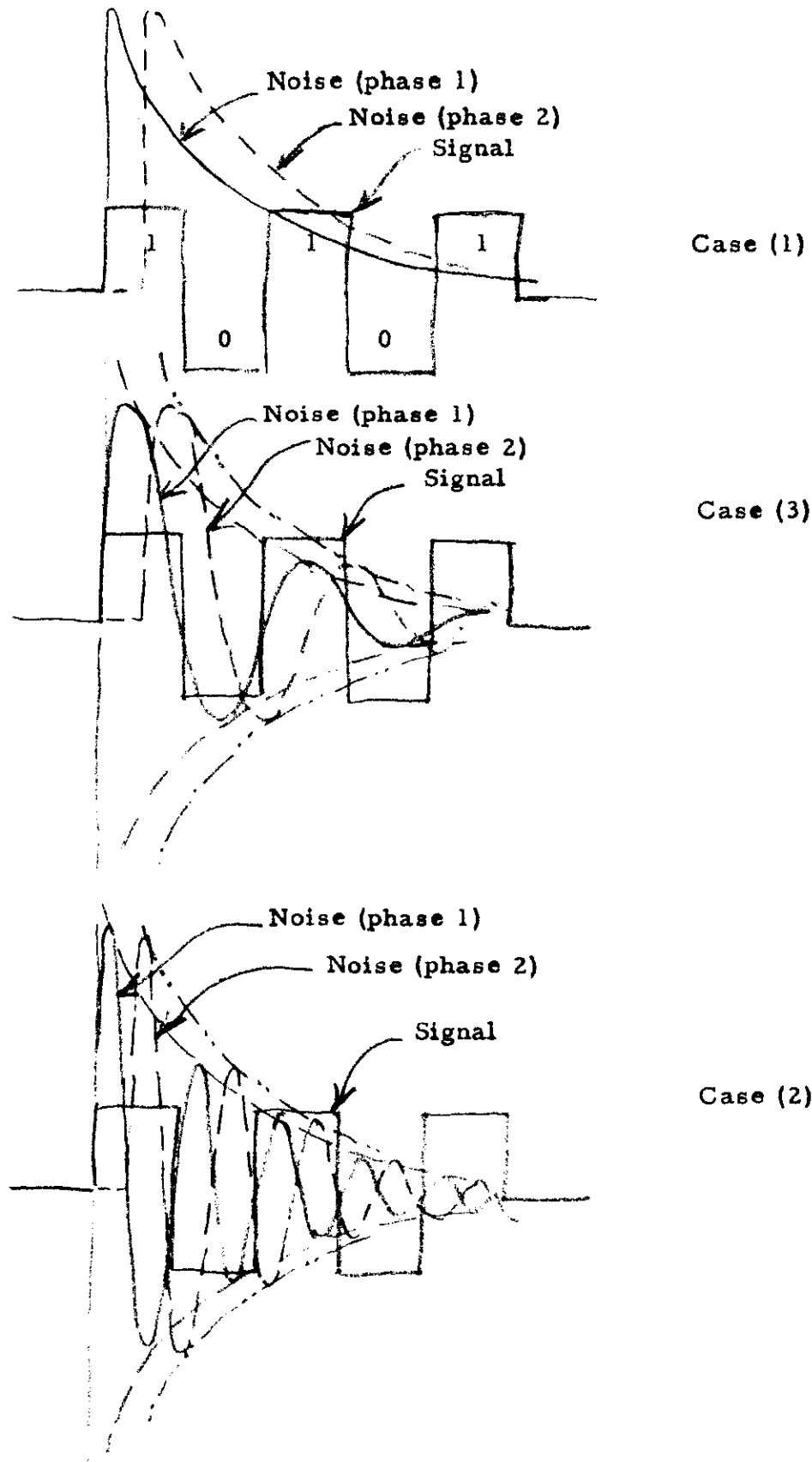


Fig. 10.0 Types of Double Errors Caused by Single, Long Noise Impulses

Table V-B

Comparison of 4-out-of-8 Code and 7-bit Code in Detecting Double Errors for Different Types of Noise						
Sending Code	Received Code and Detection Rating (D, U, N)					
	Case 1		Case 2		Case 3	
	Rec'd Code	4/8 - 7	Rec'd Code	4/8 - 7	Rec'd Code	4/8 - 7
00	11	D - U	00	N - N	10	D - D
01	11	D - D	01	N - N	10	U - U
10	11	D - D	11	D - D	10	N - N
11	11	N - N	11	N - N	10	D - D
00(0)*	11(1)	D - D	10	D - D	10	D - D
(1)	(1)	D - U				
01(0)	11(1)	D - U	11	D - D	11	D - D
(1)	(1)	D - D				
10(0)	11(1)	D - U	10	N - N	10	N - N
(1)	(1)	D - D				
11(0)	11(1)	D - D	11	N - N	11	N - N
(1)	(1)	N - N				
P(C/B)		$\frac{0-5}{16}$		0 - 0		$\frac{1-1}{8-8}$

Noise Phase 1

Noise Phase 2

D = Error detected

N = No error made

U = Error undetected

P(C/B) = Probability that a double error occurs if a single noise impulse extends over two bit periods.

* The third bit in parenthesis allows for a noise impulse extending over one full bit and two half bit periods.

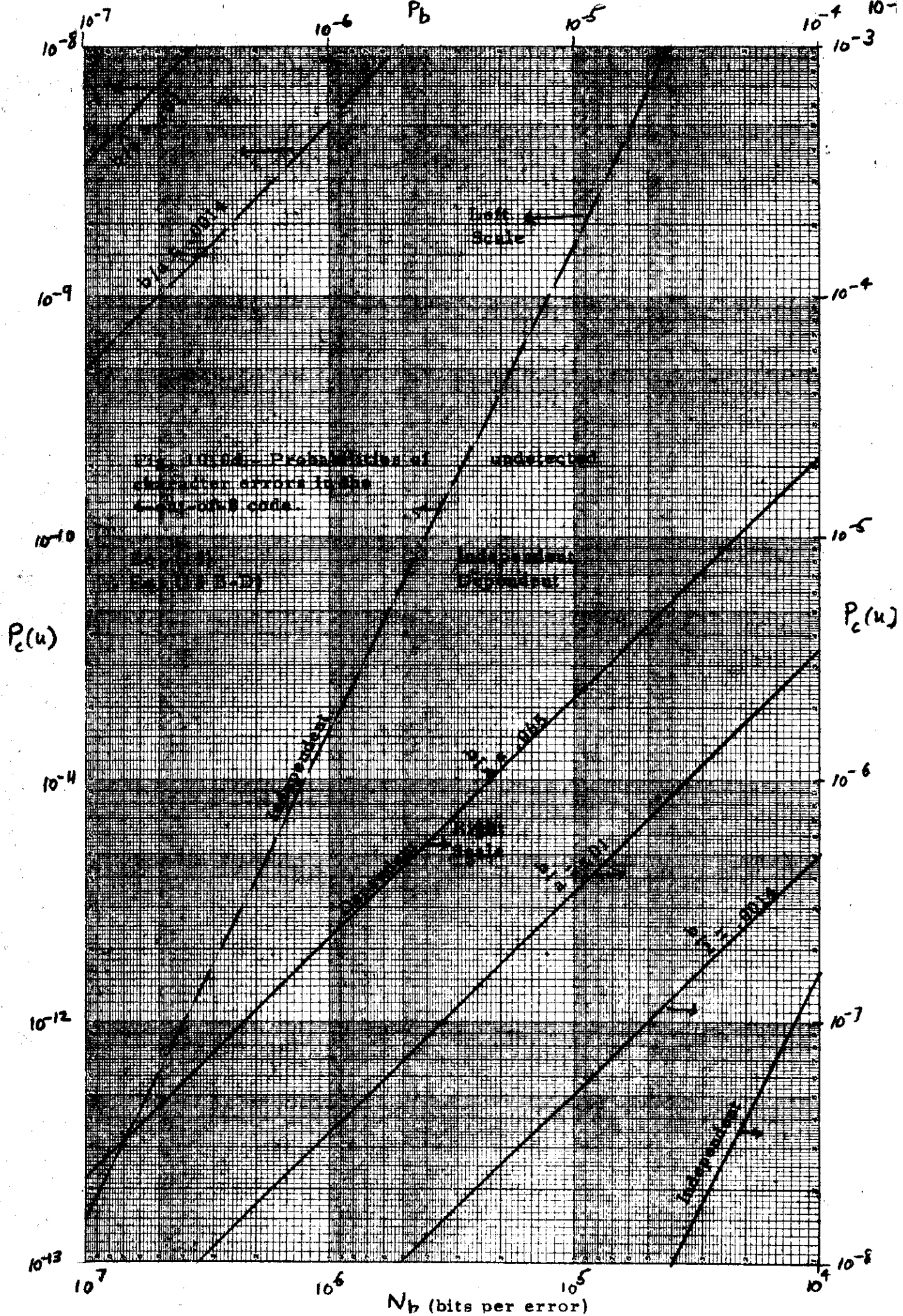


Fig. 10.10. Probabilities of undetected character errors in the transmission of 8 code.

Independent
Dependent

Independent

Dependent

Independent

VI. Case Two - Six-Bit Code with Longitudinal Block Check

A six-bit coding of a card of 80 col. is shown in Fig. 10.1. This type of error-checking is similar to some of the types discussed by R. M. Gryb* for use with the five-unit telegraph code. The analysis of this section indicates the range of accuracy attainable with a six-bit code for potential use with IBM punches, printers, and readers operating at 100 characters per second. For an example of such a system see the proposal of Dr. W. A. Christopherson**.

* R. M. Gryb, "Error Checking with Particular Reference to Telegraph Systems" Bell Telephone Laboratories, Conference Paper No. 56-844, AIEE Summer and Pacific General Meeting, June 26, 1956. Various systems of block checking using existing telegraph 5-unit code equipment are considered for the purpose of increasing the number of characters per undetected error from 44,000 to 10^8 . Some examples include both vertical and horizontal checking.

** W. A. Christopherson, "Proposal for a 100-Column-per-Second Serial Card Data Transceiver", Report RJ-DR-532.015 August 19, 1957.

The non-zero squares in Fig. 10.2, 10.3, 10.4, represent the terms that contribute to undetected errors for:

$$g = 1, 2, 3 \quad \text{or} \quad k = 2^g = 2, 4, 8$$

A first approximation to the $P_r(u/1,1)$ term in Fig. 10.2 is developed as follows:

The probability of there being at least one error in a card is by equation (1) for 6 rows x 80 columns:

$$P(m:1) = n_r \times n_c P_b (1-P_b)^{\sqrt{n_r \times n_c}} \simeq n_r \times n_c P_b \quad (16)$$

$$P_r(m:1) = 6 \times 80 P_b (1-P_b)^{480} \simeq 480 P_b \text{ for } P_b \ll 1$$

A second compensating error must be in the same row and of opposite type. This gives $\frac{1}{2} \times 80 = 40$ positions satisfying the condition, so

$$P_r(u:1,1) = 480 P_b \cdot 40 P_b = 1.92 \times 10^4 (P_b)^2$$

$$\text{then } P_r(u:1,1) = 1.92 \times 10^4 (10^{-4})^2 \simeq 2 \times 10^{-4} \quad * \quad (17)$$

When treating $0 \rightarrow 1$, and $1 \rightarrow 0$ errors separately, the value of P_b is the same if $P_b(0) = P_b(1)$ as is discussed in Section IV-C.

* In an earlier draft of this analysis a bit error rate was used which was one-half of the bit error rate used here.

SUBJECT _____

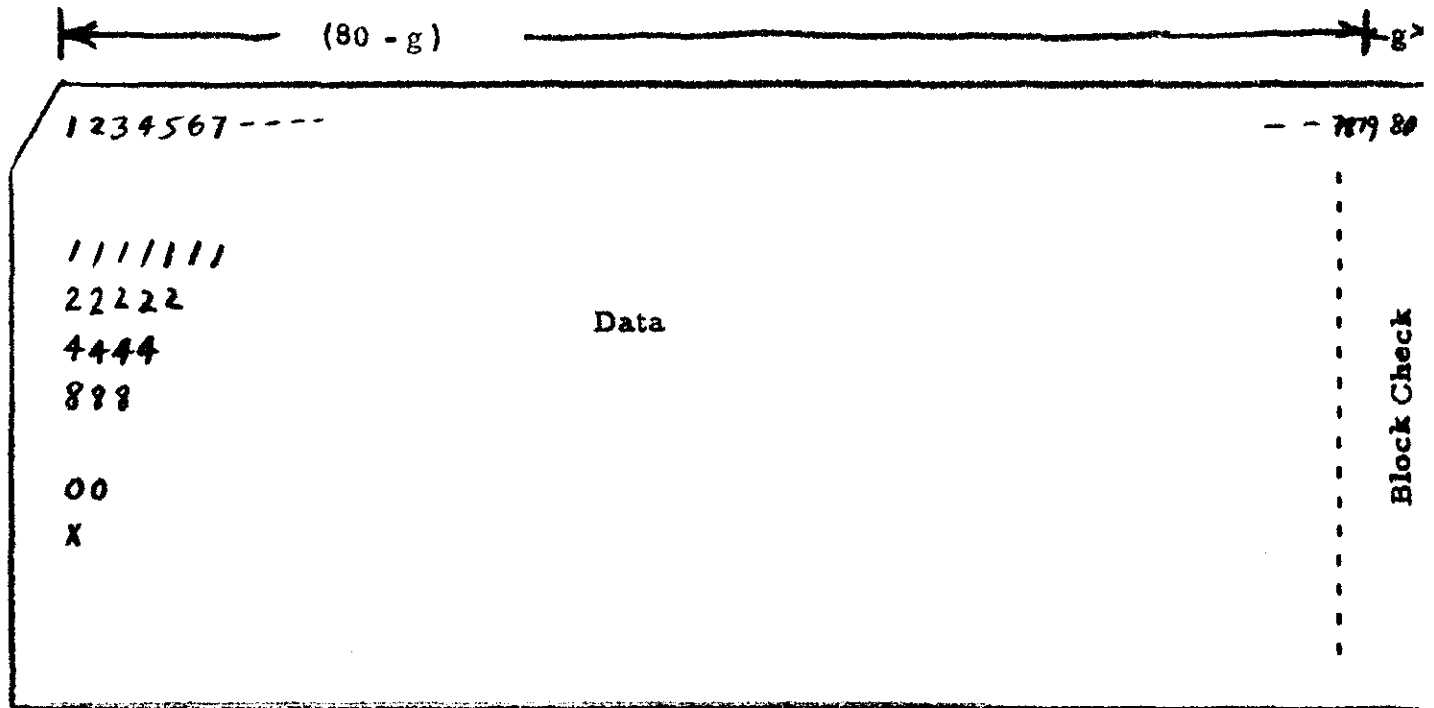


Fig. 10.1 Form of Transmission of Hollerith Card Data in a Six-Bit Code.

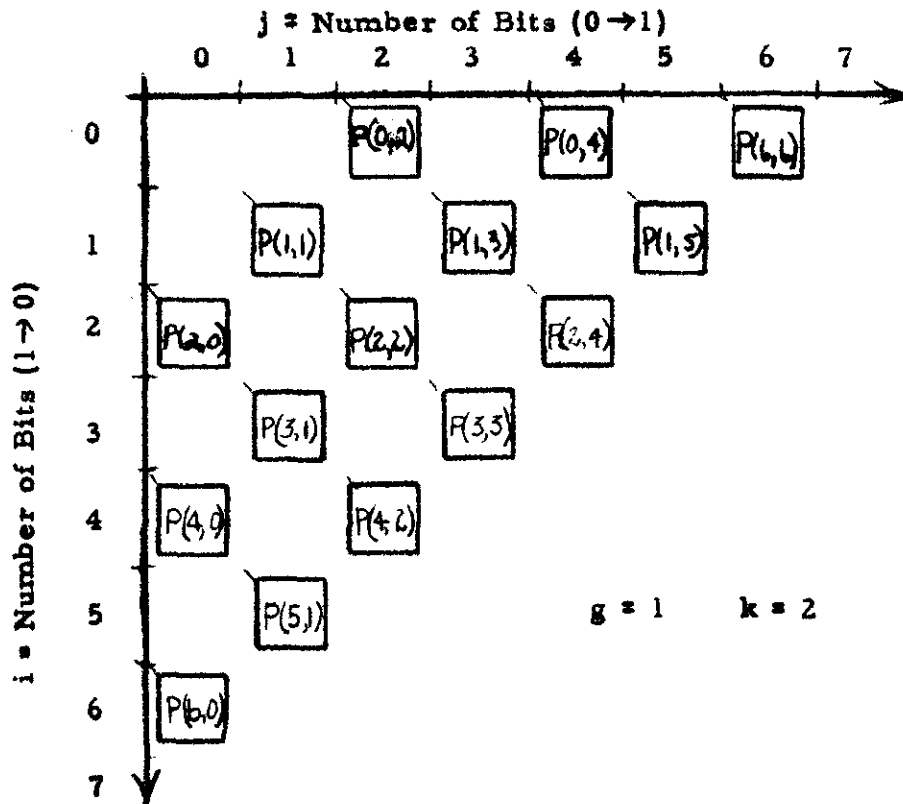


Fig. 10.2 Matrix of Possible Undetected Error Combinations in a Six-Code with Longitudinal Block Check

IBM

**SAN JOSE RESEARCH LABORATORY
TECHNICAL RECORD**

CODE _____
DISCLOSURE NO. _____
NAME _____
DATE _____
SHEET _____ OF _____

SUBJECT _____

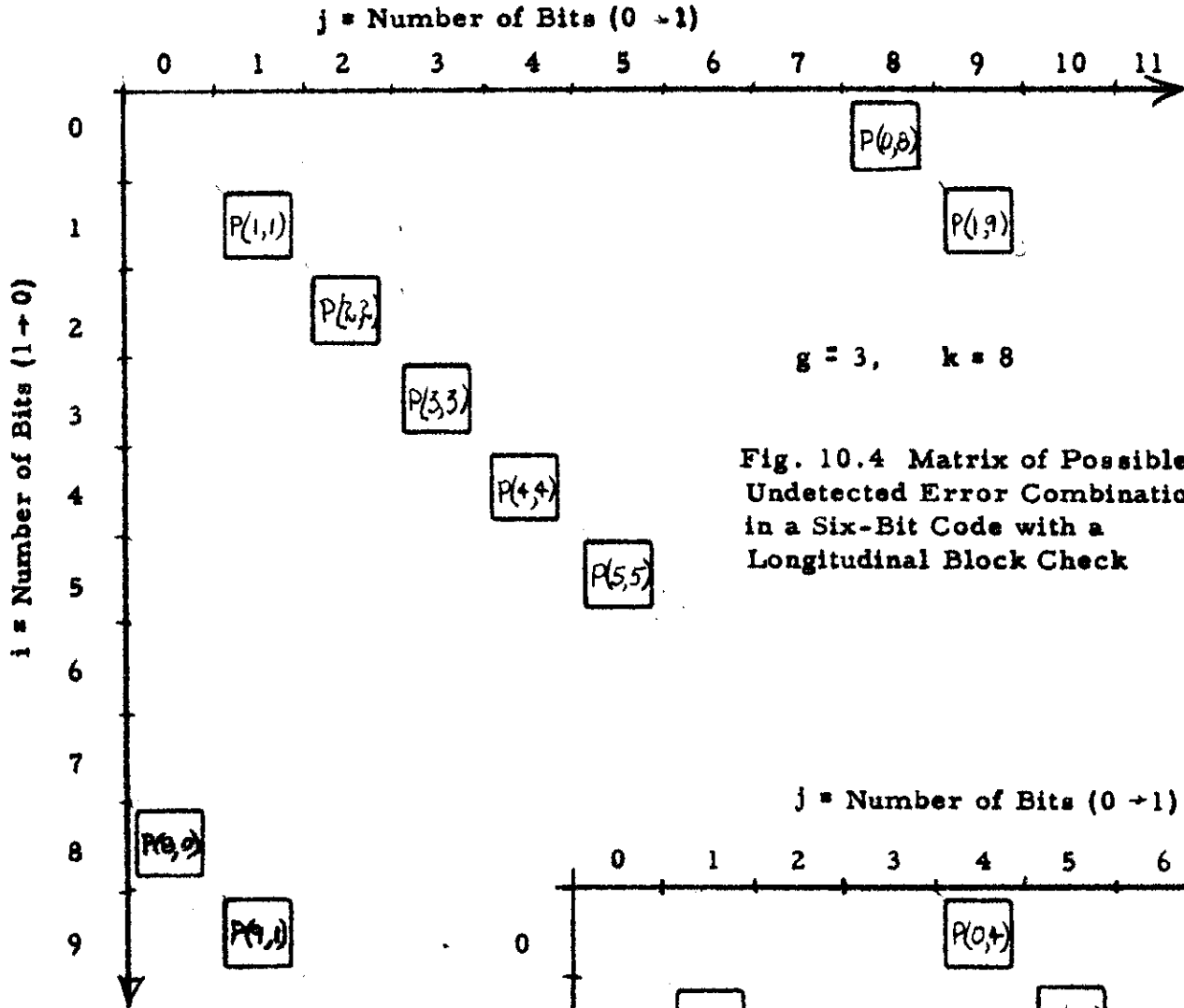


Fig. 10.4 Matrix of Possible Undetected Error Combinations in a Six-Bit Code with a Longitudinal Block Check

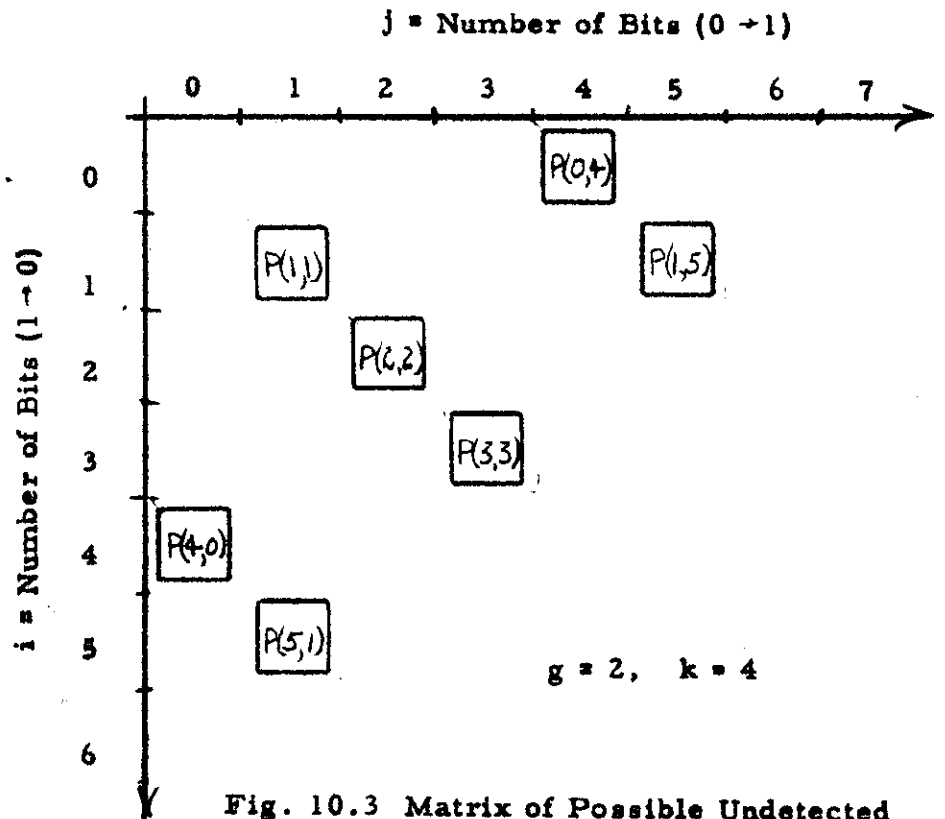


Fig. 10.3 Matrix of Possible Undetected Error Combinations in a Six-Bit Code with a Longitudinal Block Check.

The term $P_r(u/2, 0)$ for two "1" converted to "0"s derived as follows. Using the binomial distribution and assuming

$$\begin{aligned}
 P_b(1 - 0) &= P_b(0 - 1) = P_b, \text{ a row of } r \text{ bits, of } r/2 \text{ "0" bits and} \\
 r/2 \text{ "1" bits } P_r(u/2, 0) &= P_{r/2}(2) P_{r/2}(0) = \\
 &= \frac{(c/2)!}{2! (c/2-2)!} P_b^2 (1 - P_b)^{c-2} (1 - P_b)^{r/2} \\
 &\simeq \frac{\binom{c}{2} (\frac{c}{2} - 1) P_b^2}{2} \quad (17B)
 \end{aligned}$$

i. e., there are c characters or columns per card or c bits per row.

For $P_u = 10^{-4}$ a sample calculation gives:

$$P_r(u/2, 0) = \frac{40.39}{2} 10^{-8} = 7.8 \times 10^{-6}$$

For the card define $P_s(u)$, as the probability of an undetected error in a card of s rows.

$$P_s(u/2, 0) = S P_r(u) [1 - P_r(u)]^s \simeq S P_r(u; 2, 0) \quad (18)$$

For $s = 6$ and $P_r = 7.8 \times 10^{-6}$:

$$P_s(u/2, 0) = 6 \times 7.8 \times 10^{-6} = 4.67 \times 10^{-5}$$

Using the Poisson approximation for higher order terms:

$$P_s(u/i, j) = s P_r(u) [1 - P_r(u)]^s \simeq e^{-s P_r} (s P_r) \simeq s P_r(u; i, j) \quad (19)$$

$$P_r(u/i, j) = P_{r/2}(u/i) P_{r/2}(u/j) \quad (20)$$

With restriction on i and j determined from Fig. 10.2-10.4

$$P_{r/2}(u/i) \simeq \frac{C^{-c/2} P_b \left(\frac{c}{2} P_b\right)^i}{i!} \quad (21A)$$

$$P_{r/2}(u/j) \simeq \frac{C^{-c/2} P_b \left(\frac{c}{2} P_b\right)^j}{j!} \quad (21B)$$

(21) in (20) gives

$$P_r(u/i, j) = \frac{C^{-cP_b} \left(\frac{c P_b}{2}\right)^{i+j}}{(i!) (j!)} \quad (22)$$

For $P_s(u:m)$ where $i+j = m$ for example $m = 2$

$$\begin{aligned} P_s(u/2) &= P_s(u/2, 0) + P_s(u/1, 1) + P_s(u/0, 2) \\ &= P_s(u; 1, 1) + 2P_s(u; 0, 2) \end{aligned} \quad (23)$$

$$P_r(u/i, i) = \frac{e^{-cP_b} \left(\frac{cP_b}{2}\right)^{2i}}{(i!)^2} \quad (24)$$

$$P_s(u/m) = \sum_{k=0}^m P_s(u/k, m-k) \quad (25)$$

Values of terms contributing to the probability of undetected errors are tabulated in:

Table VI for $g = 1, 2, 3$

Sample Calculation for $C = 80$, $P_b = 10^{-4}$, $S = 6$:

$$P_s(u/0, 2) = \frac{6 P_r(u/0, 2) = 6 e^{-cP_b} \frac{cP_b^2}{2!}}{(0!) (2!)}$$

$$= 6 e^{-.008} (.004)^2 = 3 \times .992 \times 16 \times 10^{-6} = 47.6 \times 10^{-6}$$

$$P_s(u/2, 2) = 6 P_r(u/2, 2) = \frac{6 e^{-.008} (.004)^4}{(2!) (2!)}$$

$$1.5 \times .992 \times .256 \times 10^{-9} = .38 \times 10^{-9}$$

$$P_s(u/3, 1) = 6 e^{-.008} (.004)^4 = 6 \times .992 \times 256 \times 10^{-12} = .253 \times 10^{-9}$$

$$P_s(u/4, 0) = \frac{6 \times .992 \times 256 \times 10^{-12}}{(4!) (0!)} = .063 \times 10^{-9}$$

$$P_s(u/4) = P_s(u/2, 2) + 2 P_s(u/3, 1) + 2 P_s(u/4, 0)$$

$$= (.380 + 2 \times .253 + 2 \times .063) 10^{-9} = .992 \times 10^{-9}$$

$$P_s(u/2) = P_s(u/1, 1) + 2 P_s(u/0, 1) = .0002 + 2 \times .000048 = .000292$$

TABLE VI

Undetected Errors in Six-Bit Code with
Longitudinal Block Check

Terms Not Detected			n = 80 characters	g = 1, 2, 3	$P_b = 10^{-4}$		
g							
1	2	3					
x	x	x	$P_s(u/1,1)$.000	200		
x			$P_s(u/2,0)$.000	047		
x			$P_s(u/0,2)$.000	047		
x	x	x	$P_s(u/2,2)$.000	000	000	380
x			$P_s(u/1,3)$.000	000	000	253
x			$P_s(u/3,1)$.000	000	000	253
x	x		$P_s(u/4,0)$.000	000	000	063
x	x		$P_s(u/0,4)$.000	000	000	063
x	x	x	$P_s(u/3,3)$.000	.000	000	000 000 680

$P_s(u/g = 1) = .000\ 294$	} Probability of an undetected error in a record of 80 characters
$P_s(u/g = 2) = .000\ 200$	
$P_s(u/g = 3) \approx .000\ 200$	
$(P_s(u) \longrightarrow .000\ 000\ 8$	required accuracy)

Note on required accuracy: The requirement of 10^8 characters per undetected error corresponds to $N_r = 10^8/80 = 1.25 \times 10^6$ cards per undetected error or $P_s(u) = 1/N_r = 8 \times 10^{-7}$

10⁻³

FIG. 10-15 - Probability of Undetected Error, P_u, in a Code with Minimum Hamming Distance D_{min} = 2t + 1. Errors are assumed to be independent.

10⁻⁴

Errors are assumed to be independent

10⁻⁵

P_u(n, t)

P_u(n)

Proportional to n^t

P_u(n, t=0)
P_u(n, t=1)

10⁻⁶

Maximum P_u

For one error
in a 10⁶ block

Maximum P_u for one
undetected error

10⁻⁷

Maximum P_u for one
undetected error

10⁻⁸

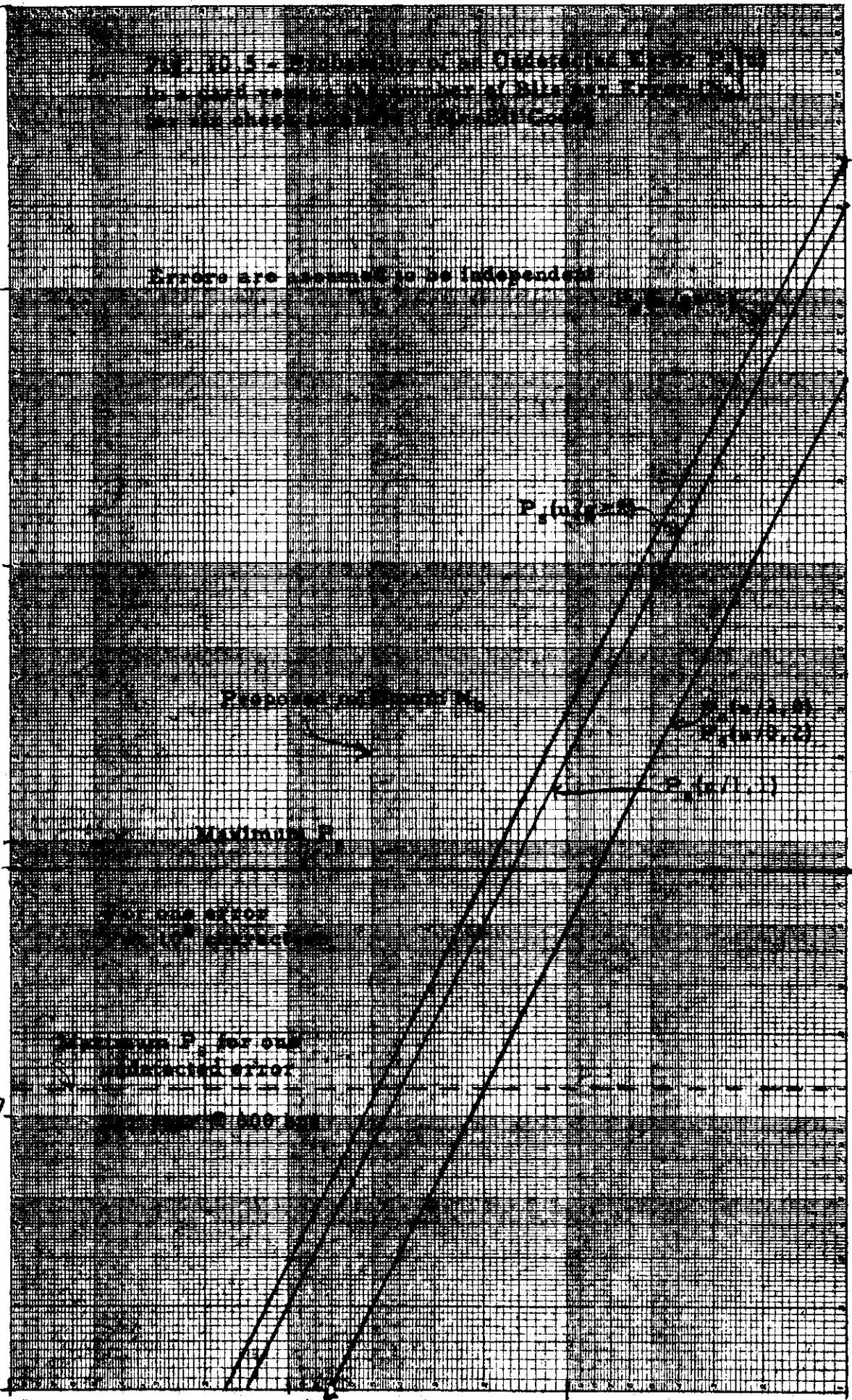
10⁷

10⁶

10⁵

10⁴

N_b (number of bits/error)



The curves of Fig 10.5 assume independent errors. In this longitudinal checking, the type of dependent error which could make the probability of undetected errors increase would be a second error. Six bits later, $P_b (b_{k+6} / b_k)$. There is little physical basis for such a conditional probability being greater than $P_b (b_k)$ because a large noise impulse affecting two bits in a character would cause single errors in two separate longitudinal checking strings. A "string" is defined as row, column, or other group of bits which are counted to form a particular parity bit in an error checking system.

VII Case Three - Conditional Probability of Undetected Errors
With Three Longitudinal Check Numbers.

For the purpose of adding vertical checking without changing to a full 7-bit code the longitudinal check is reduced to three characters leaving three characters free for a vertical check.

The three character longitudinal check is shown in Fig. 10.6.

The ratio of the terms in this case to those of the previous case is by Eq. (19):

$$\begin{aligned} \frac{P_s^1}{P_s} &= \frac{S^1 \frac{e^{-c^1 P_b} \left(\frac{c^1 P_b}{2}\right)^{i+j}}{(i!) (j!)}}{S \frac{e^{-c P_b} \left(\frac{c P_b}{2}\right)^{i+j}}{(i!) (j!)}} \\ &= \frac{S^1 e^{-c^1 P_b} (c^1)^{i+j}}{S e^{-c P_b} (c)^{i+j}} = \left(\frac{s^1}{s}\right) e^{-(c^1/c) P_b} \left(\frac{c^1}{c}\right)^{i+j} \end{aligned} \quad (26)$$

s = number of rows or groups of rows in longitudinal check

c = number of character in a row or group of rows in one longitudinal check.

In Fig. 10.1: $s = 6$, $c = 80 = L$ $sc = 480$

In Fig. 10.6: $s^1 = 3$ $c^1 = 160 = L$ $s^1 c^1 = 480$

When $c P_b \ll 1$,

$$\frac{P_s^1}{P_s} \approx \left(\frac{s^1}{s}\right) \left(\frac{c^1}{c}\right)^{i+j} \quad (27)$$

$$\text{For } \frac{s^1}{s} = \frac{1}{2}; \frac{c^1}{c} = 2 \quad \frac{P_s^1}{P_s} = (2)^{i+j-1} \quad (28)$$

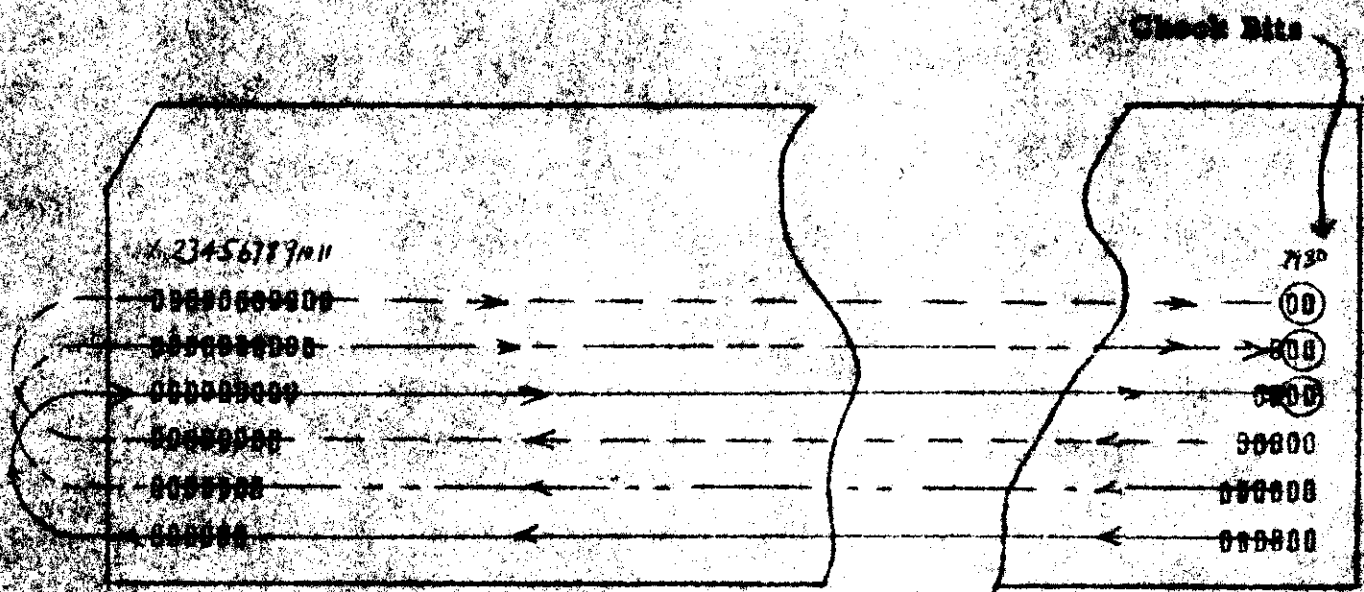


Fig. 10.6 5-Bit Code with Three Longitudinal Check Numbers

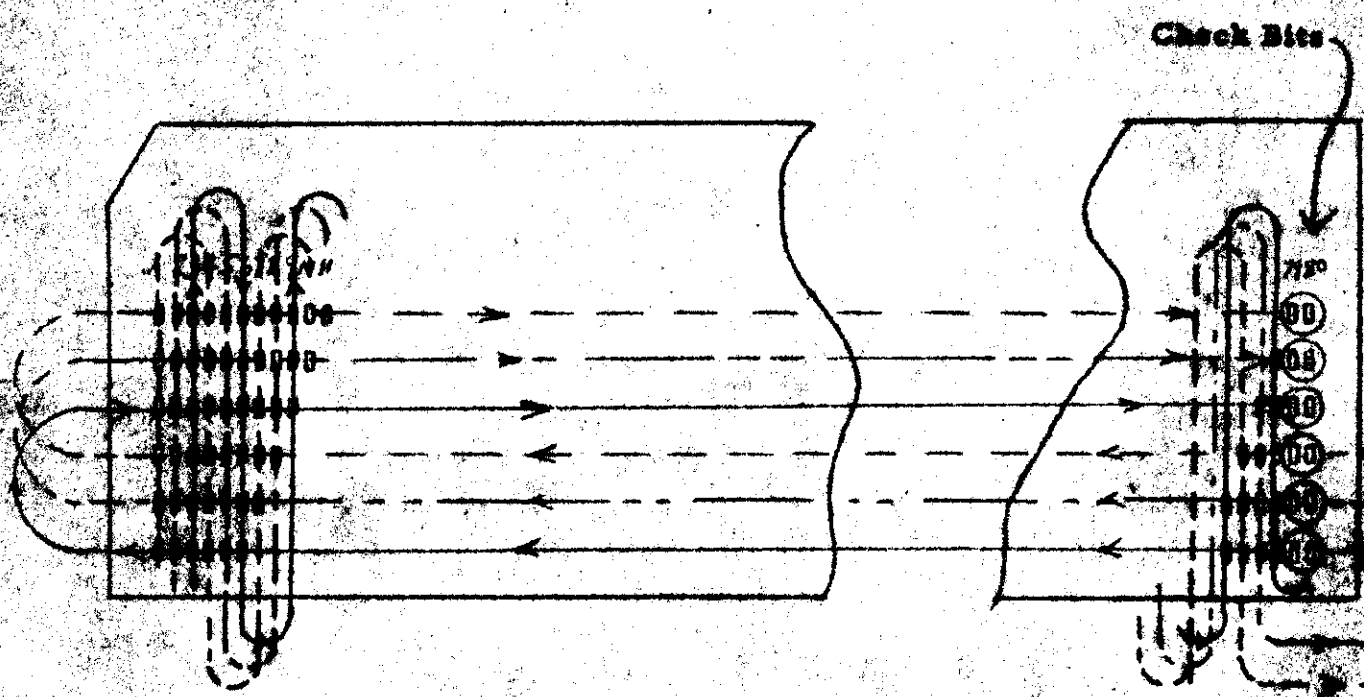


Fig. 10.7A Interlaced Longitudinal Check and Vertical Check.

Sample Calc:

$$P_s^1(u/1,1) = \left(\frac{3}{6}\right) \left(\frac{160}{80}\right)^2 \cdot 0.0002 = .0004$$

The results are tabulated in Table VII for $P_b = 10^{-4}$ or $N_b = 10^4$

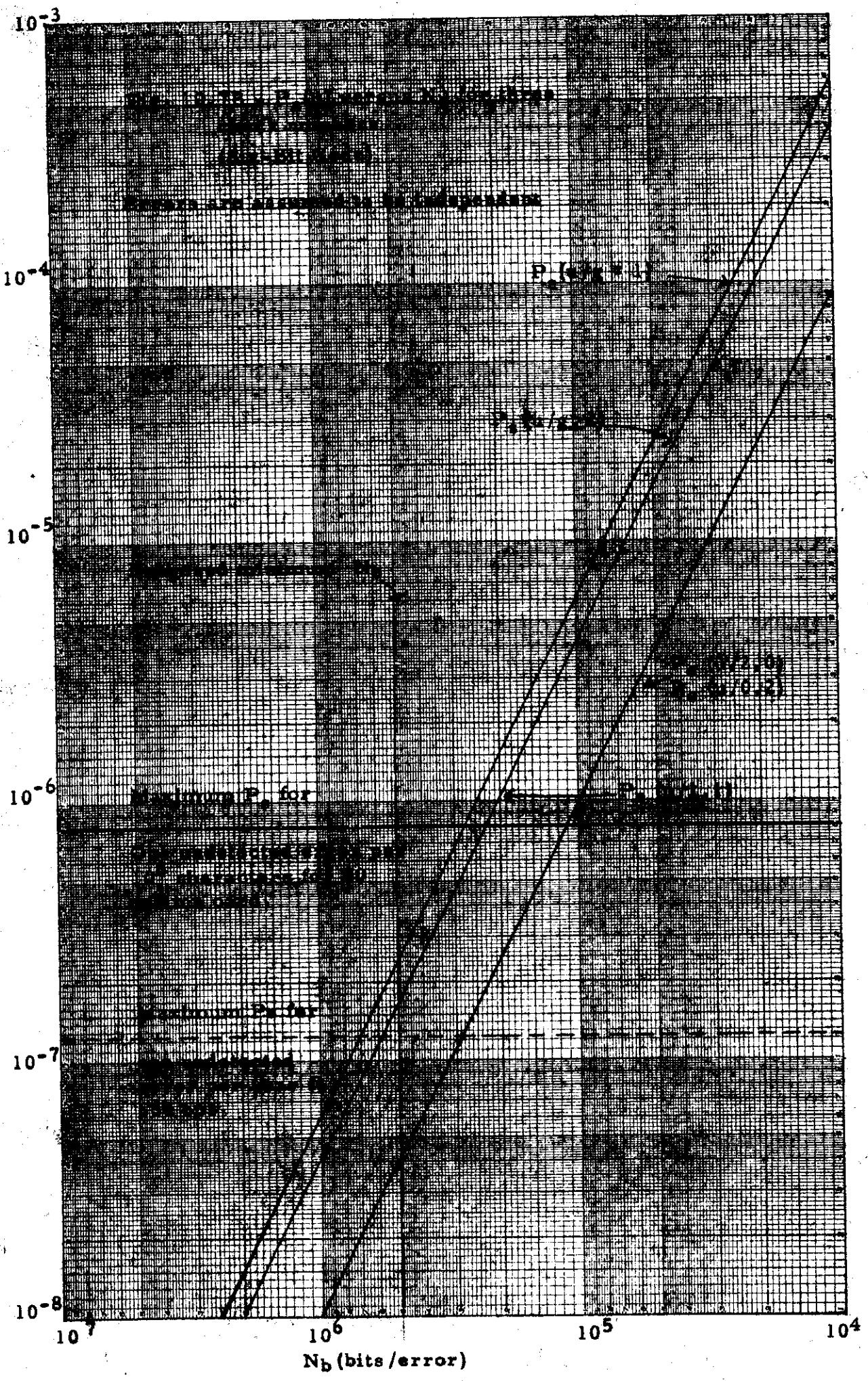
For other values of N_b , the probability of undetected errors for three longitudinal check numbers ($g = 1, 2, \dots$) can be read off of the curves of Fig. 10.7B. Independent errors are assumed.

TABLE VII

Undetected Errors in Six-Bit Code with Three
Longitudinal Block Check Bits

$n = 80$ characters $g = 1, 2$ $P_b = 10^{-4}$

Terms Not Detected		Ratio: $P_S^1(u/i, j)$	Terms in Probability of an Undetected Error
$g = 1$	$g = 2$	$P_S(u/i, j)$ or $\frac{\text{Case 3}}{\text{Case 2}}$	
X	X	2	$P_S(u/1, 1) = .000\ 4$
X			$P_S(u/2, 0) \quad .000\ 094$
X			$P_S(u/0, 2) \quad .000\ 094$
X	X	8	$P_S(u/2, 2) \quad .000\ 000\ 003$
X			$P_S(u/1, 3) \quad .000\ 000\ 002$
X			$P_S(u/3, 1) \quad .000\ 000\ 002$
X	X		$P_S(u/0, 4) \quad .000\ 000\ 000\ 5$
X	X		$P_S(u/4, 0) \quad .000\ 000\ 000\ 5$
X	X	32	$P_S(u/3, 3) \quad .000\ 000\ 000\ 000\ 022$



VIII Case Four - Multiplying Factor Due to Addition of Vertical Checking

For a six-bit code with three longitudinal check symbols, we add three interlaced vertical check symbols as shown in Fig. 10.7A. We now need the conditional probability, $P(B/A)$ where A is event of an undetected error passing through the longitudinal check system and B is the event of an undetected error passing through the vertical check system. Now the probability of an undetected error in a card (record) is:

$$P_r(u) = P(A, B) = P(B/A) P(A), \text{ from Eq. (5).}$$

$$P(A) \simeq P_s(u/g) \text{ from Table VII}$$

For symmetry take the record to cover $3 \times 26 = 78$ characters, leaving position 79 and 80 for the checking symbols*.

* This modification in Eq. (17) as follows:

$$\frac{P_r^1(u/l, 1)}{P_r(u/i, 1)} = \frac{468}{480} \times \frac{39}{40} = 0.95$$

This 5% reduction in the probability will be neglected in this section.

Let L = the number of bits in each string of bits of the longitudinal check.

Let D = the number of bits in each string of bits of the vertical check that intersect with a longitudinal check string.

For longitudinal checking only, the number of combinations of 1(0-1) and 1(1-0) errors is for $P_b(0-1) = P_b(1-0)$:

$$N(L) = C_1^{\frac{L}{2}} C_1^{\frac{L}{2}} \quad (29)$$

For vertical plus longitudinal checking the number of combinations of 1(0-1) and 1(1-0) errors is:

$$N(D/L) = C_1^{\frac{D}{2}} C_1^{\frac{D}{2}} \quad (30)$$

Then:

$$P(B/A) = \frac{N(D/L)}{N(L)} = \frac{C_1^{\frac{D}{2}} C_1^{\frac{D}{2}}}{C_1^{\frac{L}{2}} C_1^{\frac{L}{2}}} = \left(\frac{D}{L}\right)^2 \quad (31)$$

Sample calculation:

For the Fig. 10.7A: $L = 2 \times 78$, $D = 2 \times 26$

For $g = 2$, using Table VII, the first term of the probability of an undetected error is:

$$P_r(u/1,1) = P(B/A) P_s(u/1,1) = \left(\frac{26}{78}\right)^2 \cdot 0.0004 = .000044$$

The improvement in using the combination of three longitudinal with three vertical check numbers compared to six longitudinal check numbers is:

$$\frac{P_r(u/1,1)_{VIII}}{P_s(u/1,1)_{VI} \times 0.95} = \frac{.000044}{.00020 \times .95} = \frac{1}{4.5 \times .95} = \frac{1}{4.3}$$

This means that for a total of 480 bits distributed as 468 information bits and 12 checking bits, that using 6 bits for longitudinal checking and 6 bits

for vertical checking makes a reduction in the probability of undetected double errors by a factor of 4.3 compared to using the full 12 bits for longitudinal checking. Independent errors are assumed.

If we use the full 12 bits for vertical checking with $g = 2$ to make a consistent comparison, the leading term by Eq. (22), with $C = \frac{468g}{12} = 78$, is the same as the first term in table VI multiplied by 0.95. The results in the same probability of undetected errors as in Section VI. However, this system is more susceptible to dependent multiple errors. This analysis suggests an extension by putting these results in more general form and finding the ratio of D/L for minimizing the undetected errors.

IX Case Five - 7-Bit Code Without Longitudinal Check

Consider a 7-bit code for use in transmitting 80-column cards. The seventh bit is a redundant bit, so that the errors that can pass undetected are the same as shown in Fig. 10.2. For simplicity assume $P_b(0-1) = P_b(1-0)$. Since basic error rates can vary considerably the results will be plotted for a range of bit error rates. The principal terms are: $P(2,0)$, $P(1,1)$, and $P(0,2)$.

To accurately establish how these terms contribute to undetected errors, more information is needed on the detection system, noise spectrum, predetection filtering introduced by the multiplexing filters, and delay distortion. Without detailed knowledge of the detection system, only limiting values can be calculated for different conditions.

A. Independent Errors

Assume average number of bits is 4-1's and 3-0's. The terms are as follows from Eq. ()

$$P_c(2,0)_I = \frac{4!}{2!2!} P_b^2 (1-P_b)^2 (1-P_b)^3 \simeq 6 P_b^2$$

$$P_c(1,1)_I = 3P_b (1-P_b)^3 2P_b(1-P_b)^2 \simeq 6 P_b^2$$

$$P_c(0,2)_I = (1-P_b)^4 \frac{3!}{2!1!} P_b^2 (1-P_b) \simeq 3 P_b^2$$

$$P_c(u:2) = P_c(2,0) + P_c(1,1) + P_c(0,2) = 15P_b^2 \quad (32)$$

Comparing with a less exact analysis

$$P_c(u:2) = \frac{7!}{2! 5!} P_b^2 (1-P_b)^5 = \frac{6 \times 7}{2} P_b^2 = 21 P_b^2 \quad (33)$$

For 80 characters the probability of a record having an undetected error is:

$$P_r \approx 80 P_c(u:2) = 80 \times 15 P_b^2 = 1200 P_b^2 \quad (34)$$

Sample calculation for $P_b = 10^{-4}$

$$P_r(u) \approx 1200 \times 10^{-8} = .000012 = \frac{1}{83,000 \text{ cards}} \quad (34A)$$

Comparing with the general formula of Ref. 5.

2 errors in card $7 \times 80 = 560$ bits/card

$$P_x = e^{-560 \times 10^{-4}} \frac{(560 \times 10^{-4})^2}{2!} \approx 15.7 \times 10^{-4}$$

$$P_u = \frac{C_2^7 P_1^L}{C_2^L} = \frac{C_2^7 P_1^{80}}{C_2^{560}} = \frac{\frac{7!}{2! 5!} \frac{80!}{79!}}{\frac{560!}{2! 558!}} = \frac{6 \times 7 \times 80}{559 \times 560}$$

$$P_u = 1.34 \times 10^{-2}$$

$$P_u/P_x = P_u \cdot P_x = 1.34 \times 10^{-2} \cdot 15.7 \times 10^{-4} = 21 \times 10^{-6} \quad (34B)$$

This difference between the two methods poses a problem to be investigated. Is it necessary to treat (0 1) and (1 0) errors separately?

Considering the two types of errors gives:

$$P_r(u) \simeq .000012 \quad (34A)$$

Using the simple formulation of Ref. 5 which does not distinguish between the two types gives:

$$P_r(u) = p_u/p_x = .000021 \quad (34B)$$

This is a problem to be resolved.

When using the formulas of this paper, the probability of detected errors is:

$$P_r(d:1) \simeq 80 \times 7 P_b = 560 P_b = .0560 = \frac{1}{18 \text{ cards}}$$

B. Dependent Errors

$$P_c(2)_D \cong (a-1) P_b(b_k) P_b(b_m/b_k) \quad (35)$$

a = number of bits per character

$$\begin{aligned}
 P_c(2) &\simeq P \left(\begin{array}{c} P(c) \\ \text{two adjacent identical bits are changed} \\ \text{by the noise impulse} \end{array} \right) \cong P \left(\begin{array}{c} P(C/B) \\ \text{given one bit in} \\ \text{error the next bit} \\ \text{is in error} \end{array} \right) \\
 &\times P \left(\begin{array}{c} \text{a bit is the same as the} \\ \text{previous one} \\ P(B/A) \end{array} \right) \times P \left(\begin{array}{c} \text{there is a bit in} \\ \text{error in the character} \\ P(A) \end{array} \right) = \\
 &= P(C/B) P(B/A) P(A)
 \end{aligned}$$

As defined above, case (1) of Fig. 10.0 is implied in Eq. (35B).

$$P(A) = aP_b = 7P_b \quad (35C)$$

$$P(B/A) = \frac{2.74}{a} = \frac{2.74}{7} \quad (35D)*$$

$P(C/B)$ is not known, so plot curve for $P(C/B) = 0.1, 0.01$

$$P(C/B) = 0.01 \quad P_c(2) = 2.74 \times .01 P_b = .0274 \times 10^{-4} = 2.74 \times 10^{-6}$$

$$P(C/B) = 0.1 \quad P_c(2) = 2.74 \times 10^{-5}$$

$$P_r(u:2:01) = 80 P_c(2:01) = 80 \times 2.74 \times 10^{-6} = 2.20 \times 10^{-4}$$

$$P_r(u:2.1) = 80 P_c(2.1) = 80 \times 2.74 \times 10^{-3} = 2.20 \times 10^{-3}$$

* The factor 2.74 is the average number of sets of 00 or 11 in a character in the EDPM 7-bit code as determined by a sample count on p. 2-2.1.

X - Case Six

7-Bit Code with Longitudinal Block Check

In this case the full 7-rows have block check redundant bits. The exact analysis of this case is not complete.

Using the formulas of Schatzoff and Harding to get the order of magnitude, shows that for independent errors of one error in 10,000 bits as follows:

$P_R(u:2)$ = Probability that an undetected error consisting of two bits in error occurs in a record

$$P_R^d(u:2) = P(C/B) P(B/A) P(A) P(D/C) \quad (\text{dependent}) \quad (36)**$$

$P(B)$ = Probability that two bits are in error in a record

$P(C/B)$ = Probability that given two bits, in error, that a second set of two bits are in error

$P(B/A)$ = Probability that given an impulse noise pulse has occurred. It has an amplitude greater than A_B , sufficient to affect two bits.

$P(A)$ = Probability that one bit is in error.

$$P_R^i(u:2) = P(D/C) P^i(C) \quad (\text{Independent}) \quad (37)$$

$$P^i(B) = C_2^n [P(A)]^2 [1 - P(A)]^{n-2} \quad (\text{independent}) \quad (38)$$

*Formerly p 10-19 (8-1); 8-24

** Here a record is 80 characters or 560 bits.

$P_R(D/C)$ = Probability that gives four bits are in error in a record, that they are compensating.

$P_R^i(C)$ = Probability that four bits are in error in a record.

A. Independent Errors

For independent errors we can use the formulas of Schatzoff and Harding⁵

$$P_4 = P_r^i(C) = \frac{e^{-nP} (nP)^4}{4!} \quad (39)$$

The conditional probability that the error is undetected, is for four errors with vertical and longitudinal checking:

$$P_u = P^i(D/C) = \frac{C_2^7 C_2^2 \frac{P_2^L}{2!}}{C_4^{7L}} \quad (40)*$$

Same calculation for $P_b = P(A) = 10^{-4}$ given by (39) and (40) for

$$L = 80, x = 4, n = 7L = 560$$

$$P_p^L = L(L-1) \dots (L-p+1)$$

$$P_2^{80} = 80 \times 79$$

$$P_r^i(C) = \frac{C^{-.056} (.056)^4}{4!} = 4.1 \times 10^{-7}, \text{ by Eq. (39)}$$

$$P_r^i(D/C) = \frac{\frac{7!}{2!5!} \frac{2}{0!2!} \frac{80 \times 79}{2}}{\frac{560!}{4!556!}} = \frac{3.4.6.7 \times 80 \times 79}{555.558.559.560} =$$

$$\frac{.6}{555} \frac{42}{558} \frac{79}{559} \frac{80}{560} = 1.09 \times 7.5 \quad 14.1 \quad 14.3 \times 10^{-8} = 1650 \times 10^{-8}$$

$$= 16.5 \times 10^{-8} \text{ by Eq. (40).}$$

6. same as Ref. 5.

* P_2^L = permutation of 2 in a set of L

$$P_2^L = C_2^L$$

$$P_R^i(D) = P_R^i(D/C) P_R^i(C) =$$

$$p(u/x) = p_u \cdot p_x = 16.5 \times 10^{-6} \times 4.1 \times 10^{-7} = 6.7 \times 10^{-12}$$

No. years per undetected error at 700 bits per second:

$$Y_e(\text{years/error}) = \frac{C_c(\text{char/card})}{C_y(\text{char/year}) P_{u/x}(\text{error/card})} = \frac{80}{7.5 \times 10^{-8} \times 6.7 \times 10^{-12}}$$

$$= 16,000 \text{ years}$$

B. Dependent Errors

Using Eq. (36), let $P(A) = 10^{-4}$, let $0.01 P(B/A) = .01$ and consider

impulse noise pulses of Cases (1) of Fig. 10.0 (1), so $P(E/B) = 5/16$

Then the range of $P(E)$ is: $P_b(E) \text{ min} = \frac{5}{16} \times .01 \times 10^{-4} = 0.313 \times 10^{-6}$

Per character: $P_C^1(E) = (n-1) P_b(E) = (7-1) \times .313 \times 10^{-6} = 1.88 \times 10^{-6}$ (1st char)

Second character*: $P_C(E) = 2P_b(E) = .626 \times 10^{-6}$

Per record:** $P_R(u) = C \frac{L}{2} P_C^1(E) P_C^1(E) = \frac{79.80}{2!} 1.88 \times 10^{-6} \times .626 \times 10^{-6} =$

$$370 \times 10^{-12} = 3.7 \times 10^{-10}$$

The number of undetected errors per year is:

$$Z = \frac{7.5 \times 10^8 \text{ char/yr}}{80 \text{ char/card}} \times 3.7 \times 10^{-10} \frac{\text{error}}{\text{card}} = 0.35 \times 10^{-2} = .0035 \text{ year}^{-1}$$

$$= \frac{1 \text{ error}}{286 \text{ years}} \text{ for } P(B/A) = 0.01, P(E/B) = 5/16$$

$P(E/B)$ = Probability that given two bits in error in a character, that they are compensating errors which would not be detected by a vertical check. This corresponds to $P(C/B)$ of Table V-B, page 10-8.4

*Note: The above is based on Eq. (10). The second character double error bits must be in the same rows and either identical or opposite to the errors in the first character.

** The coefficient $C \frac{L}{2} = \frac{80.79}{2}$ is derived by saying the first double error

can occur in any of the eighty characters. The average character position would be the middle, or position 40. This leaves on the average $80/2 = 40$ positions in which the second double error can occur.

If $P(B/A) = 0.1$

Then $P_b(E) = 0.313 \times 10^{-5}$

$$P_c^{1*}(E) = 1.88 \times 10^{-5}$$

$$P_c^{11*}(E) = 0.626 \times 10^{-5}$$

$$P_r(u) = 3.7 \times 10^{-8}$$

$$Z^{1*} = \frac{1 \text{ undetected error}}{2.86 \text{ years}}$$

Comparison with analysis*, assumes $P(C/B) = 1$ instead of $5/16$, and $P(B/A) =$

$$P_r(u) = C_2^L P_c^1(E) P_c^{11}(E) = C_2^L (n-1) P_b(E) 2 P_b(E) =$$

$$C_2^L a(n-1) P_b(E)^2 = 2C_2^L (n-1) P(E/B) P(B/A) P(A)^2 \quad (41)$$

$$P_r^s(u) = 2 \frac{80!}{2!78!} (7-1) [1 \times .02 \times 10^{-4}]^2 = 79 \times 80 \times 64 \times 10^{-12}$$

$$\left[\text{For } P(E/B) = 1, P(B/A) = .02 \right] = 4 \times 3.75 \times 10^{-8} = 15 \times 10^{-8}$$

$$Z^s = \frac{7.5 \times 10^8 \text{ char/year}}{80 \text{ char/record}} \quad 15 \times 10^{-8} \text{ error/record}$$

$$= \frac{7.5 \times 3.75 \times 4 \text{ undet. error}}{80 \text{ year}} = 4 \times 0.352 = 1.4 = \frac{1 \text{ undetected error}}{0.707 \text{ years}}$$

Comparison with a previous calculation

*The analysis of Page 8-26 used a correction factor:

$$p(i, i+1) = p(i+1, i) \quad p(i) = S p_0^2$$

$$S = \frac{bn}{a^2} \left\| P_r^d(D) = S^2 P_i^d(D) = (200)^2 6.7 \times 10^{-12} \text{ where } S = \frac{2}{100} \times \frac{10\%}{100} = 200 \right.$$

$$P_r^d(D) = 26.8 \times 10^{-8}$$

The years per undetected error is:

$$Y_e = \frac{C_c}{C_y P_r^d(D)} = \frac{80 \text{ (char/card)}}{7.5 \times 10^8 \text{ (char/year)} 26.8 \times 10^{-8} \text{ (error/card)}} =$$

$$= 0.40 \text{ (year/error)} = \frac{1 \text{ year}}{2.5 \text{ errors}}$$

XI NOTE ON ERROR CORRECTING CODE FOR REDUCING NUMBER OF MESSAGE REPEATS ON HUMAN INPUT

Previous studies have dealt with setting criterion for the allowable undetected errors. ¹ In a machine operated system or a batch system, a high rate of detected errors can be tolerated. In an inline system where an operator keys in questions and waits for answers, an error detection system may not be adequate because an operator may become irritated at having to repeat an inquiry when an error is detected. Single error correction with double error detection codes have been described by Hamming² and others.³

To construct such a code for alphanumeric characters, six information bits plus C checking bits are required. The formulas of reference (2) ~~and~~ ~~indicated~~ (3) indicate C = 5. ~~This~~ This means an eleven bit code will provide single error correction plus double error detection.

The probability of characters being in error are for independent and dependent error rates. The eleven bit code:

Independent:

$$P_C(2) = C^{11} P_B^2 (1-P_B)^9 \approx 55 P_B^2 \quad (51)$$

Dependent: P(B/A) is the conditional probability.

Event A is the occurrence of one error in a character.

Event B is the occurrence of a second error in a character.

$$P_C(2) = C^{11} P_B (1-P_B)^{10} P(B/A) \approx 11 P_B P(B/A) \quad (52)$$

1. J. A. McLaughlin's memorandum on Allowable Error Frequencies in Data Transmission, 8/28/57

2. R. W. Hamming "Error Detecting and Error Correcting Codes" BSTJ 26 147 (1950)

3. Peter Elias "Error-Free Coding" Trans. I. R. E. PGIT-4, p29 (Sept. 1954) See bibliography in this article and other articles in same issue of Trans. I. R. E.

* For C = 4 it appears that partial double error detection is included.

** J. A. McLaughlin has proposed an 11-bit single error detecting code which is compatible with the 7-bit magnetic taper codes.

Sample calculation for $P_b = 10^{-4}$:

<u>P(B/A)</u>	<u>$P_c^{(2)}$</u>	<u>Curve</u>	<u>C_o (Characters/double error)</u>
P(B) = P(A)	$55 \times (10^{-4})^2$		1.8×10^6
.02	$11 \times 10^{-4} \times .02$	A	4.55×10^4
.10	$11 \times 10^{-4} \times .10$		9×10^3

Next we consider the probability of a word having to be repeated because of a double error being detected. The average input message size described in reference (4) is about twenty alphanumeric characters. For $m = 20$ characters in a record having a double error is:

$$P_r = m P_c (1 + P_c)^{m-1} \quad m P_c \quad (10-3)$$

Sample calculations:

<u>P(B/A)</u>	<u>$P_c^{(2)}$</u>	<u>P_r</u>	<u>R_c ($\frac{\text{No. of Rec.}}{\text{Rec. in Error}}$)</u>	<u>Curve</u>
P(B)=P(A)	5.5×10^{-7}	1.1×10^{-5}	9×10^4	B
.02	2.2×10^{-5}	4.4×10^{-4}	2.27×10^3	C
.10	1.1×10^{-4}	2.2×10^{-3}	450	D

Comparison with 4-out-of-8 code

For comparison the probability of a repeat order in the 4-out-of-8 code is calculated:

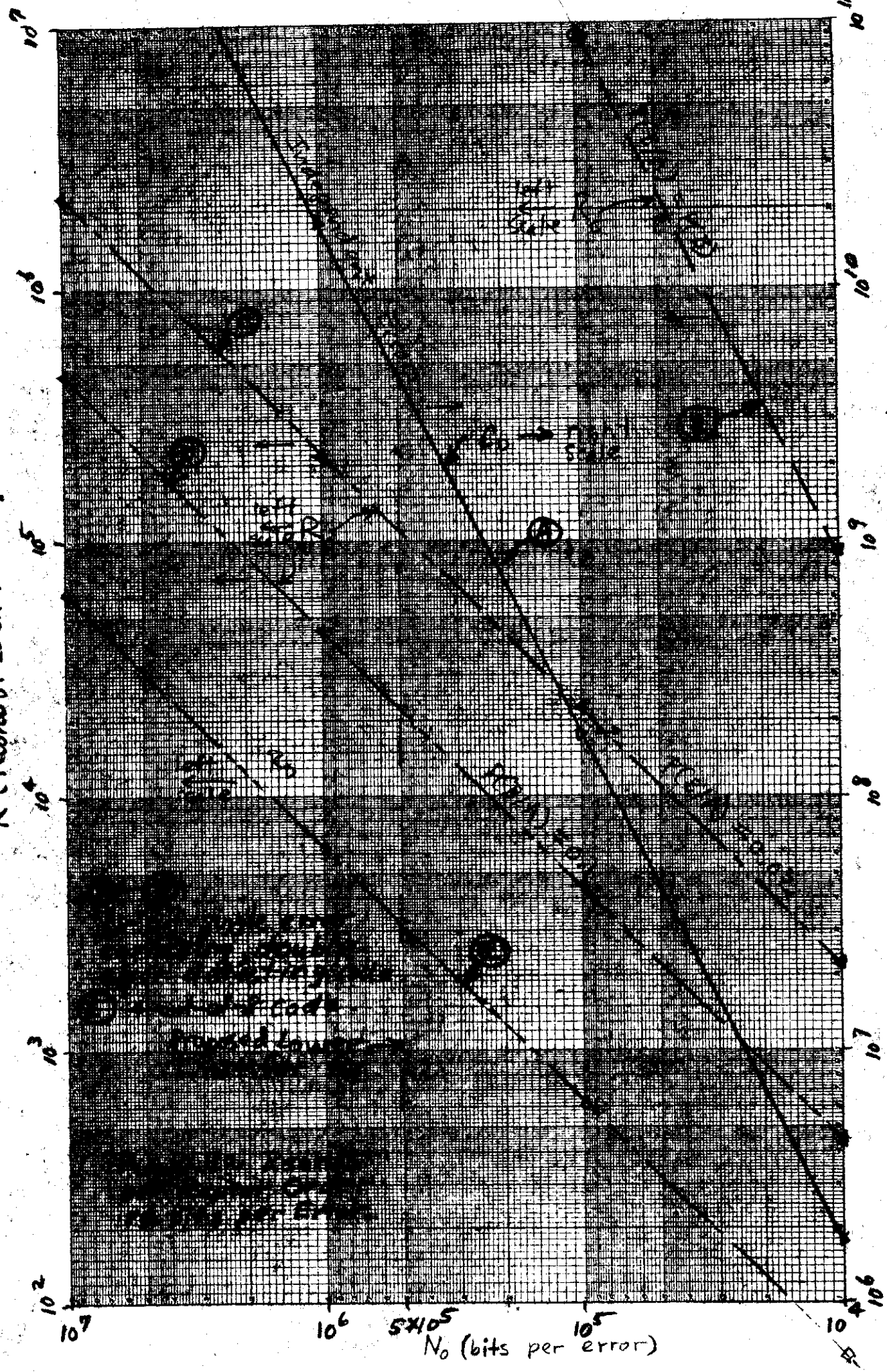
$$P_c^{(1)} = n P_b = 8 \times 10^{-4}$$

$$P_r = 20 \times 8 \times 10^{-4} = .0160 \quad R_D = (P_r)^{-1} = 62.5$$

This result is used to plot curve E.

10-3-57 HBW

R (records of 20 char each per double error)



C₀ (characters per double error)

No (bits per error)

**COMPARISON OF DIFFERENT SYSTEMS AT THE PROPOSED
STANDARD OF 5×10^5 BITS PER ERROR**

Code	Conditions	Records per Repeated Record (20 characters)
11-Bit SEC, DED	Independent Errors Only	$R_C = 2 \times 10^8$ Records
11-Bit SEC, DED	Dependent Errors $P(B/A) = .02$	$R_C = 1/1 \times 10^5$ Records
11-Bit SEC, DED	Dependent Errors $P(B/A) = .10$	$R_C = 2.2 \times 10^4$ Records
4-out-of-8 SED, half DED	Only Single Errors Counted in This Approximation	$R_D = 320$ Records

F. B. Wood
October 4, 1957

Comparison of 7-Bit and 4-out-of-8 Codes

A Independent Errors (Symmetrical)

(1) 7-Bit Code (p. 10-21.4)

$$P_c(u/2) = C_2^7 P_e^2 (1-P_e)^5 \quad (60A)$$

$$P_c(u/4) = C_4^7 P_e^4 (1-P_e)^3 \quad (60B)$$

$$P_c(u/6) = C_6^7 P_e^6 (1-P_e)^1 \quad (60C)$$

$$P_c(u) = \sum_{k=2,4,6} P_c(u/k) \approx 21 P_e^2 + 35 P_e^4 + 7 P_e^6 \quad (61)$$

$$\text{For } P_e \geq 10^{-4}: P_c(u) \approx 21 P_e^2 \quad (62)$$

(2) 4-out-of-8 Code (p. 10-8)

$$P_c(u/1,1) = \{C_1^4 P_e [1-P_e]^3\}^2 \approx 16 P_e^2 \quad (63A)$$

$$P_c(u/2,2) = \{C_2^4 P_e^2 [1-P_e]^2\}^2 \approx 36 P_e^4 \quad (63B)$$

$$P_c(u) = \sum P_c(u/c,i) = 16 P_e^2 + 36 P_e^4 \approx 16 P_e^2 \quad (64)$$

(3) Comparison (Transition from 7-bit to 4/8)

$$\Delta e = \frac{21 P_e^2 - 16 P_e^2}{21 P_e^2} = -\frac{5}{21} \quad \Rightarrow \quad -24\% \text{ change in undetected error rate}$$

Paid for by decrease of character rate of $-\frac{8-7}{7} = -\frac{1}{7} \Rightarrow -14\%$ in information rate.

Note: By hypothesis "B"

$$\Delta e' = \frac{21 P_e^2 \times \frac{1}{4} - 16 P_e^2 \times \frac{7}{128}}{21 P_e^2 \times \frac{1}{4}} = \frac{\frac{21}{4} - \frac{16 \times 7}{128}}{\frac{21}{4}} = \frac{21 - 16 \times 7/16}{21} = \frac{21 - 16 \times 0.7}{21} = \frac{21 - 11.2}{21} = \frac{9.8}{21} = 0.467 \approx 46.7\%$$

or no loss in undetected error rate to $(\frac{1}{4})$.

(4) Unsymmetrical Errors

	$\frac{P_e(0-1)}{P_e(\text{total})}$	$\frac{P_e(1-0)}{P_e(\text{total})}$	<u>Double Errors</u>		
(a)	$\frac{1}{2}$	$\frac{1}{2}$	00	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$	} $\frac{1}{2} = \frac{16}{3}$
			01	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$	
			10	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$	
			11	$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$	
				<u>$\frac{1}{4} = 1$</u>	
(b)	$\frac{1}{4}$	$\frac{3}{4}$	00	$\frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$	} $\frac{3}{8} = \frac{1}{3}$
			01	$\frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$	
			10	$\frac{3}{4} \cdot \frac{1}{4} = \frac{3}{16}$	
			11	$\frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16}$	
				<u>$\frac{16}{16} = 1$</u>	
(c)	$\frac{1}{8}$	$\frac{7}{8}$	00	$\frac{1}{8} \cdot \frac{1}{8} = \frac{1}{64}$	} $\frac{7}{32} = \frac{1}{3}$
			01	$\frac{1}{8} \cdot \frac{7}{8} = \frac{7}{64}$	
			10	$\frac{7}{8} \cdot \frac{1}{8} = \frac{7}{64}$	
			11	$\frac{7}{8} \cdot \frac{7}{8} = \frac{49}{64}$	
				<u>$\frac{64}{64} = 1$</u>	

Hypothesis "A":

Fraction of double errors which are compensating (used in 04-out-of-8 code)

Note these fractions are to used with:

$$C_2^8 P_e^2 (1-P_e)^6 \approx 28 P_e^2 \quad \text{for 4/8 code, } (6)$$

except note discrepancy between $\frac{1}{2} \cdot 28 P_e^2 = 14 P_e^2$ and the $16 P_e^2$ of 4-10-27. 16

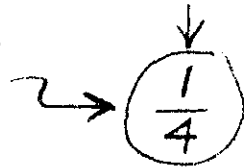
Should the above fraction be weighted?
 Consider revised table, etc.

Hypothesis "B":

Example of weighting calculation:

$\frac{1}{8}$	$\frac{7}{8}$	0 0	$\frac{1}{64} \times \frac{1}{4} = \frac{1}{256}$	$\left. \begin{matrix} \frac{1}{256} \\ \frac{7}{256} \\ \frac{7}{256} \\ \frac{19}{256} \end{matrix} \right\} \frac{14}{256} = \left(\frac{7}{128} \right)$
		0 1	$\frac{7}{64} \times \frac{1}{4} = \frac{7}{256}$	
		1 0	$\frac{7}{64} \times \frac{1}{4} = \frac{7}{256}$	
		1 1	$\frac{49}{64} \times \frac{1}{4} = \frac{49}{256}$	
				<hr/> $\frac{64}{256}$

Conditional probability due to lack of symmetry for 7-Bit code



Conditional probability due to lack of symmetry in 4/8 code

The above analyses are not consistent so far. Which is correct? Hypothesis "A" or "B"?

The definition: $\frac{P(\text{error})}{P(\text{total})}$ may be a source of trouble, i.e. these fractions may have to be redefined as: $P(\text{error}/0, \text{noise})$ and $P(\text{error}/1, \text{noise})$.

Note: The discrepancy of $10-21.4 / 10-21.5$ giving $P_c(u/2) = 15 P_A^2$ (32)

and $P_c(u/2) = 21 P_A^2$ (33) has been resolved, equation (33) is correct. FEW

B. Dependent Errors

(1) Noise Model

$P_A(A_0)$ = Probability that a noise impulse of amplitude $\geq A_0$ occurs in a bit time.

T = Bit period in seconds

α = Decay constant of envelope of noise.

$\alpha \approx \frac{\omega_0}{2Q}$, where ω_0 is $2\pi f_0$ of transmission circuit and Q is effective Q of transmission circuit.

Let $\beta = e^{\alpha T}$, then if A_1 is the noise signal amplitude required to cause an error in one bit, then the amplitudes to cause multiple errors in adjacent bits are:

No. of bits	Amplitude	Probability
1	A_1	C_1
2	$A_2 = \beta A_1$	$2k C_1$
3	$A_3 = \beta^2 A_1$	$3k^2 C_1$
4	$A_4 = \beta^3 A_1$	$4k^3 C_1$
n	$A_n = \beta^{n-1} A_1$	$n k^{n-1} C_1$

V/L indicates not yet used in this analysis.

C_1 is the fraction of the total bit errors which are single independent errors.
 k is a number, $k < 1$

$$P_A(A_0) = \sum_n n k^{n-1} C_1 P_n$$

$$\frac{1}{C_1} = \sum_n n k^{n-1} = \frac{1}{(1-k)^2}$$

See 758 (1929 ed)
 (67)

C_1	h
.33	.425
.50	.293
.90	.050
.99	.005

For $C_1 = .33$ the unweighted fractional weights of errors are:

$$C_1 = .33 \quad C_2 = hC_1 = .425 \times .33 = .14 \quad C_3 = hC_2 = .06$$

$$C_4 = .025 \quad C_5 = .017 \quad C_6 = .0046$$

$$\sum m C_m = \begin{array}{r} .33 = .33 \\ 2 \times .14 = .28 \\ 3 \times .06 = .18 \\ 4 \times .025 = .10 \\ 5 \times .017 = .085 \\ 6 \times .0046 = .0276 \\ \hline .968 \end{array}$$

(2) 4/8 Code = Undetected Errors in Block

$$P(u/2, A_2', A_2'') = C_2^m [P(u/2)] P(A_2'', A_2') \quad (68)$$

$m =$ number of characters in a block

$$P(u/2) = 2 \frac{P_b^{(0-1)} P_r^{(1-0)} g}{P_r^2} \text{ if hypothesis "A" is used.}$$

$g = 1$ for "A" $g = \frac{1}{4}$ for "B" (Divide by 4 if Hypothesis "B" is used)

$$P(A_2'', A_2') = P(A_2'') P(A_2')$$

$$P(A_2') = 2 h C_1 (m-1) P_r [1 - P_r]^{m-1}$$

$$m = 8$$

$$P(A_2'') = 2 h C_1 P_r [1 - P_r]^{m-1}$$

↑ These 2's normalize double errors to a per bit basis (~~bits~~)

$$P(u/A_2'', A_2') =$$

$$= C_2^m \left[\frac{2 P_A(0-1) P_A(1-0) g}{P_b^2} \right]^2 4^{(m-1)} k^2 C_1^2 P_A^2 [1-P_A]^{2m-2} \quad (69)$$

Sample Calculation: Let $g=1$ $P_b=10^{-5}$ $m=8$
 $C_1=.33$ $k=.425$ $m=1000$
 $P_A(0-1)/P_A = 1/4$

$$P(u/A_2'', A_2') = \frac{1000!}{2! 998!} \left[2 \cdot \frac{1}{4} \cdot \frac{3}{4} \right]^2 4^{(8-1)} \cdot .7 \cdot .425^2 \cdot .33^2 \cdot 10^{-10} \approx$$

"record"

$$= 0.5 \times 10^6 \cdot \frac{9}{64} \cdot 28 \cdot .0116 \times 10^{-10} = .0316 \times 10^{-4}$$

$$1 \times P_{\text{rec}}(u/4) = 3.86 \times 10^{-6}$$

$$P_{\text{char}}(u/4) = P_{\text{rec}}(u/4) / m = 3.86 \times 10^{-9} = 0.39 \times 10^{-8}$$

↑
4 bits in 2 char.

The true $P_{\text{char}}(u) = 2 P_{\text{char}}(u/4) = 0.8 \times 10^{-8}$ Hyp¹
 For next term see later section

(3) 7-Bit Code = Undetected Transmission Block (1st term)

$$P(u/2) = g \quad \text{from pp 10-27.1 / 27.2 hyp "A" & "B"}$$

C_2^m is the same

$P(A_2'', A_2')$ is the same except $m=7$ instead of 8

$$P(u/A_2'', A_2') = C_2^m g^2 4^{(m-1)} k^2 C_1^2 P_A^2 [1-P_A]^{2m-2} \quad (70)$$

Sample Calculation: same conditions as above, except $m=7$

$$P(u/4)_{\text{rec}} = 0.5 \times 10^6 \cdot 1 \times 24 \cdot .425^2 \cdot .33^2 \cdot 10^{-10} = .236 \times 10^{-4}$$

$$= 23.6 \times 10^{-6}$$

$$P_{\text{char}}(u/4) = P_{\text{rec}}(u/4) / m = \frac{23.6 \times 10^{-6}}{7} = 3.37 \times 10^{-6}$$

2 4 bits in 2 char

$$\text{True } P_{\text{char}}(u) = 2 P_{\text{char}}(u/4) = 6.74 \times 10^{-6}$$

$$\text{Ratio } \frac{P_{\text{char}}(u/7\text{-Bit})}{P_{\text{char}}(u/4/8)} = \frac{6.74 \times 10^{-6}}{0.8 \times 10^{-8}} = 8.425$$

C. Notes on Error Probabilities

(1) Criterion
If $P_c \leq 10^{-8}$ is specified, then what is the max. P_r ?

If an undetected error is caused by a single character then

$$P_r = n P_c \quad (71)$$

If it is caused by a minimum of two characters, then:

$$P_r = \frac{n P_c}{2} \quad (72)$$

This may be more complicated when higher terms are considered.

For $n = 1000$, $P_c = 10^{-8}$, using eq (72):

$$P_r \leq \frac{1000 \times 10^{-8}}{2} = 5 \times 10^{-5} = .00005$$

(2) Changing Block Length

Consider using the 8th bit used in converting the 7-bit code to 4-out-of-6 code as additional longitudinal check.

To transmit in 7-bit code with every 8th character, a check character.

$$n' = 7$$

$S =$ the number of sub-140-bits.

$$S = \frac{n}{n'+1} = \frac{1000}{8} = 125$$

The factor C_2^{1000} in the probability of undetected errors is now changed to:

$$C_2^n \rightarrow S C_2^{n'} \quad (73)$$

Hypothesis "A"

10⁻²

10⁻³

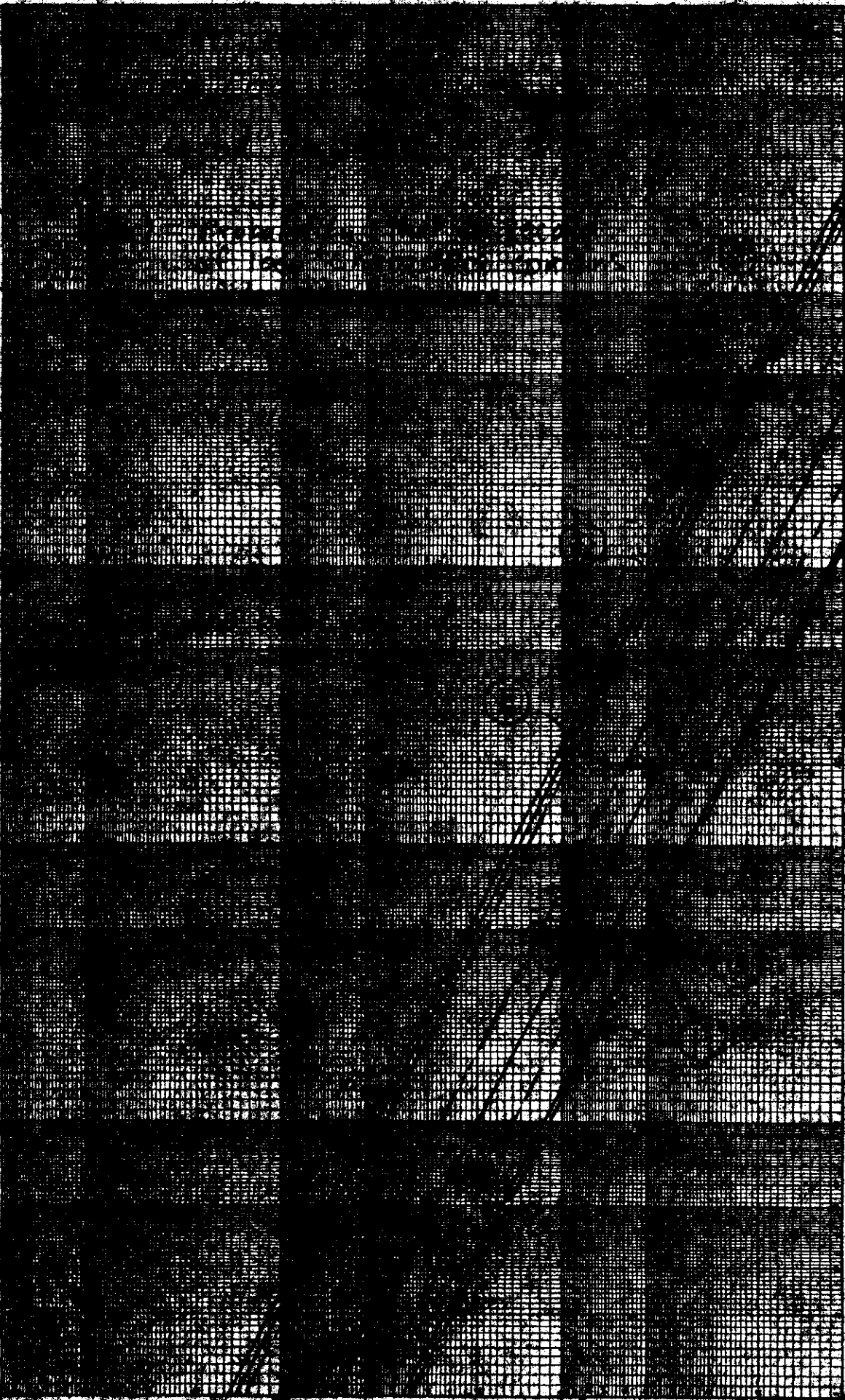
10⁻⁴

P(0)

10⁻⁵

10⁻⁶

10⁻⁷



Nb. (pulses/error)

10⁴

Sample calculation:

$$C_2^{m'} / C_2^m = 5 C_2^{m'} / C_2^m = \frac{\left(\frac{125 \cdot \frac{7!}{2 \cdot 5!}}{1000'} \right)}{\left(\frac{2 \cdot 998!}{1000'} \right)} = \frac{125 \times 21}{998 \times 500}$$

$$= .00525$$

For $P_0 = 10^{-4}$, 7 Bit Code: $P_1(u) = .00525 \times 238 \times 10^{-6}$
 $= 12.5 \times 10^{-8}$

(3) Interlocking or Transmission by Rows

If the transmitting core buffer is read out by rows, so that successive bits in the serial transmission are from different characters the probability of undetected errors in a character become independent.

The probability of an undetected error becomes: (independent)

$$P(u) = C_2^m \left[\frac{2 P_b(1-P_b) P_b(1-P_b)}{P_b^2} g \right]_{m=8}^{(m-1)} \left[\frac{P_b^2 (1-P_b)^{m-2}}{P_b^2} \right]^2 \quad (74)$$

$$P(u) = C_2^m \left[g \right]_{m=7}^{(m-1)} \left[\frac{P_b^2 (1-P_b)^{m-2}}{P_b^2} \right]^2 \text{ for 7-bit } (75)$$

* Correction substitute for $(m-1)$

in eq (74): $[C_1^4]^2 = 16$

in eq (75): $C_2^7 = 21$

Let $g = 1$, $\frac{P_b(1-P_b)}{P_b} = \frac{1}{2}$, $\frac{P_b(1-P_b)}{P_b} = \frac{1}{2}$

Sample calculations:

$$4/8 : (74)$$

$$P(u) = 0.5 \times 10^6 \times \frac{1}{4} \times 16 (10^{-4})^4 = 2 \times 10^{-10}$$

As correct below

$$1.8 \times 10^{-10}$$

$$7-8.2 : (75)$$

$$P_r(u) = 0.5 \times 10^6 \times 1 \times 21 (10^{-4})^4 = 10.5 \times 10^{-10}$$

$$9.95 \times 10^{-10}$$

These can be corrected for multiple errors, approximately by multiplying by:

$$\left(\frac{1}{C_1}\right)^2 \text{ of } 29 (67)$$

or $C_1 = 0.33$ multiply the above by $3^2 = 9$

(A) Unrelated Errors with N Block Tests (Imitating

$$P(u) = C_1^m \left[\frac{2P_1(1-P_1)^{m-1}}{P_1^2} \right] 16 P_1^2 \quad 4/8 \quad (76)$$

$$P(u) = C_1^m [\quad] 21 P_1^2 \quad 1 \quad (77)$$

Sample Calculation:

$$4/8 : P(u) = 1000 \times \frac{1}{2} \times 16 (10^{-4})^2 = 8 \times 10^{-5} \quad \left| \begin{array}{l} \times 9 \\ 72 \times 10^{-5} \end{array} \right.$$

$$7 : P(u) = 1000 \times 21 (10^{-4})^2 = 21 \times 10^{-5} \quad \left| \begin{array}{l} \times 9 \\ 189 \times 10^{-5} \end{array} \right.$$

J. B. Wood
November 27, 1957

11-27-51

Summary of Comparison of 7-Bit and 4-out-of-8 Codes

Methods of coding information for serial transfer between 1000 character core-buffers are considered. Information is received from magnetic tape in 7-bit code and is stored in a 7-plane 7-bit core-buffer.

The distribution of multiple errors is assumed to be:

	Fraction of Total Errors	Weight	Weighted Fraction
Single	0.333	1	.333
Double (Adjacent)	0.142	2	.284
Triple	0.06	3	.181
Quadruple	0.025	4	.103
5-tuple	0.0107	5	.054
6-tuple	0.0046	6	.028
7-tuple	0.00195	7	.014
			<u>0.997</u>

Only the most significant terms have been retained in the calculations. Curves of $P_r(u)$ are plotted on page 10-32 as a function of the number of pulses per error.

$P_r(u)$ = Probability that a record (r) of 1000 characters contains undetected errors. (Note that this cannot be arbitrarily converted to the probability of a character in error, because under different conditions plotted, the leading term is 1 or 2 character

Curve (A) 7-Bit Code Transmission with One-Bit Block Check on each row of 1000 bits. This applies for both symmetric and non-symmetric probabilities of $0 \rightarrow 1$ and $1 \rightarrow 0$ errors.

Symmetry defined on page 10-27.1

Curve (B) 4-out-of-8 code with one-bit per row block check inserted before translation to 4/8 code. 1000 character blocks.
 $P(0 \rightarrow 1) - P(1 \rightarrow 0) : \frac{1}{2} - \frac{1}{2}$ i.e. Symmetric.

Formulas for (A) and (B) are eqs (70) and (69).
 There is an ambiguity in the definition of the non-symmetric ratios which is lumped in the factor "g". In comparing the 7-Bit code and the 4/8 code under comparable conditions the g or g² terms cancel out in the comparison.

Curve (C) 7-Bit Code, ^{check} No Block Check, Interlaced by a simple read-out and write-in of core buffers on the data link side by rows, while the original input and final output is by columns (characters).

The check point is from eq (77) p. 10-34.

Curve (D) 4-out-of-8 code, No Block Check, Interlaced as in curve C, but translated to 4/8 code for transmission.

Curve (E) 4-out-of-8 code, Block as in curve B, except
 $P(0 \rightarrow 1) - P(1 \rightarrow 0) : \frac{1}{4} - \frac{3}{4}$ i.e. Non-symmetric.

Curve (F) Limiting Curve for 7-Bit code with special interlacing patterns. Eq (77) without multiple error factor. No block check.

Curve (G) 4-out-of-8 code, Block check as in curve B, except
 $P(0 \rightarrow 1) - P(1 \rightarrow 0) : \frac{1}{8} - \frac{7}{8}$.

curve (H) Example of Sub-Block Checking: The same number of redundant check bits as are added in translating from 7-Bit to 4/8 code are used to provide block checking every 7-characters within the 1000 character message. This curve applies for both Symmetric and non-Symmetric error probabilities

F B Wood
11-27-57

10.13 Review of Work Related to
Data Processing Problem 06-062-00
"Simulation Study of Communication Networks"
Paragraph 65. Optimum codes as a
function of

(1) Bit Rates:

Experimental error rates have been measured with an experimental modulation-demodulation system as a function of bit-rate for random thermal noise and impulse noise. (E. Haysner, IBM Jour Trans p 84). Curves of transmission efficiency and optimum block length for 7-bit code as a function of bit-rate have been calculated for fixed error probabilities. (F. B. Wood, AIEE Preprint 55-1181, Fig. 8 and 9.

(2) Types of transmission media:

Intra-plant cables: Studies of the primary cable constants, secondary cable constants, and characteristic impedance as a function of frequency have been made. The approximate bandwidth and channel capacity have been calculated as function of length. (F. B. Wood, RJ-134 plus additional notes which are being

edited for reports, sec. 11)

Local loops: Data published by non-ITM sources has been used to find distribution of line losses (F.B. Wood, notes* p. 11-89.2)

N-Carrier: Impulsive noise on a noisy line has been compared with other possible noise sources. (F.B. Wood notes* p. 22-6).

Microwave: The equivalent noise due to fading and equipment failure for extreme cases reported by Bell Telephone Laboratories is plotted for comparison with other noise sources affecting data transmission. (F.B. Wood notes*, sec 22).

(3) Class of Application

The probability of undetected errors for a few different codes have been plotted for different message lengths and other restrictions. (F.B. Wood, notes*, sec. 10)

This study was directed to the evaluate the hypothesis that error detecting with transmission could be tolerated on both transmission.

* These notes are in the process of being converted to memorandum reports.

while real time computing might require error correctors to prevent operators from blaming system for errors.

(4) Error frequency distribution by class of Error:

Curves of the probability of undetected errors as a function of bits per error for different ratios of double to single errors have been calculated for different codes. (S.B. Wood, note, Sec 10)

Dr. N.M. Abramson, Stanford University, has developed a SEC-AEC code for us. (IBM Patent is nos 80779, 80780 and Stanford Electronics Laboratories, Tech Report No 51, Dec 30, 1958). The logic to implement this class of codes is being studied.

Mr. Dangle of AEC has advised us that statistics on noise distribution on telephone lines will be available at the AIEE Summer Meeting.

(5) Error Correction Codes - Performance of Data and Error:

A comparison of 7-bit and 10-bit codes is included in the optimum block length

analysis. (See also AFEF Program 58-1181,
figure 2).

Supplement to (2):

Tables of inter-plant cable material
and installation costs have been prepared
(F.B. Wood 47-OR-523-020)

F.B. Wood
2/20/59

SAN JOSE LABORATORY
ADVANCED SYSTEMS DEVELOPMENT

August 12, 1959

FILE MEMORANDUM: 5720-10.14

Problem: What is the optimum error-detection system for single character, 10 character, 80 character, 100 character, and 1000 character block lengths in data transmission with decision feedback?

Assume six information bits per character, and allow two redundant bits per character. Consider different combinations of vertical and horizontal check bits.

For example, take an 80 character block of 640 bits. These bits could be used alternatively in the following cases:

- (1) 4-out-of-8 code, 8 bits/character, 80 characters.
- (2) 7-bit code vertical check, 7 bits/character, 80 characters. Eleven block check characters, distributed as 8 sub-block check character in sub-block of ten characters, each with 3 super-block check characters.
- (3) 4-out-of-8 code, 8 bits/character, 78 characters, 2 horizontal block check characters.
- (4) Other error-detection codes using the same amount of redundancy.

F. B. Wood
8-12-59

FBW:jp