

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.

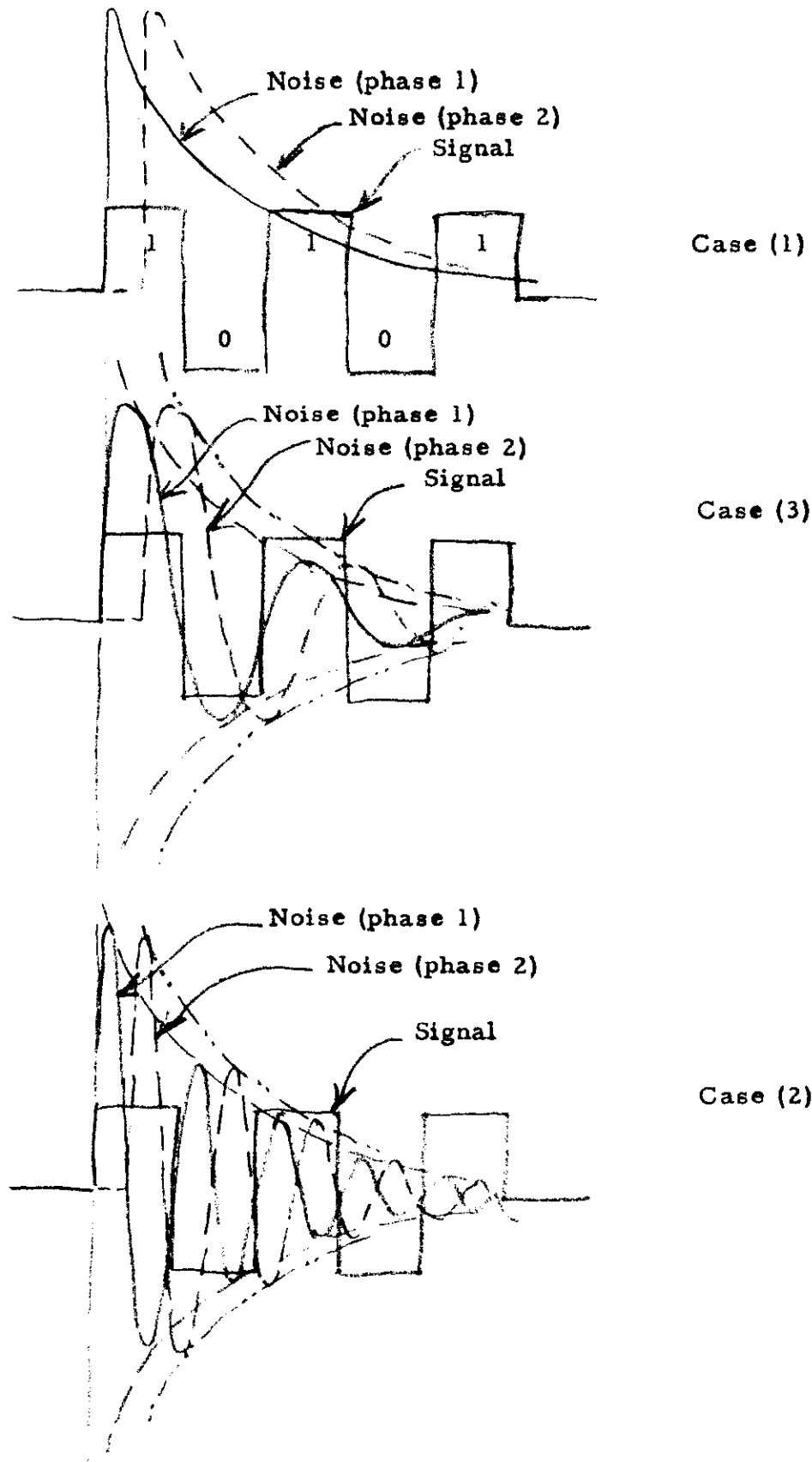


Fig. 10.0 Types of Double Errors Caused by Single, Long Noise Impulses

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.

