8.0 Analysis of Block Length in Error Checking

(References p. 7-11)

See pp 8-1 → 8-3 for general statement of the problem.

The optimum block length may be different for the different sections (sub-spaces of Fig 2 in SJC-12) of a business data system. The optimum block length may be different for:

a) human reading
b) human keying
c) data transmission
d) computer logic
e) machine reading of documents.

This study at present is limited to the data transmission area. To start out the optimum block length formula of A Cohen (AD-11919) will be analyzed (See p. 8-4).

AREA: Communications Research

TITLE: 8.1 Study of Error Detection and Correction

The objective will be to establish criteria for determining as a function of bit error rates: (1) the relative advantages of error checking codes and error correction codes, and the optimum block length for longitudinal checking.

An example of a system well protected against errors would be an eleven bit code which provides single-error correction and double error detection with 64 possible characters with a specified block length for a repeat signal. From the allowed signals and the probability of: (1) random noise, (2) impulse noise, and (3) accidental breaking of the line, the probabilities of single errors and double errors (including breaking of the line) could be estimated. Then the net transmission rate could be calculated by subtracting the time lost in waiting for the double error signal and then repeating the block having a double error.

A curve of transmission speed as a function of the bit-error probabilities and block length could be calculated. Then a similar curve for an error detecting code with repeating of blocks containing detected single errors could be calculated for comparison of the two systems. The cost per bit of information could be estimated from the cost of the different error detecting and correcting circuits, the cost of transmission at the specified error rates, and the transmission line rental.

Significance to IBM: This investigation would simplify the determination of the type of code needed for specified bit-error rates and give the optimum block length for block checking in data transmission systems.

F. B. Wood

FBW:hp

8. 2 / 8. 30 Optimum Block Length

This section is superseded by report:

RJ - DR - 532. 016, " Optimum Block Length for Data Transmission with Error Checking."

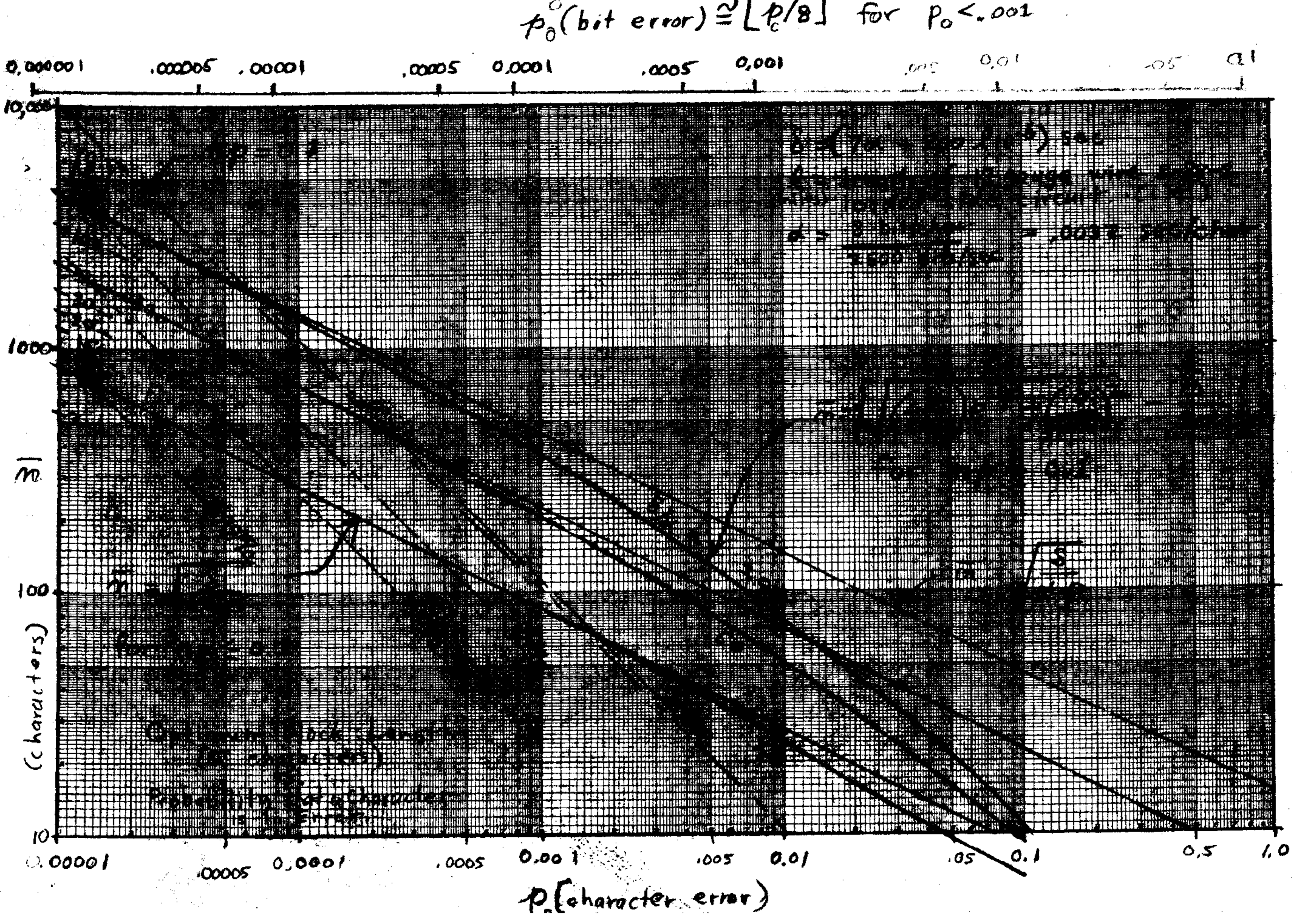
Sample calculations not used in the above report referred to in abbreviated form as "016," are contained in sections:

8. 4 Application of Block Length Formula

8. 5 Optimum Block Length for Experimental Line

8. 51 Efficiency Curves

6-13-57 FBW Rev. 7-17-57



8-1

8.6 - 8.12 Error Probability for Undetected Errors
(these sections moved to Sec. 10)

8.13 Datacom Calculation

For a Datacom batch transmission system using a 1000 character buffer and a four-wire data channel at 1000 bits per second a sample calculation has been made of the maximum error-rate acceptable. If the 4-out-of-8 code is used on a 3000 mile link, what is the acceptable error rate to operate at 95% efficiency?

The definitions used in the above question and the following analysis are taken as defined in Report RJ-DR-532-016.

Using procedure outlined on page 2 of Report 016, except that $\overline{E}$ and $\overline{n}$ are fixed, leaving P_c to be determined.

1. <u>Time Interval per Character:</u>

Eight bits per character at 1000 bits/sec gives

$$a = .008 \text{ sec.}$$

2. <u>Reply Time Delay:</u>

Since there are no echo suppressors on a four-wire link,

$$\delta = (k_1 + k_2 + k_3)\alpha + 2\ell/v + b$$

from Report 016, page 5.

Since synchronization time is negligible in a core buffer, $b = o$

Set $k_1 = o,\ k_2 = k_3 = 1$

We note in Report 016 that v varies from 8000 miles/sec for some loaded cables up to 128,000 miles/sec for non-loaded cable. For microwave links v approaches 186,000 miles/sec. For this example use a 19 guage H-44-S loaded cable having v = 20,000 miles/sec (RDRE p 822).

$$\delta = 2 \times .008 + \frac{2 \times 3000}{20,000} = .316 \text{ sec}$$

$$\delta/a = .316/.008 = 39.5 \simeq 40$$

3. <u>Optimum Block Length and Minimum Characters per Error:</u>

Using Report 016, Fig. 4, p. 12, finding intersection of 95% maximum efficiency and = 40 gives:

$\bar{n}$ - 1600 characters per block (graphically determined)

N_o - 4 5 x 10^5 bits per error

4. Determination of Efficiency at Non-Optimum Block Length n 1000:

Converting N_o = 4.5 x 10^5 bits per error to character error probability:

$$P_c = \frac{8}{N_o} - \frac{8}{4\ 5} \times 10^{-5} \quad .000018 \simeq 00002$$

Examining sample calculations of Report 016, p. 16 for same P_c, but slightly different other parameters, we then interpolate on Fig 6 near point A and find the efficiency at the non-optimum block length:

[n] = 1000

Efficiency $\simeq$ 94%

5. Examination of Undetected Error Rate:

Referring to p. 10-8 1 of some unpublished* notes, Table IV:

For N_o - 5 x 10^5; $P_b - \frac{1}{N_o} = 2 \times 10^{-6}$

For independent bit errors in the 4-out-of-8 code:

P_c (u) - 6 4 x 10^{-11} or,

there are 1 56 x 10^{10} characters per undetected error for a basic rate of 5 x 10^5 characters per single independent error.

What is not yet known is the conditional probability of double errors at 1000 bits/sec.

At 2500 bits/sec, the conditional probability of changing a "01" to a "10" has been found to be:

$$02 < P\ (b_{k+1}/b_k) < \ 10$$

At 1000 bits/sec, the conditional probability is expected to be at least an order of magnitude lower.

Since $P(b_{k+1}/b_k)$ is not known for 1000 bits/sec, the best we can do, before obtaining experimental data is to calculate what the maximum tolerable value is: Sampling 28 of the 54 characters used in the 1-2-4-7-R-R_1-O-X code of SJA-16, p 11 (also notes p 2-2), we find that the average number of bit positions susceptable to compensating double errors in adjacent bits is 3.5 per character. Experimental data at 2500 bits/sec. indicates it is the double errors in adjacent bits that are significant. Therefore, the probability of a double error changing an "01" to "10" or vice versa is:

$$P_c\,(b_{k+1}, b_k) = 7\,P_b\,\frac{3.5}{7}\;P\,(b_{k+1}/b_k)$$

Solving for $P(b_{k+1}/b_k)$ and taking $P_c\,(b_{k+1}, b_k) = 10^{-8}$ from the criterion set by Mr. J. A. McLaughlin gives as a limit:

$$P(b_{k+1}/b_k) = \frac{P_c(b_{k+1}, b_k)}{3.5\,P_b} = \frac{10^{-8}}{3.5 \times 2 \times 10^{-6}} \approx .0014$$

The detection system must be specified to give this conditional probability a physical reality.

6. Summary:

The result of the above analysis indicates that for a Datacom system as follows:

Buffer: n = 1000 characters
Code: 4-out-of-8
Speed: 1000 bits/sec
Efficiency: 94% (defined in Report 016)
Distance: 3000 miles; v = 20,000 miles/sec

The error rates must meet the following requirements:

$P_b\,(b_k) = 2 \times 10^{-6}$ (at least 500,000 bits/error)

$P_b\,(b_{k+1}/b_k) = 0.0014$ (for every 10,000 errors, there must be less than 14 double compensating errors which could change a "01" to a "10" or vice versa)

Double errors changing a "00" to a "11" and vice versa would be detected by the 4-out-of-8 code.

F. B. Wood
10/9/57
Revised 3/24/58

8:14 Comparison of Feedback With Error-Correcting Code (Problem)

Problem: Is grouping messages into blocks with a feedback signal from the error checking logic at the receiver more desirable than an error-correcting code?

Bishop and Buchanan have shown that the cost of decreasing the uncertainity by information feedback is greater than the cost of doing the same thing by means of redundancy in the one way path [1]

Our analysis have not gone as far as theirs for a comparative analysis We have dealt with "Optimum Block Lenth" in information feedback separately from the problem of "Undetected Errors " Our problem now is to synthesize these two parts of the problem

A tentative hypothesis for our problems is that: A finite but very small probability must be assigned to unusual events such as the channel being cut With this assumption I guess that there may be some cross-over point between redundant error-checking and information feedback

F B Wood
12-13-57

[1] Walton B Bishop and Bobby L. Buchanan. "Message Redundancy vs Feedback for Reducing Message Uncertainity." IRE Nat Conv. Rec , Vol 5, Part 2, pp 33 - 39, March, 1957.

8.15 - 8.16 Direct Derivation of Optimum Block Length

The result of these analyses are included in report:

RJ-MR-11 "Optimum Block Length for Data Transmission With Error Checking." February 28, 1958

8 17 Expectation of Number of Times a Message is Sent Due to Dependent Errors Due to Line Failure

This analysis has been replaced by a more general analysis made by J M Heyning.

F. B Wood
3/24/58

FILE MEMORANDUM: FBW-8.19

Comparison of Two Transmission Efficiency Analyses

After preparing (I) Report RJ-MR-11, February 28, 1958, (a revision of RJ-DR-532-016, September 20, 1957. derived from page 8 - 11, June 13, 1957), some notes (II) were brought to my attention which overlap in subject matter My analysis (I) uses the error probability of a character being in error, while Norris (II) uses the number of block errors per hour Te determine the relation between these two analyses, I have had Mr P. R. Daher trace a sample set of curves. Starting with a given p_c in (I), calculating the block errors/hour for a particular case gives a series of points on the curves of (II) shown in Fig. 8.19a The efficiency curves for p_c - 0.0001 is curve A in Fig. 8 19b

To reduce this curve "A" to the same base as used in (II) the value of p_c is multiplied by 2 25 to account for the backing and overpunching time in punched tape operation analyzed in (II). This modified curve is plotted as "B" in Fig 8 19b.

Curve (I) from Fig. 8 19a is replotted in Fig. 8.19b as "C" Comparison of curves "B" and "C" indicates that this slide rule and graphical comparison shows agreement within two per cent for efficiency vs block length curves derived from the two analyses

The analysis of (I) covers a range of $3 \leq (\delta / \alpha) \leq 192$, while the sample picked out of (II) is for $\delta / \alpha = 97.5$, which is close enough to $\delta / \alpha = 96$ (in I) for comparison The differences in terminology and definitions in the two analysis relate to the orientation of (I) toward high speed data transmission and the starting point of (II) being the slower speed punched tape

F. B Wood
3/28/58

I [illegible]

II [illegible]

I

II

[illegible]

I

II

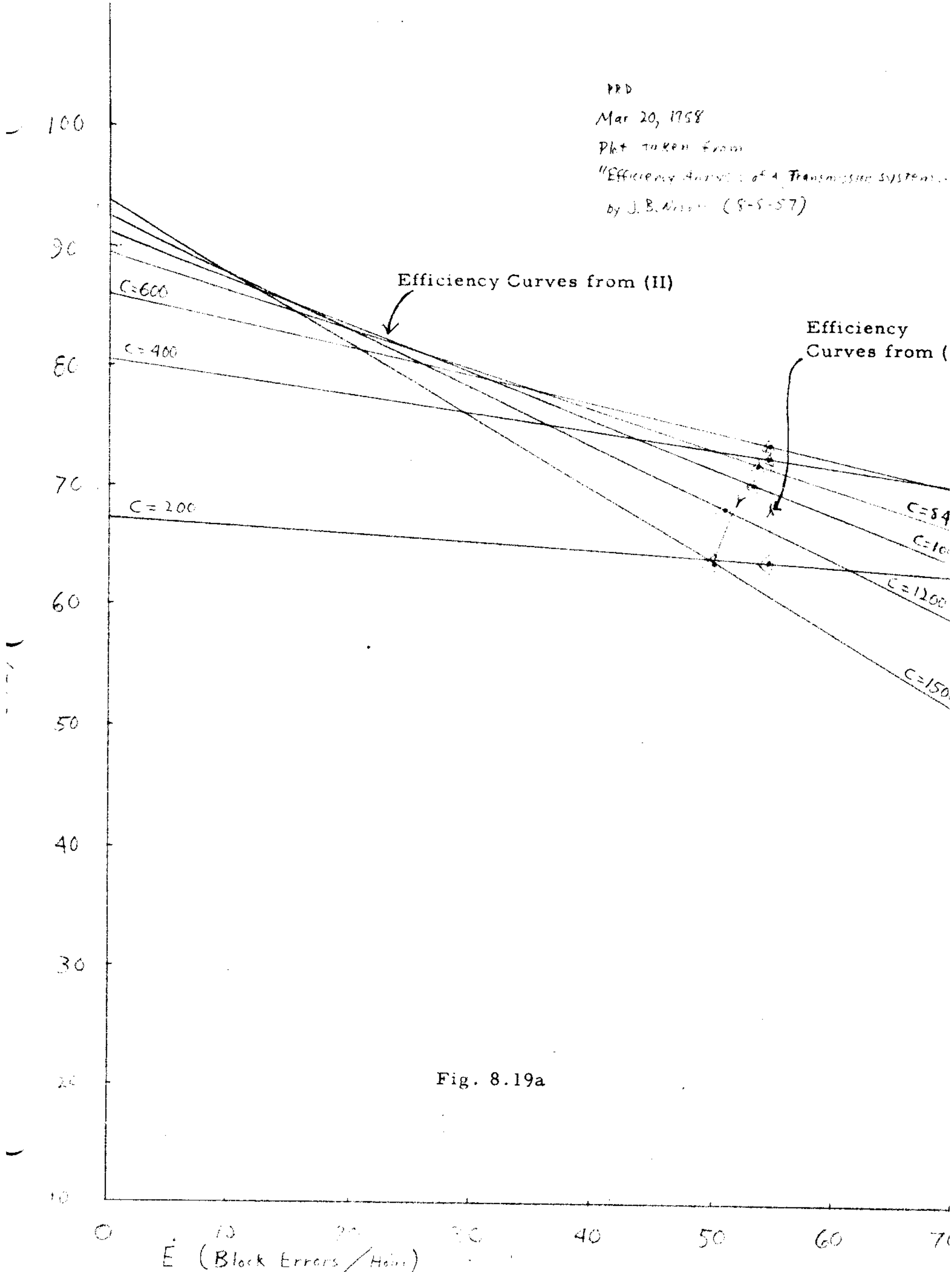

PRD
Mar 20, 1958
Plot taken from
"Efficiency ... of a Transmission System
by J. B. ... (8-5-57)
Efficiency Curves from (II)
Efficiency
Curves from (
C=600
C=400
C=200
C=1200
C=150
100
90
80
70
60
50
40
30
0
40
50
60
Ė (Block Errors / Hour)

Fig. 8.19a