ESD-TR-66-110

MTR-189

OPTIMUM MESSAGE LENGTH FOR A DATA COMMUNICATIONS SYSTEM USING RETRANSMISSION ERROR CONTROL

JUNE 1966

L. L. Stine

Prepared for

DIRECTORATE OF AEROSPACE INSTRUMENTATION
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Distribution of this document is unlimited.

Project 705B
Prepared by
THE MITRE CORPORATION
Bedford, Massachusetts
Contract AF19(628)-5165

ESRIS

AD0635462

ESD-TR-66-110 MTR-189

OPTIMUM MESSAGE LENGTH FOR A DATA COMMUNICATIONS SYSTEM USING RETRANSMISSION ERROR CONTROL

JUNE 1966

L. L. Stine

Prepared for

DIRECTORATE OF AEROSPACE INSTRUMENTATION
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Distribution of this document is unlimited.

Project 705B
Prepared by
THE MITRE CORPORATION
Bedford, Massachusetts
Contract AF19(628)-5165

ABSTRACT

In this report, a mathematical model of a data communications system using retransmission error control is developed. System performance is specified by information throughput, the amount of error-free data transferred per unit of time. Information throughput is determined as a function of message length, taking into account transmission delay, channel error rate, and overhead words in the data message.

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.

C. V. Horrigan

C.V.HORRIGAN
Acting Director
Aerospace Instrumentation

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

SECTION I

INTRODUCTION

In digital data communications, error detection and retransmission is one method of obtaining essentially error-free transfer of data over a bursty channel between a data source and a remotely located data sink. A performance measure of a data communications system is information throughput, the amount of error-free data transferred per unit of time. In this report, a previously developed mathematical model[1] is extended to permit determination of information throughput as a function of message length, taking into account transmission delay, channel error rate, and overhead words in the data message. Kuhn[2], who previously examined this problem, did not include in his analysis the effects of transmission delay or overhead words.

The data communications system, the mathematical model of the system for which the throughput is calculated, the error characteristics of the channel, and the graphical solutions to the throughput equation, which permits determination of optimum message length for maximum throughput under changing operating parameters of channel error rate, delay, and message overhead, are described in this report.

SECTION II

THE DATA COMMUNICATIONS SYSTEM

A functional block diagram of a typical data communications system is shown in Figure 1. A data source sends a sequence of digital information via a noisy channel to a remotely located sink. The data are encoded into information blocks called words. The data words are assembled into message blocks containing N words. A data word is considered in error if one or more errors appear in the received data word. A message is considered in error if one or more of the N words in the message are received in error.

The sink is assumed to have the capability always to detect errors in the received data. This is a reasonable assumption since very powerful error-detecting codes are available. If a message in error is detected, the sink requests the source to go back in the data sequence to the message received in error and recommence transmission. The sink then ignores all incoming messages until the retransmitted data message is received correctly.

The time interval between the detection of a message in error and reception of the retransmission is the sum of the detection time and the round-trip transmission time of the communications system. The error-detection time is the time it takes for the sink to receive the message or N word intervals. The round-trip transmission time of the communications system is D word intervals.

Of the N words in each message, I words are overhead, i.e., they carry no information to the data sink. These I words are used to identify

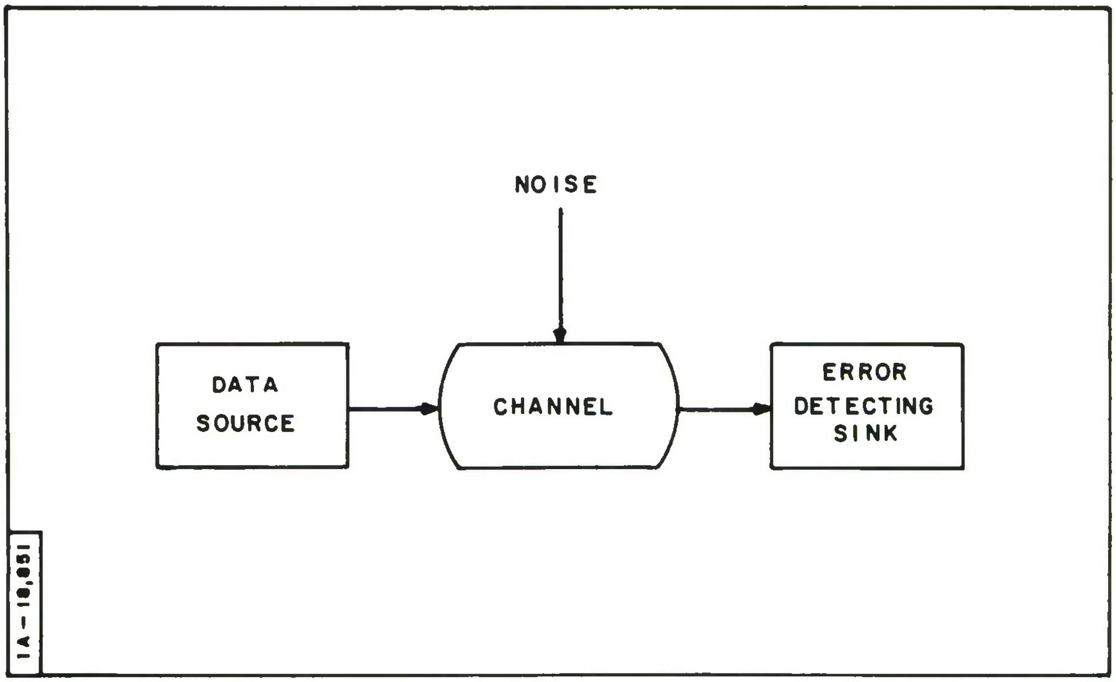

Figure 1. Block Diagram of the Data Communications System

the message, provide error detection, and maintain system synchronization. The remaining N-I words of the message contain the useful information transferred over the communications system.

SECTION III

ERROR CHARACTERISTICS OF THE CHANNEL

Error measurements on various communication media, such as high-frequency radio and troposcatter, indicate that these communications channels are bursty, i.e., the error occurrences are highly correlated. Attempts have been made to model these channels and to list their relevant statistical characteristics[3,4] in order to describe the behavior of the channel error patterns.

A quantity commonly used to describe the channel is average word-error rate, P. This number is found by averaging the number of words in error in a fixed interval of the data stream over the total number of words in the interval. For example, if 10 words were in error in an interval of 10^4 words, the average word-error rate would be 10 divided by 10^4 or 10^{-3}.

For a channel with random, uncorrelated error patterns, average word-error rate is a good measure of transmission quality. For bursty channels, however, average word-error rate is not a meaningful measure unless additional higher order probabilities describing the random channel behavior are given.

In this report, however, average word-error rate alone is used to illustrate the method. Higher order probability terms would vary the calculated results, but not the method employed.

SECTION IV

THE MATHEMATICAL MODEL

GENERAL

The data transmission system attempts to send messages over the channel. Errors occur in the messages at an average rate, P, and cause retransmissions which interrupt the flow of information.

Expressions for the average length of a sequence of error-free data messages, the average number of retransmissions necessary to receive an error-free message, and the information throughput of the communications system are derived below.

AVERAGE LENGTH OF AN ERROR-FREE SEQUENCE

Let $\bar{P}$ equal 1-P, where P is the average word-error rate. From the theory of combinatorial analysis, it can be shown that the probability of a sequence of n words being transmitted without any errors, L(n), is given by

$$L(n) = P \, \bar{P}^n \ . \tag{1}$$

The average sequence length, $\bar{n}$, is given by

$$n = \sum_{n=0}^{\infty} n \, L(n) \ . \tag{2}$$

Substituting (1) into (2) gives

$$\bar{n} = \sum_{n=0}^{\infty} n \, P \, \bar{P}^n \ . \tag{3}$$

Equation (3) reduces to

$$\bar{n} = \frac{\bar{P}\ P}{(1-\bar{P})^2} = \frac{\bar{P}}{P} \quad . \tag{4}$$

Therefore, the average length of the errorless sequence is $\bar{P}/P$ words.

AVERAGE NUMBER OF RETRANSMISSIONS WITH ERRORS

From the mathematics of combinatorial analysis, it can be shown that the probability of x retransmissions to receive an error-free retransmission $K(x)$ is given by

$$K(x) = \bar{P}\ P^{x} \ . \tag{5}$$

The average number of retransmissions to receive an error-free retransmission is given by

$$\bar{x} = \sum_{x=0}^{\infty} x\, K(x) \ . \tag{6}$$

Substituting (5) into (6) yields

$$\bar{x} = \bar{P} \sum_{x=0}^{\infty} xP^{x} \ . \tag{7}$$

Equation (7) is similar to Equation (3) and reduces to

$$\bar{x} = \frac{P}{\bar{P}} \quad . \tag{8}$$

From Equation (4), therefore,

$$\bar{x} = \frac{1}{\bar{n}} \ . \tag{9}$$

TRANSMISSION OF N WORD MESSAGES

When data are transmitted in groups of N words, all N words must be received without errors or a retransmission is requested. The probability of successful transmission $\bar{P}(N)$ is given by

$$\bar{P}(N) = \bar{P}^N \ . \tag{10}$$

The probability that the N word block contains at least one error, $P(N)$, is given by

$$P(N) = 1-\bar{P}^N \ . \tag{11}$$

This relationship is plotted in Figure 2, which shows $P(N)$ as a function of N and $\bar{P}$ or $1-P$.

The average number of blocks transmitted without errors is represented by $\bar{n}_N$, and, in order to receive a message which is retransmitted without errors, the average number of retransmissions of N word blocks is represented by $\bar{x}_N$. The transmission with errors which initiated the retransmission sequence is not included in $\bar{x}_N$. Substituting Equations (9) and (10) into Equations (4) and (8) yields

$$\bar{n}_N = \frac{\bar{P}^N}{1-\bar{P}^N} \ , \tag{12}$$

and

$$\bar{x}_N = \frac{1-\bar{P}^N}{\bar{P}^N} \ . \tag{13}$$

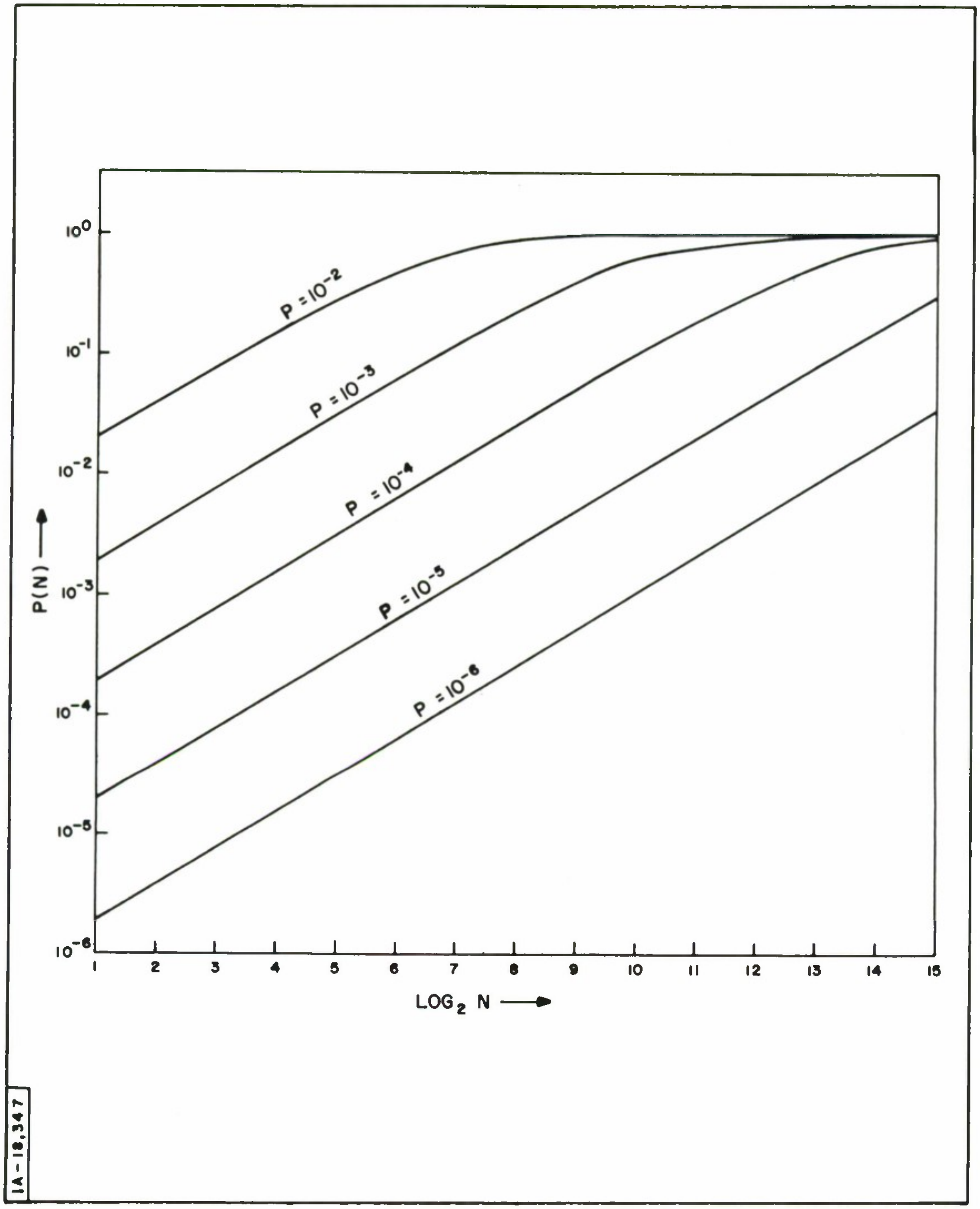

Figure 2. Probability Error of N Word Messages P(N) Versus N and Word-Error Probability P

INFORMATION THROUGHPUT

Transmission of an error-free message results in N data words being accepted by the sink. An error-free sequence of $\bar{n}$ messages results in $N\bar{n}$ data words being accepted by the sink. A retransmission caused by the occurrence of an error and the termination of a sequence of errorless messages results in a transmission interruption of $(D+N)$ words, which is the sum of the round-trip transmission time and the error-detection time. The average interruption duration is $(D+N)$ times the average number of times $(1+\bar{x})$ that a message must be transmitted and retransmitted in order to be received without errors.

Information throughput, T_N, is defined as the amount of error-free information transferred per unit time. This quantity is the ratio of the average information transferred between retransmissions to the information which would be passed if there were no retransmissions. Thus,

$$T_N = \frac{N\bar{n}}{N\bar{n} + (D+N)\ (1+\bar{x})} \quad . \tag{14}$$

Substituting Equations (11) and (12) into Equation (13) yields

$$T_N = \frac{N\ \bar{P}^N/1 - \bar{P}^N}{N\left(\bar{P}^N/1-\bar{P}^N\right) + \left(D+N\right)\left[1+\left(1-\bar{P}^N/\bar{P}^N\right)\right]} \quad . \tag{15}$$

Equation (14) may be rewritten as

$$T_N = \frac{N\ \bar{P}^{2N}}{N\left(\bar{P}^{2N}\right) + \left(D+N\right)\left(1-\bar{P}^N\right)} \quad . \tag{16}$$

Since each N word message contains I overhead words, which results in a relative fraction, R, of noninformation-carrying words per message where

$$R = \frac{I}{N}, \tag{17}$$

then the actual information throughput, T, is given by

$$T = T_N \ (1-R) \ . \tag{18}$$

Substituting Equations (15) and (16) into Equation (17) yields

$$T = \frac{N \ \bar{P}^{2N} \ \left(1 - {}^{I}\!/\!{}_{N}\right)}{N \ \bar{P}^{2N} + \left(D + N\right) \ \left(1 - \bar{P}^{N}\right)} \ . \tag{19}$$

Equation (18) has been calculated by a computer using the values of $\bar{P}$, N, D, and I shown in Table I. Figures 3 through 18 are plots of Equation (18). Each plot shows information throughput, T, versus message length, N, for various values of $\bar{P}$ and constant values of D and I.

By inspecting Figures 3 through 18 it is obvious that for given values of P, D, and I there is a unique message length that maximizes the information throughput. Message lengths greater or less than the optimum will result in decreased throughput. Tables II through VI show the optimum message length to yield maximum throughput versus $\bar{P}$ for various values of D and I.

Table I

Values of $\overline{P}$, N, D, and I
Used in Computer Solution of Equation (19)

$\overline{P}$	N	D	I
$1 - 10^{-2}$	2^1	10	1
$1 - 10^{-3}$	2^2	40	2
$1 - 10^{-4}$	2^3	160	4
$1 - 10^{-5}$	2^4	640	8
$1 - 10^{-6}$	2^5		
	2^6		
	2^7		
	2^8		
	2^9		
	2^{10}		
	2^{11}		
	2^{12}		
	2^{13}		
	2^{14}		
	2^{15}		

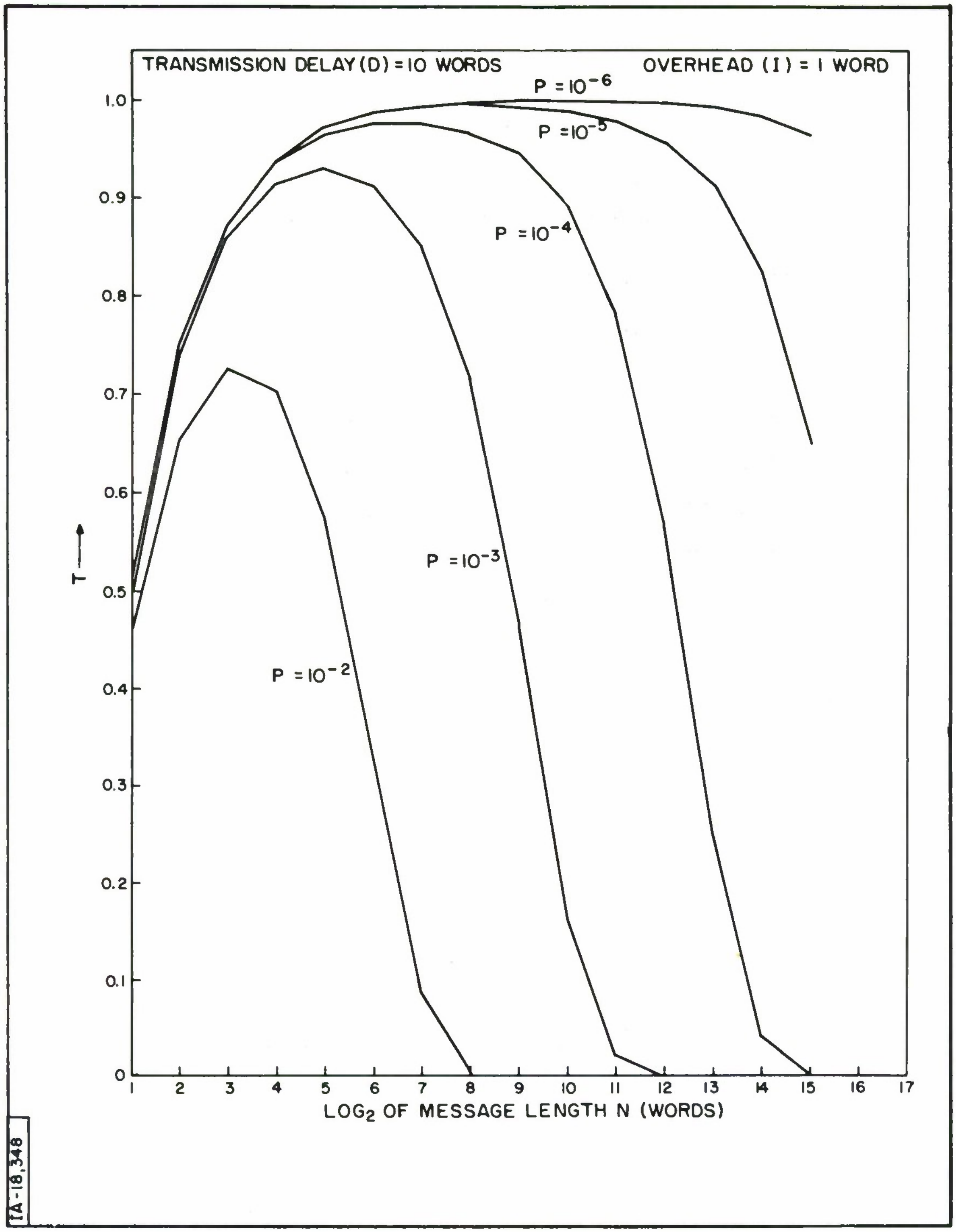

Figure 3. Information Throughput T Versus Message Length N and Probability of Error P

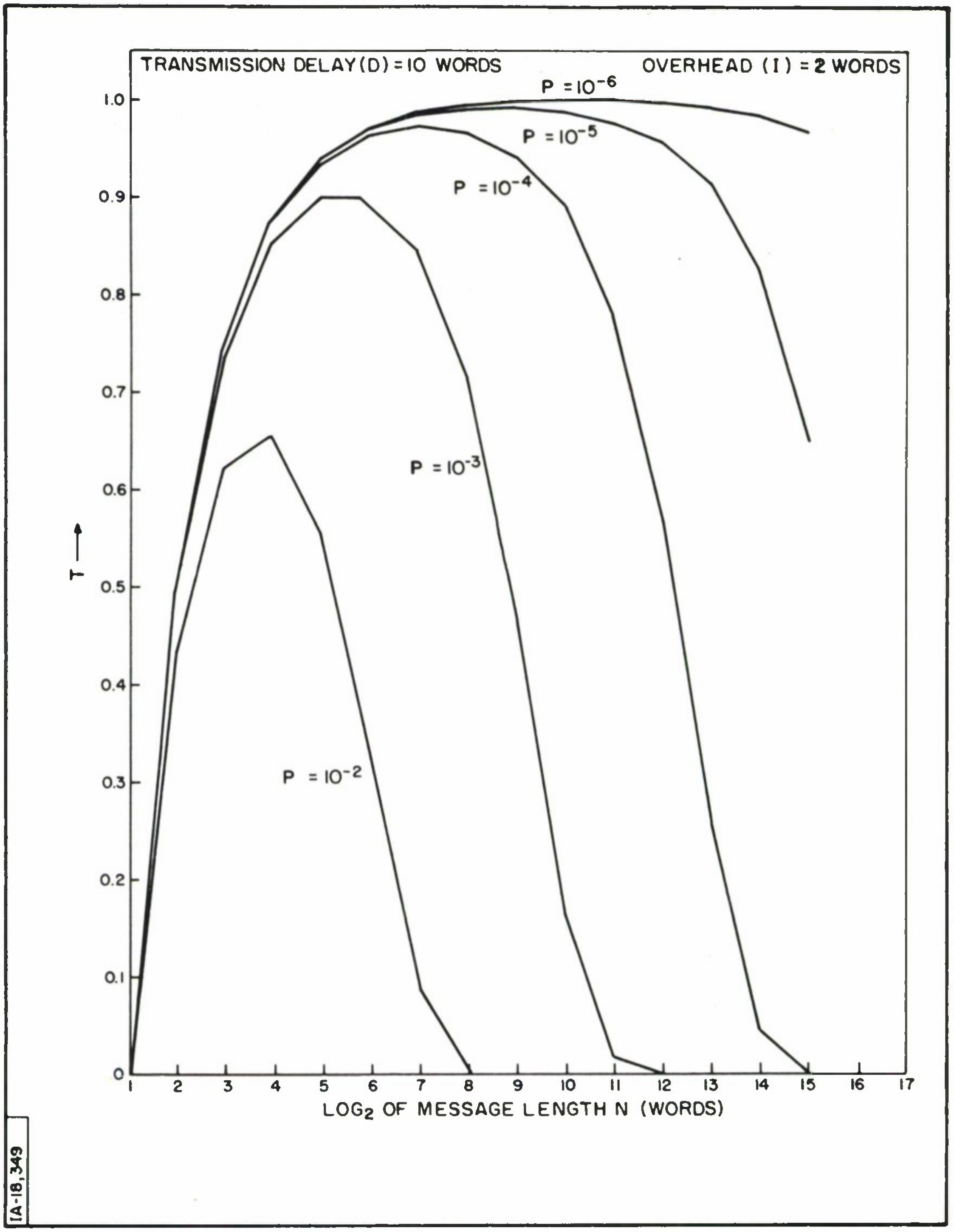

Figure 4. Information Throughput T Versus Message Length N and Probability of Error P

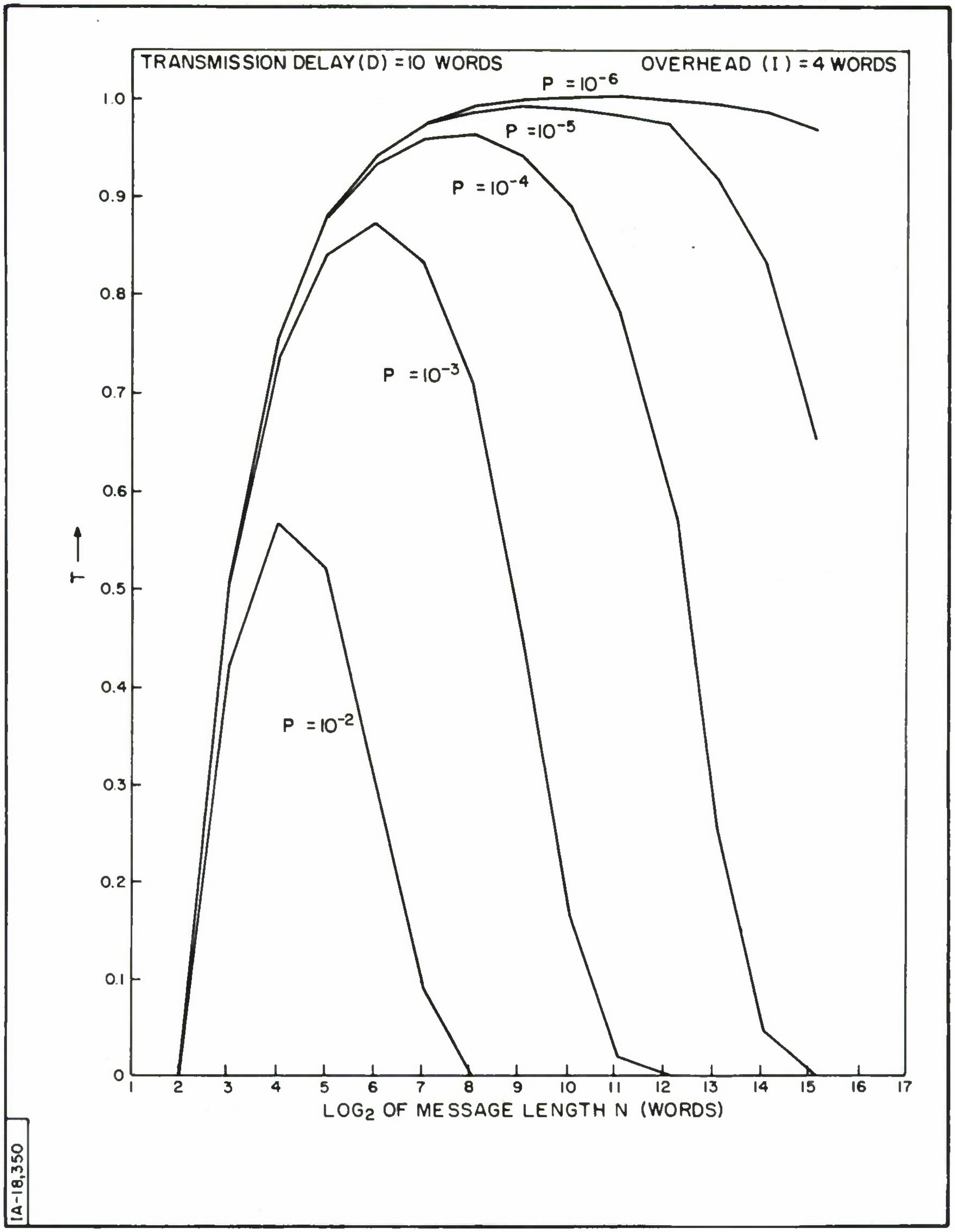

Figure 5. Information Throughput T Versus Message Length N and Probability of Error P

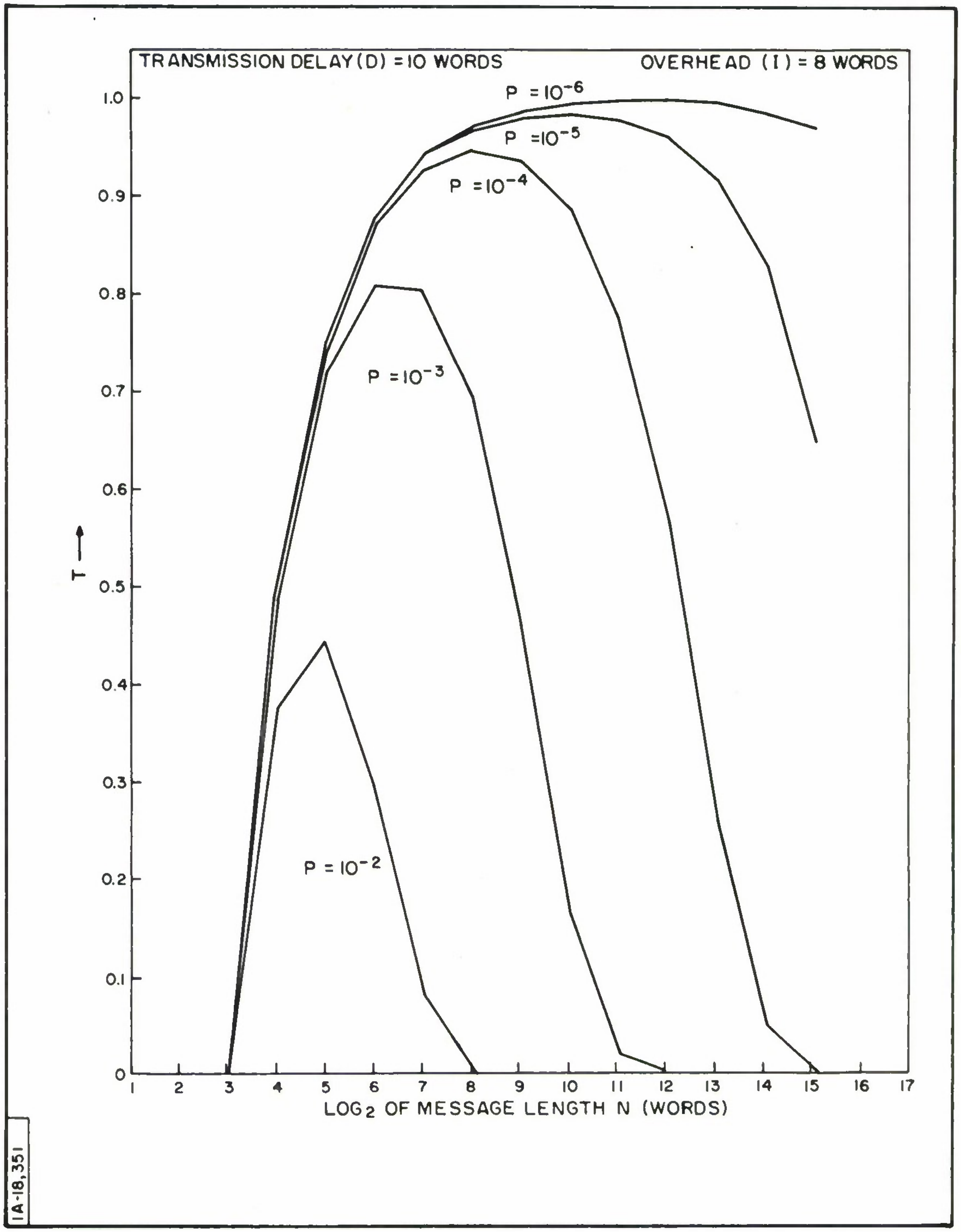

Figure 6. Information Throughput T Versus Message Length N and Probability of Error P

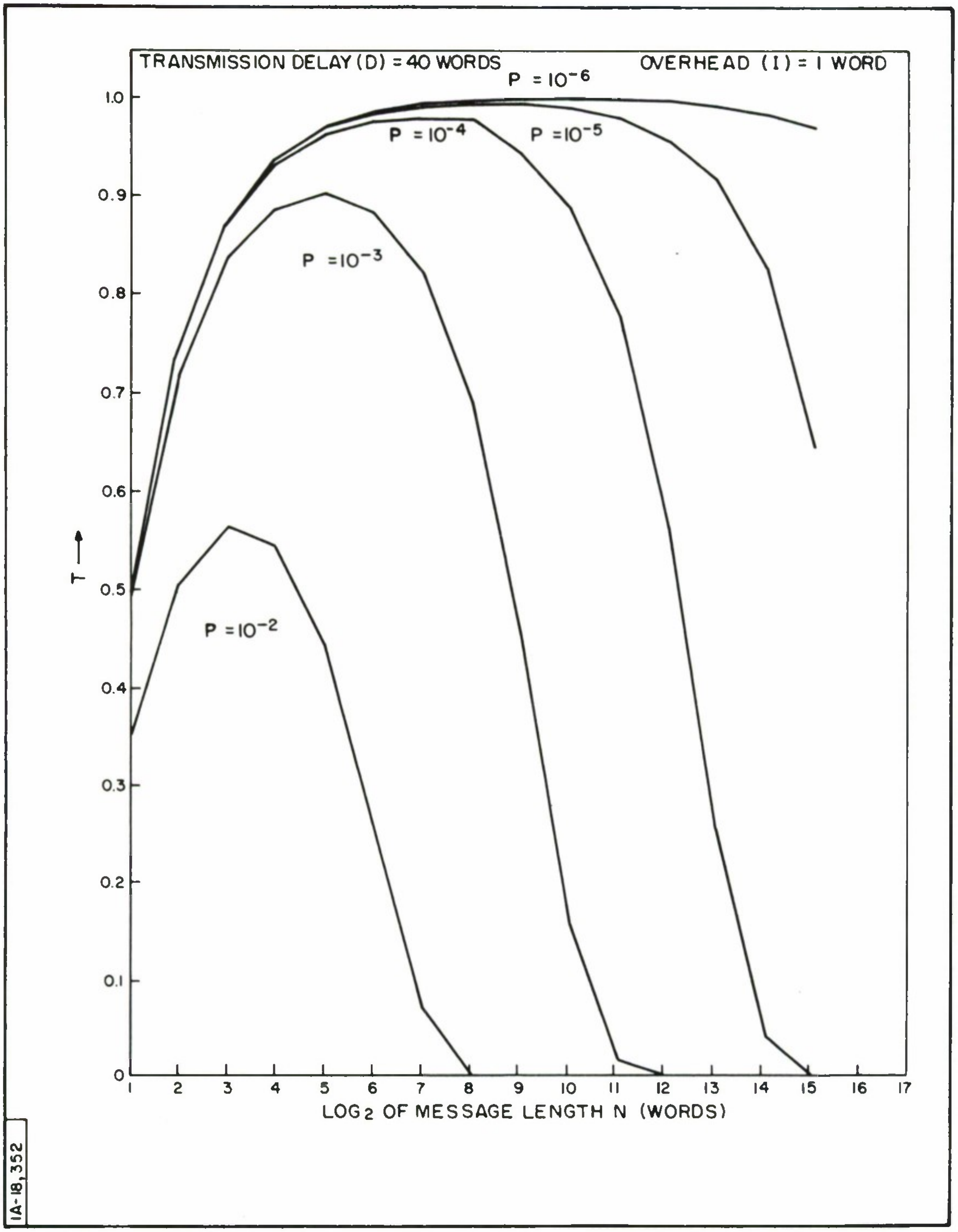

Figure 7. Information Throughput T Versus Message Length N and Probability of Error P

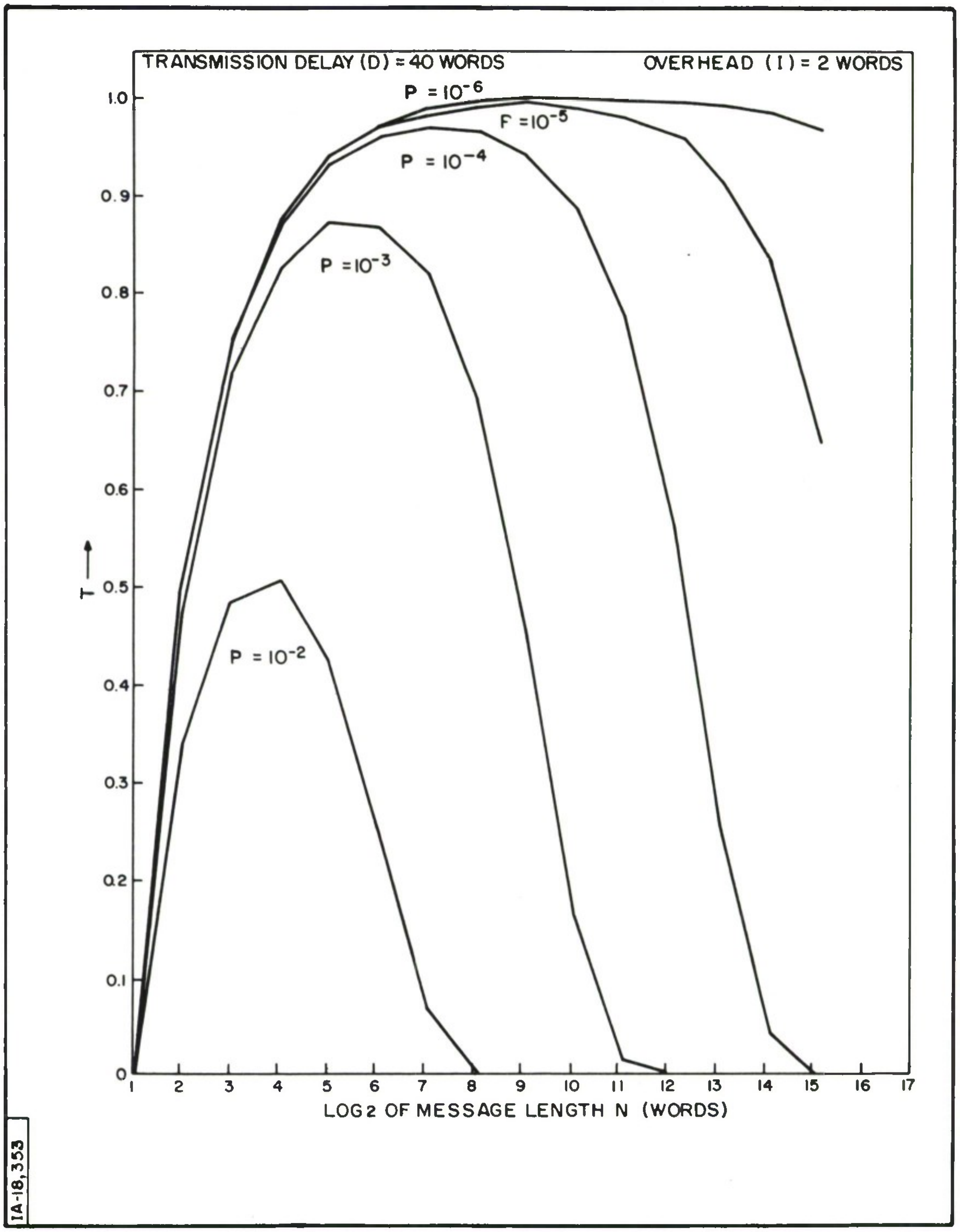

Figure 8. Information Throughput T Versus Message Length N and Probability of Error P

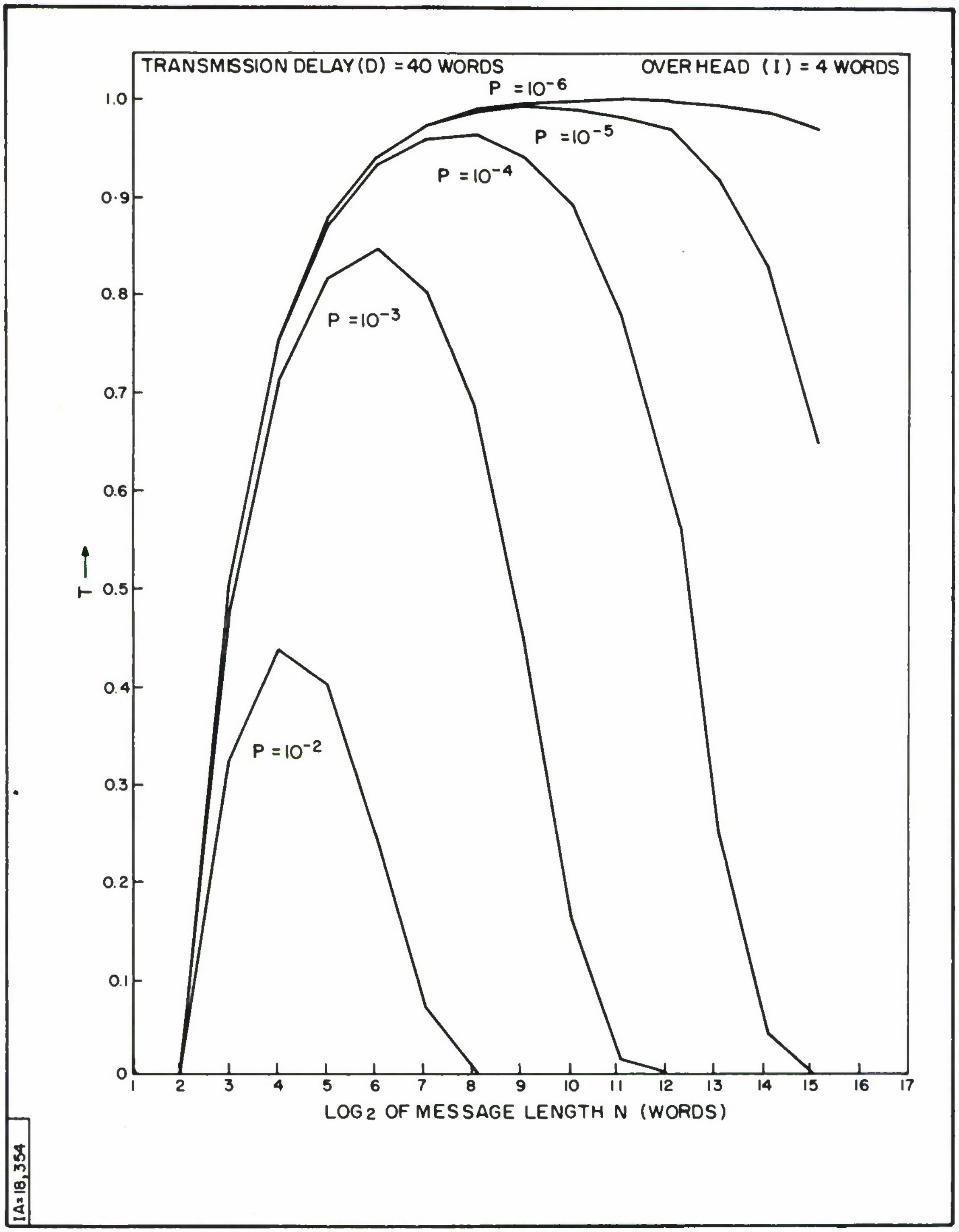

Figure 9. Information Throughput T Versus Message Length N and Probability of Error P

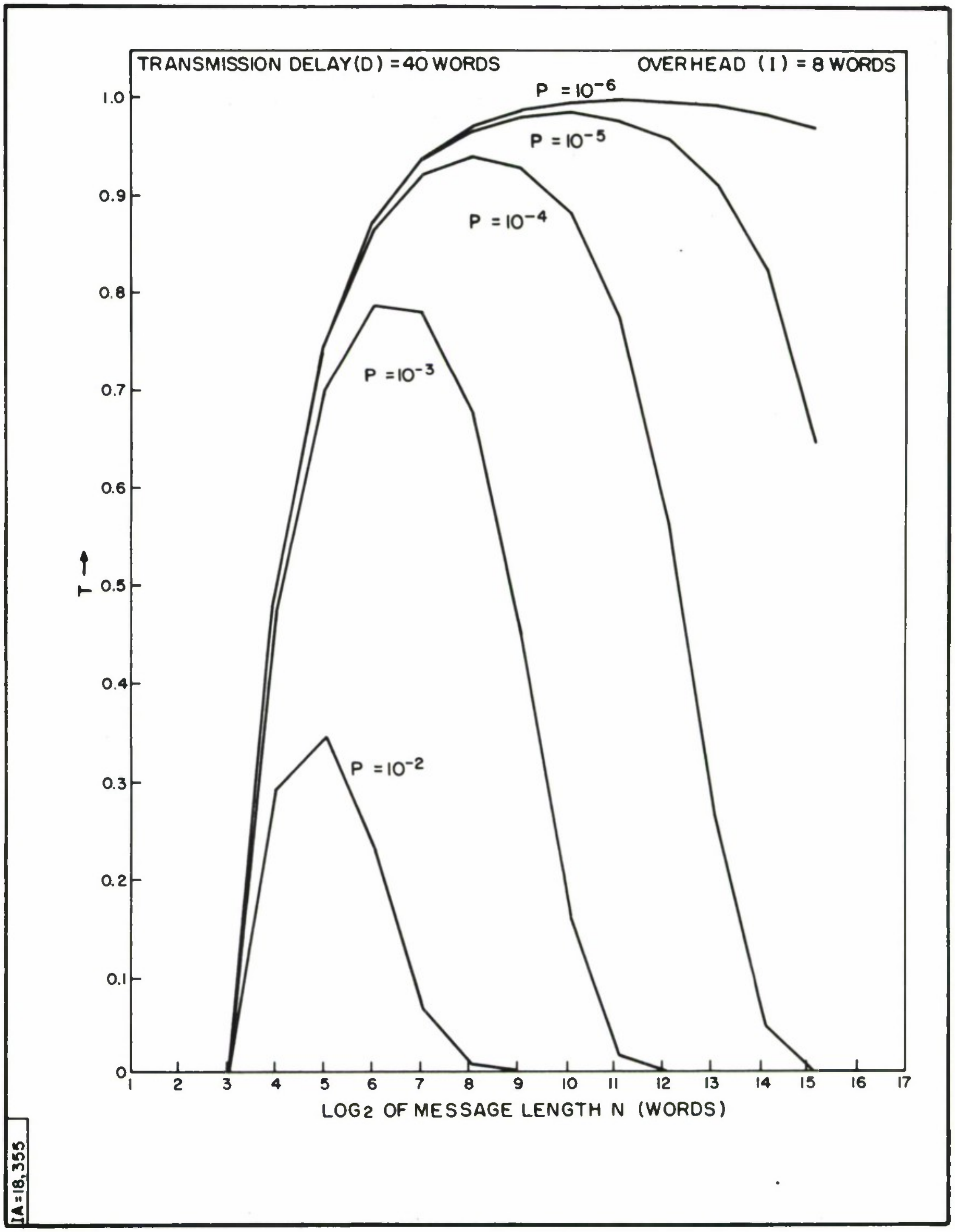

Figure 10. Information Throughput T Versus Message Length N and Probability of Error P

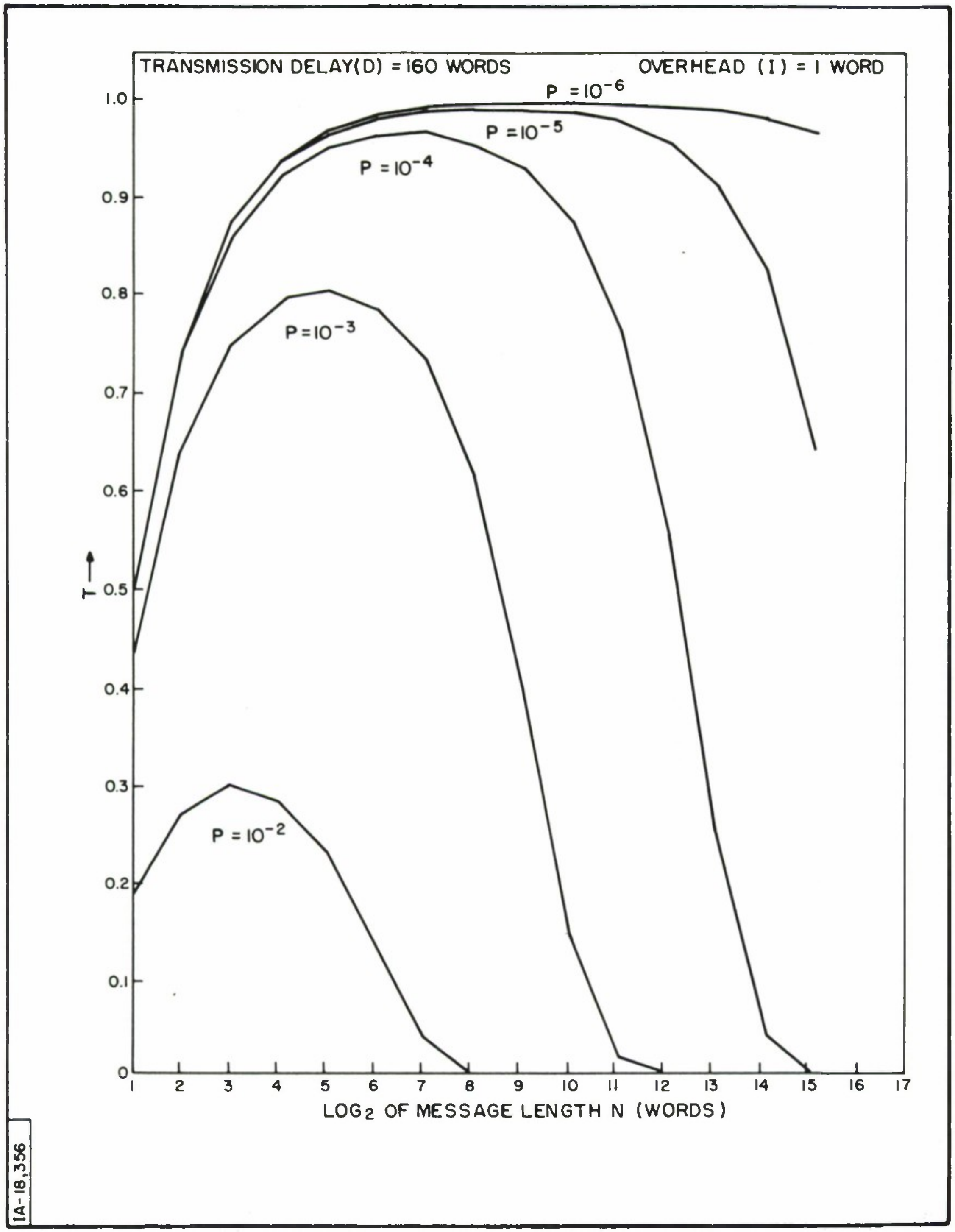

Figure 11. Information Throughput T Versus Message Length N and Probability of Error P

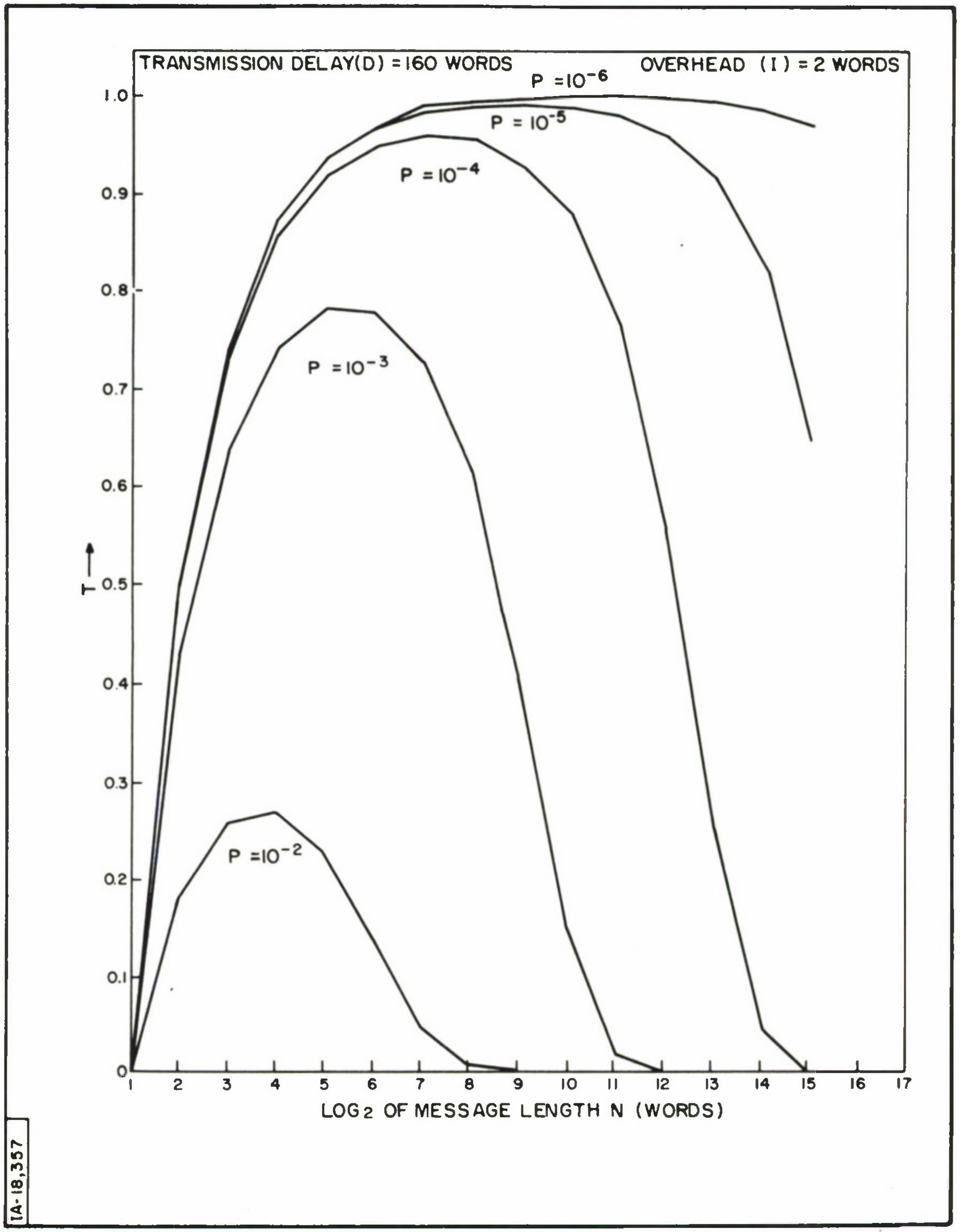

Figure 12. Information Throughput T Versus Message Length N and Probability of Error P

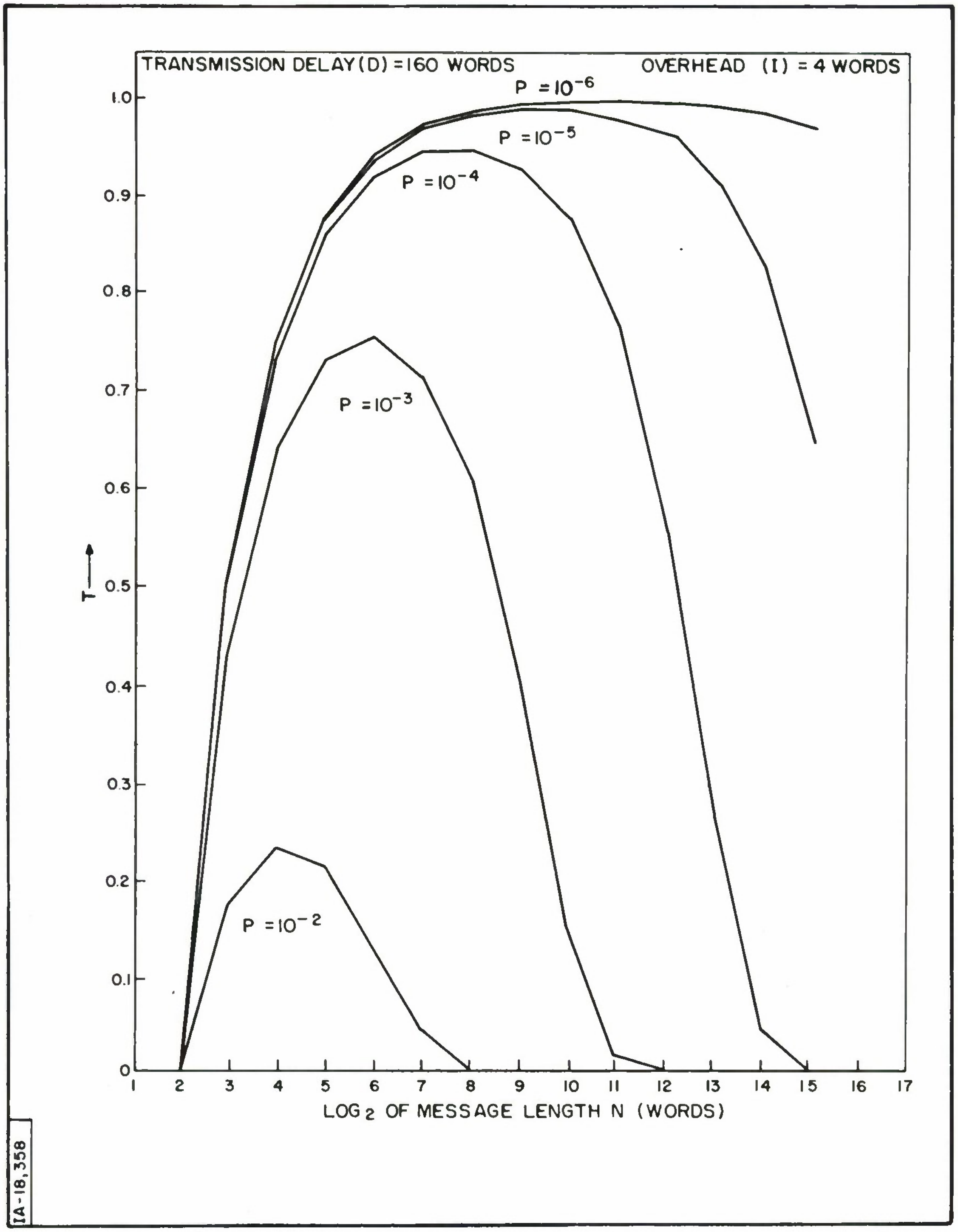

Figure 13. Information Throughput T Versus Message Length N and Probability of Error P

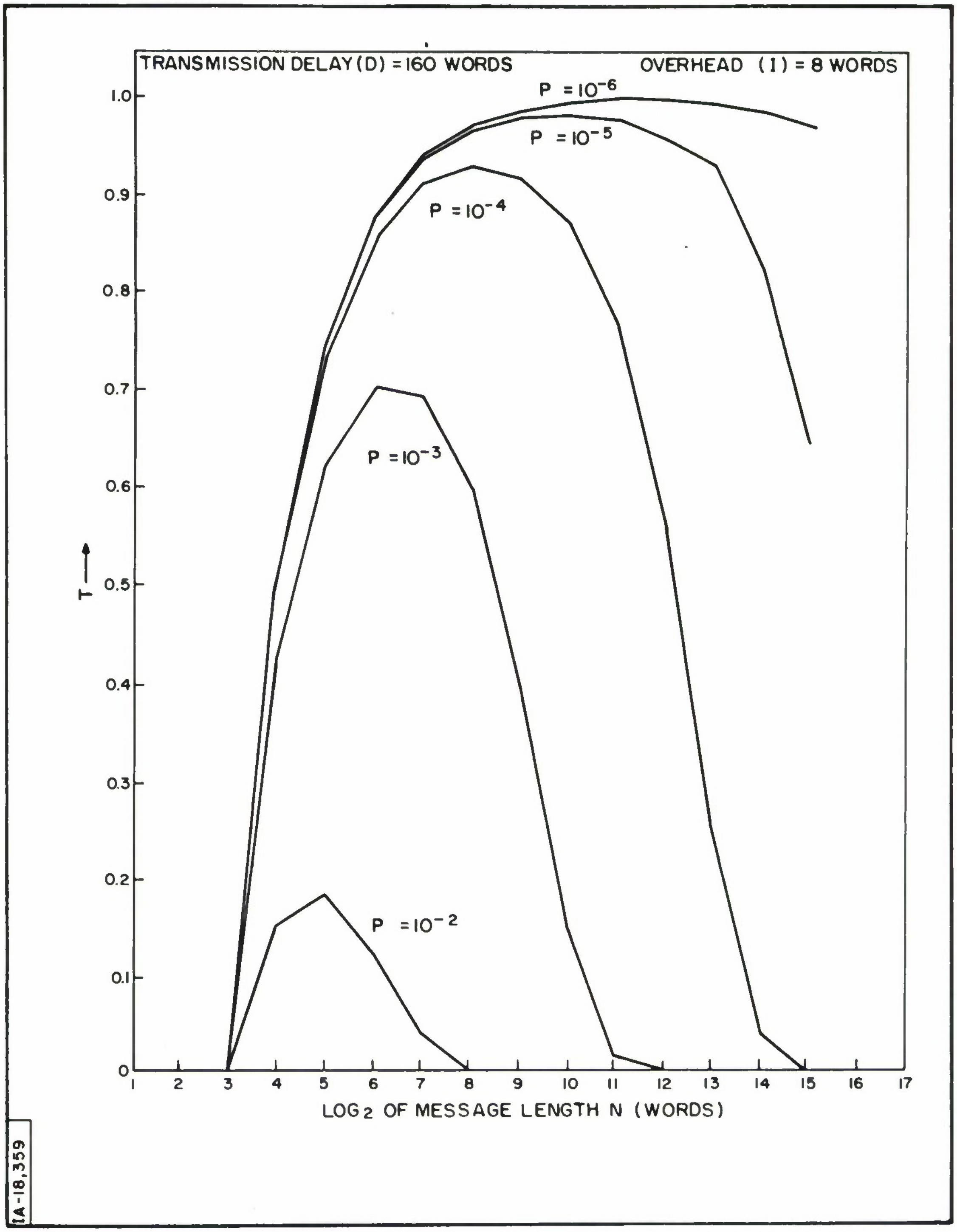

Figure 14. Information Throughput T Versus Message Length N and Probability of Error P

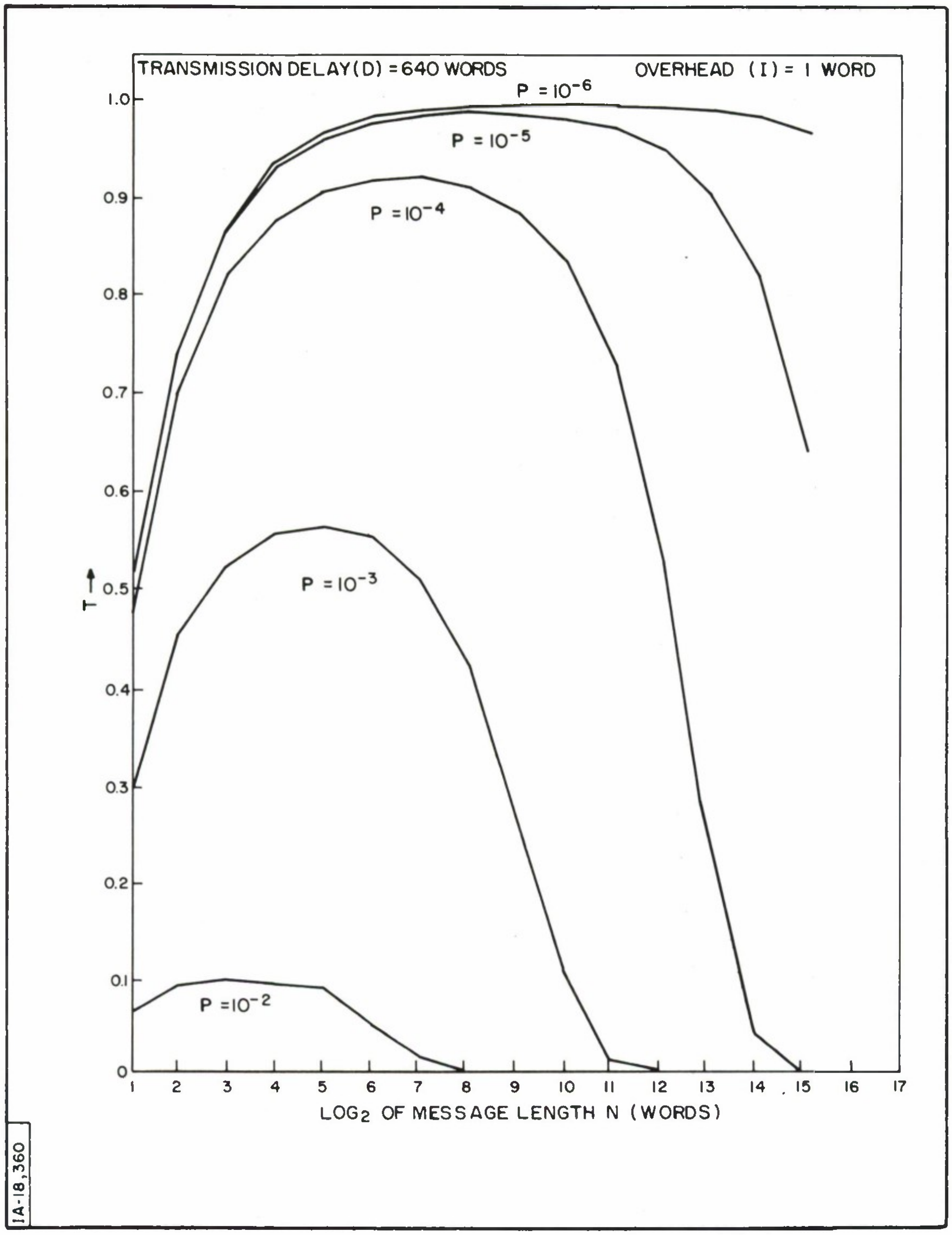

Figure 15. Information Throughput T Versus Message Length N and Probability of Error P

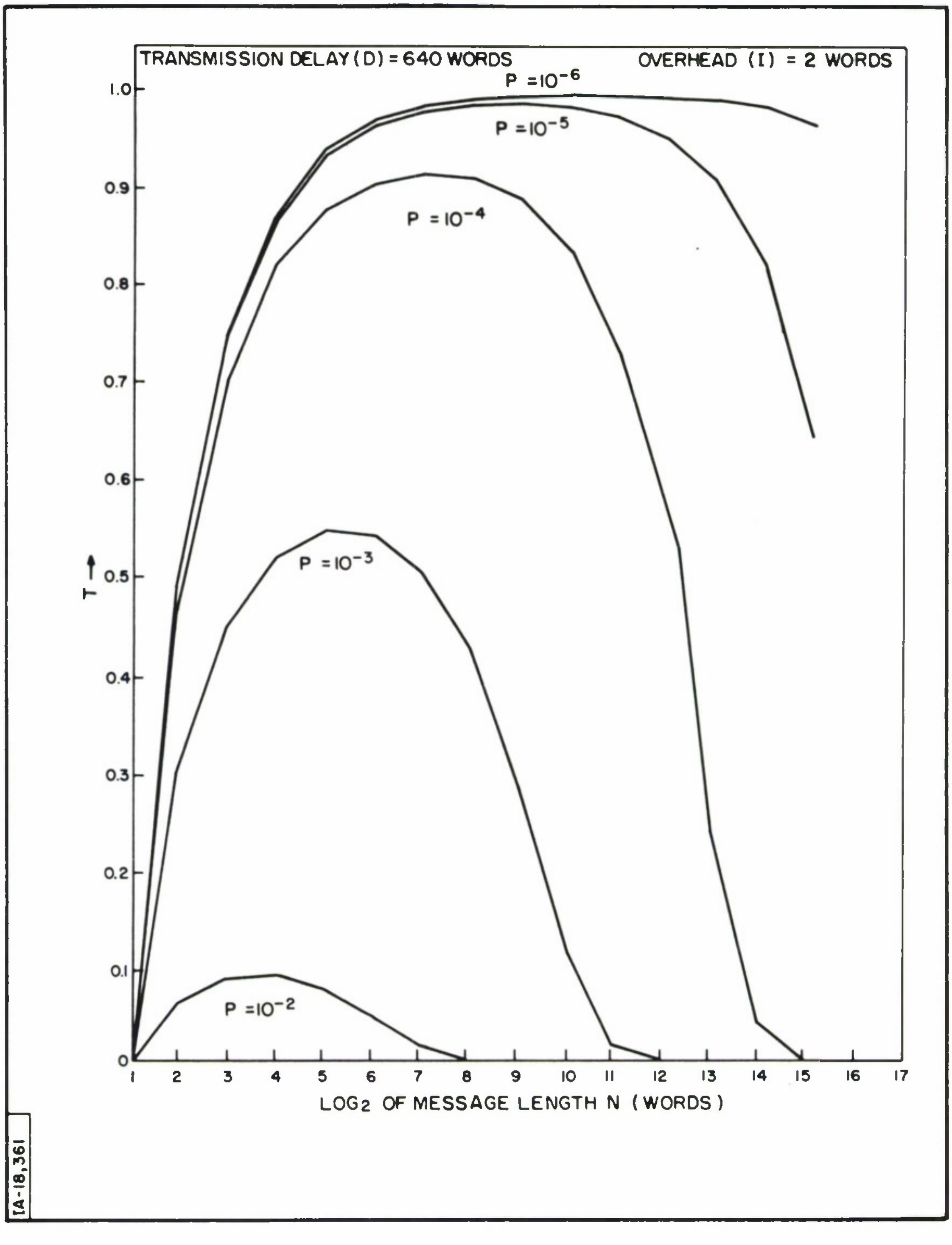

Figure 16. Information Throughput T Versus Message Length N and Probability of Error P

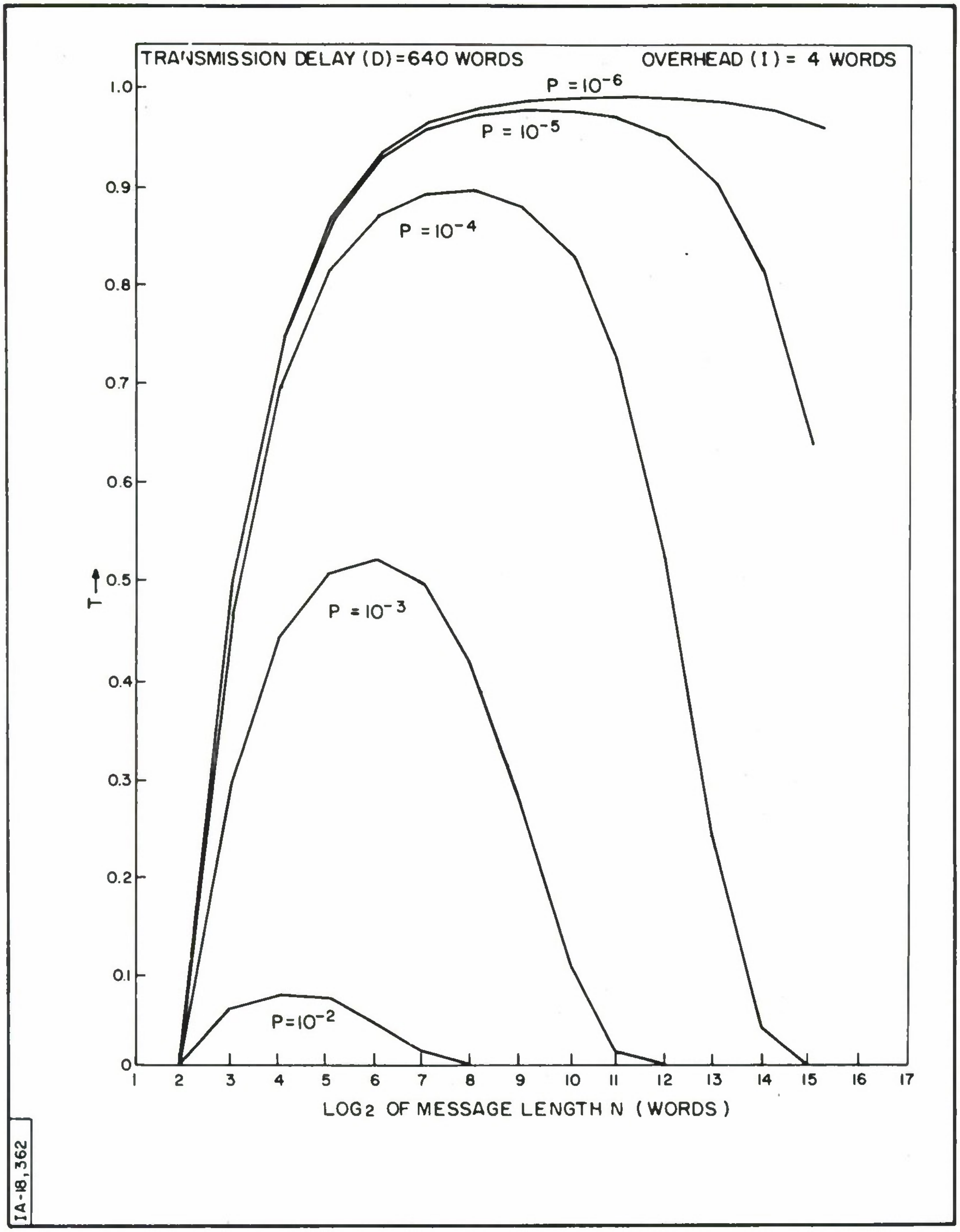

Figure 17. Information Throughput T Versus Message Length N and Probability of Error P

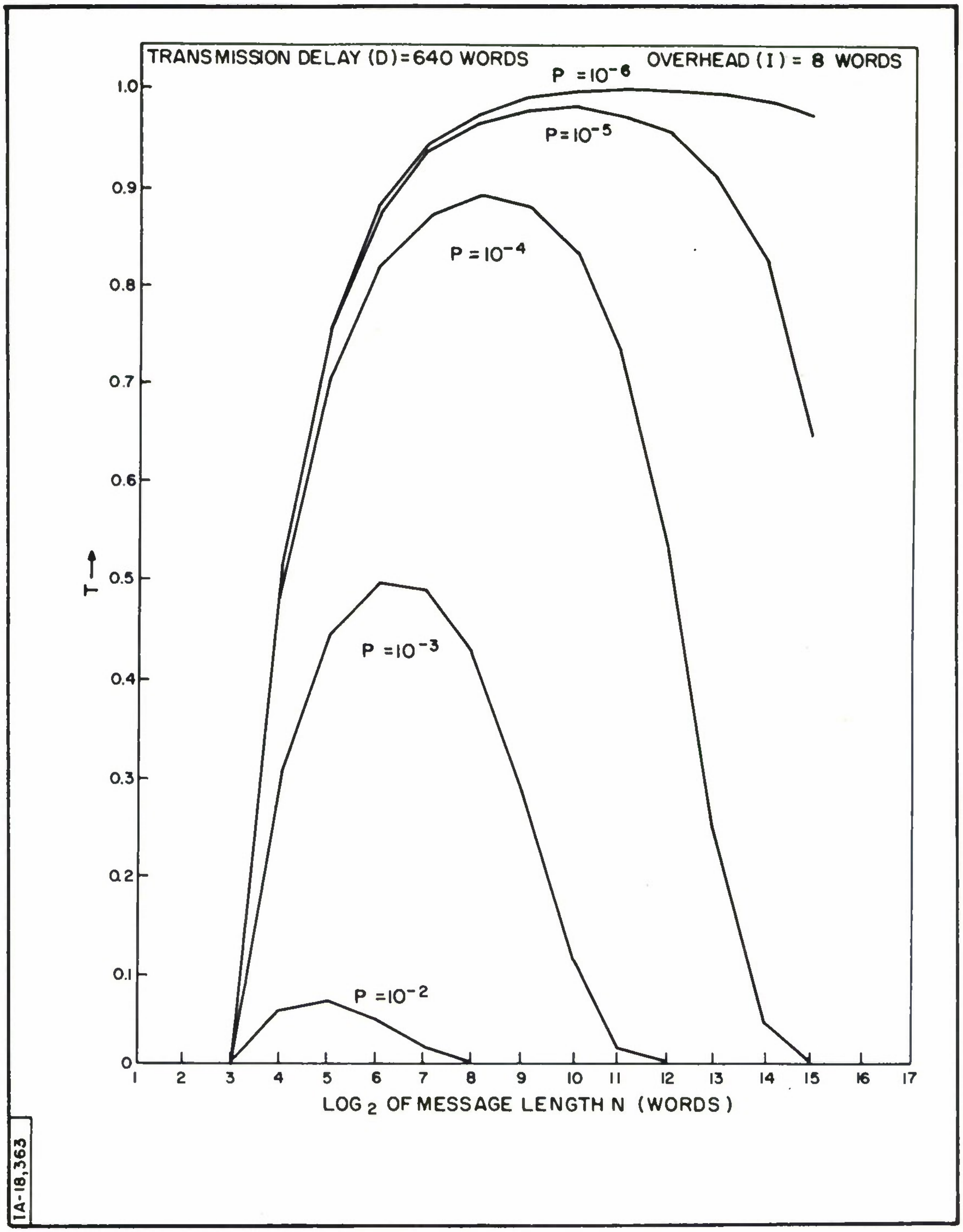

Figure 18. Information Throughput T Versus Message Length N and Probability of Error P

Table II

Optimum Message Length in Words for $\overline{P}$ Equal to $1\text{-}10^{-2}$ Versus D and I

I D	1	2	4	8
10	8	16	16	32
40	8	16	16	32
160	8	16	16	32
640	8	16	16	32

Table III

Optimum Message Length in Words for $\overline{P}$ Equal to $1\text{-}10^{-3}$ Versus D and I

I D	1	2	4	8
10	32	32	64	64
40	32	32	64	64
160	32	32	64	64
640	32	32	64	64

Table IV

Optimum Message Length in Words for $\bar{P}$ Equal to $1\text{-}10^{-4}$ Versus D and I

D \ I	1	2	4	8
10	128	128	256	256
40	128	128	256	256
160	128	128	256	256
640	128	128	256	256

Table V

Optimum Message Length in Words for $\bar{P}$ Equal to $1\text{-}10^{-5}$ Versus D and I

D \ I	1	2	4	8
10	256	512	512	1024
40	256	512	512	1024
160	256	512	512	1024
640	256	512	512	1024

Table VI

Optimum Message Length in Words for
$\bar{P}$ Equal to $1-10^{-6}$ Versus D and I

D \ I	1	2	4	8
10	1024	1024	2048	2048
40	1024	1024	2048	2048
160	1024	1024	2048	2048
640	1024	1024	2048	2048

SECTION V

CONCLUSIONS

In this report, a mathematical model of a data communications system using retransmission error control was developed. System performance is specified by information throughput, the amount of error-free data transferred per unit of time. Information throughput, T, is determined as a function of message length, N, taking into account transmission delay, D, channel error rate, P, and overhead words in the data message, I. The throughput expression given in Equation (19) was plotted in Figures 3 through 18 for various values of N, D, P, and I.

It is obvious from inspecting the curves that throughput is dependent on error rate and message length. If the error rate is 10^{-5} or less, then message length is very close to optimum over a broad range of values. This optimum is also greater than 0.95. Lower rates than 10^{-5} would bring the optimum closer to 1.0. For example, a nominal error rate of wire line might be 10^{-5}. To attain a lower error rate would require forward acting error-correction codes which would add a fixed percent of redundancy, C, to each data word. Typical values of C range from 0.33 to 0.5. Since the attainable increase in throughput is only 0.05 or less, the use of these codes, however, would actually decrease the throughput.

Typical optimum throughputs at an error rate of 10^{-2}, however, range from 0.75 to 0.05. If error-correction coding can decrease the error rate to 10^{-5} or better, a net gain in information throughput can result.

A study of the advantages of error-correction coding is just one example of the usefulness of the model presented in this report. Other uses will depend on the problems of the communications system designer.

REFERENCES

1. L. L. Stine, Optimal Delay for Forward Error Correction Equipment, the MITRE Corporation, WP-346, Bedford, Mass., 12 Nov. 1965.

2. T. G. Kuhn, Retransmission Error Control, IEEE Trans. on Communications Systems, June (1963).

3. S. Berkovits, et al., A Model for Digital Error Distributions, First IEEE Annual Communications Convention, June 7-9, 1965, Boulder, Colorado.

4. S. Sussman, Communication Channel Attributes as Related to Error Control, First IEEE Annual Communications Convention, June 7-9, 1965, Boulder, Colorado.

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)* The MITRE Corporation Bedford, Massachusetts	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Optimum Message Length for a Data Communications System Using Retransmission Error Control

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

N/A

5. AUTHOR(S) *(Last name, first name, initial)*

Stine, Larry L.

6. REPORT DATE June 1966	7a. TOTAL NO. OF PAGES 36	7b. NO. OF REFS 4

8a. CONTRACT OR GRANT NO. AF19(628)-5165	9a. ORIGINATOR'S REPORT NUMBER(S) ESD-TR-66-110
b. PROJECT NO. c. 705B d.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* MTR-189

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Directorate of Aerospace Instrumentation; Electronic Systems Division, L.G. Hanscom Field, Bedford, Massachusetts

13. ABSTRACT

In this report, a mathematical model of a data communications system using retransmission error control is developed. System performance is specified by information throughput, the amount of error-free data transferred per unit of time. Information throughput is determined as a function of message length, taking into account transmission delay, channel error rate, and overhead words in the data message.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Data Communications						
Error Control						
ARQ						
Throughput						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (*corporate author*) issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (*either by the originator or by the sponsor*), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
______________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
______________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
______________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (*paying for*) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.