NRL Report 8140

Throughput Performance of Data-Communication Systems Using Automatic-Repeat-Request (ARQ) Error-Control Schemes

JOEL M. MORRIS

Systems Integration and Instrumentation Branch
Communications Sciences Division

September 8, 1977

NAVAL RESEARCH LABORATORY
Washington, D.C.

Approved for public release; distribution unlimited.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NRL REPORT 8140	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THROUGHPUT PERFORMANCE OF DATA-COMMUNICATION SYSTEMS USING AUTOMATIC-REPEAT-REQUEST (ARQ) ERROR-CONTROL SCHEMES		5. TYPE OF REPORT & PERIOD COVERED Interim report on a continuing NRL Problem 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Joel M. Morris		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, D.C. 20375		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NRL Problem R01-90 28010N Task X0720
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Electronic Systems Command Washington, D.C. 20360		12. REPORT DATE September 8, 1977 13. NUMBER OF PAGES 37
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Automatic repeat request	Data communications
ARQ	Error detection
Data throughput	Error correction
Throughput efficiency	Error control

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The throughput performance of several ARQ schemes was evaluated. The purpose is to provide data-communication design engineers with information to help them choose error-control techniques or to assess the performance of their proposed ARQ or hybrid designs. The schemes considered included the stop-and-wait, the go-back-N, and the selective-repeat schemes and some variations of these. In addition, the performance of a hybrid scheme that uses the best properties of both the ARQ scheme and the forward-error-correction (FEC) scheme was assessed. Most of

(Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (Continued)

the equations derived for computing throughput efficiency η or optimal blocklength were obtained as direct modifications or extensions or equations found in the published literature. The equations were derived primarily for random-error channel models, although some burst-error conditions were considered. Among items yet to be studied are the optimal transmission rate and blocklength to maximize the throughput (in bits per second).

CONTENTS

INTRODUCTION 1

VARIOUS ARQ SCHEMES 2

Basic Stop-and-Wait ARQ Scheme 2
Variation 1 of the Stop-and-Wait ARQ Scheme 3
Variation 2 of the Stop-and-Wait ARQ Scheme 3
Continuous-ARQ Schemes 3
Go-Back-*N* Continuous-ARQ Scheme 4
Variation of the Go-Back-*N* Continuous-ARQ Scheme 4
Selective-Repeat Continuous-ARQ Scheme 5
Hybrid ARQ Schemes 5

THROUGHPUT PERFORMANCE OF THE VARIOUS ARQ SCHEMES 5

Preliminary Notation, Definitions, and Assumptions 6
Throughput of the Basic Stop-and-Wait ARQ Scheme 7
Throughput of Variation 1 of the Stop-and-Wait ARQ Scheme 8
Throughput of Variation 2 of the Stop-and-Wait ARQ Scheme 12
Throughput of the Go-Back-*N* Continuous-ARQ Scheme 12
Throughput of the Variation of the Go-Back-*N* Continuous-ARQ Scheme 13
Throughput of the Selective-Repeat Continuous-ARQ Scheme 15
Throughput Error of the Hybrid ARQ Scheme With Separate Error Detection and Correction Encoding 15

OPTIMAL BLOCKLENGTHS FOR ARQ SYSTEMS 19

Optimal Blocklength for the Basic Stop-and-Wait ARQ Scheme 20
Optimal Blocklength for Variation 1 of the Stop-and-Wait ARQ Scheme 21
Optimal Blocklength for Variation 2 of the Stop-and-Wait ARQ Scheme 26
Optimal Blocklength for the Go-Back-*N* Continuous-ARQ Scheme 26
Optimal Blocklength for the Variation of the Go-Back-*N* Continuous-ARQ Scheme 27
Optimal Blocklength for the Selective-Repeat Continuous-ARQ Scheme 29

SUMMARY AND CONCLUSIONS 29

REFERENCES 33

THROUGHPUT PERFORMANCE OF DATA-COMMUNICATION SYSTEMS USING AUTOMATIC-REPEAT-REQUEST (ARQ) ERROR-CONTROL SCHEMES

INTRODUCTION

An automatic-repeat-request (ARQ) scheme [1-6] is a highly reliable error-control technique for the transmission of messages or blocks of data in data-transmission systems such as message or packet-switched networks. This high reliability is a result of the continuous retransmission (repeats requested) of those messages or blocks found in error by the receiver. Consequently, as the channel error rate increases, ARQ schemes yield lower and lower throughput (received message or block acceptance rate) while maintaining high reliability.

In contrast, forward-error-correction (FEC) schemes [7], maintain a constant message or block acceptance rate at the receiver but suffer decreased reliability by accepting larger and larger numbers of uncorrected messages or blocks as the error rate increases. The FEC schemes however do not need feedback communications for the implementation of their error-correction techniques, and this absence of feedback channels is an operational requirement in many applications.

Both ARQ and FEC schemes generally use coding to detect, correct, or both detect and correct some finite number of errors in the received message or block (a human operator may perform the operation mentally). The price paid for coding is reduced throughput, since the ratio of information bits to transmitted bits is reduced. In addition, in any practical system, there is a nonzero probability — although usually acceptable with proper design — of some accepted messages or blocks being in error.

Generally the choice of ARQ over FEC schemes depends primarily on the availability of a feedback link and the desirability of constant high reliability rather than constant throughput performance. Yet some combination of ARQ and FEC [6-8] schemes have been shown effective in moderate- and high-error-rate environments. In this situation the error-correcting properties of the code may considerably reduce the error rate experienced by the error-detecting properties of the code if acting alone. The result is that the ARQ component of this scheme operates in an effective low-to-moderate error environment. Thus the scheme yields improved throughput performance overall.

In this report we evaluate the throughput performance of several ARQ schemes. The purpose is to provide data-communication design engineers with information to help them choose error-control techniques or to assess the performance of their proposed ARQ or hybrid designs. The schemes considered include the stop-and-wait, the go-back-N, and the selective-repeat schemes and some variations of these. In addition the performance of

Manuscript submitted May 31, 1977.

ARQ-FEC hybrid schemes, which use the best properties of both, is assessed. Most of the equations presented for computing throughput efficiency η or optimal blocklength are obtained as direct modifications or extensions of equations found in the published literature. The equations are derived primarily for random-error channel models, although some burst-error conditions are considered.

In the next section of this report we discuss in general terms the different ARQ schemes considered. Hybrid schemes are also included. This section provides a heuristic presentation of these schemes for the interested reader who is not too knowledgeable about ARQ techniques. The respective throughput efficiency equations are developed and evaluated in the third section. Throughput curves for some of the ARQ schemes are presented. The fourth section contains the development of optimal-blocklength equations and algorithms for those ARQ schemes that segment messages and data into blocks. Explicit results on optimal blocklengths are presented in this section for some of the considered schemes. The summary, concluding remarks, and some recommendations for additional study are presented in the final section.

VARIOUS ARQ SCHEMES

The various automatic-repeat-request (ARQ) schemes commonly considered are variations of two basic ARQ error-control techniques: the stop-and-wait ARQ scheme and the continuous-ARQ scheme. However hybrid ARQ schemes which incorporate error-correcting techniques are becoming popular in the current literature.

Basic Stop-and-Wait ARQ Scheme

The stop-and-wait ARQ scheme [1-6] is probably the most popular error-control scheme. This scheme is in widespread use because it is relatively easy to implement, it requires only one block of temporary storage, and, since transmission of information is in only one direction at a time, half-duplex data links are appropriate.

In the stop-and-wait ARQ scheme the transmitter delays (*stops*) transmission of a succeeding block and temporarily stores the last block (*waits*) until receiving some form of acknowledgment from the receiver indicating the acceptability of the last transmitted block. If the received block is error-free and therefore *acceptable*, the receiver indicates this by sending an ACK signal to the transmitter. The transmitter then proceeds to transmit the next block. If the received block contains errors and therefore is *unacceptable*, the receiver indicates this by sending a NAK signal to the transmitter. The transmitter responds to this *repeat request* by retransmitting the last block.

The throughput efficiency of the stop-and-wait ARQ scheme is highly dependent on the channel quality and the *waiting* time between transmission of a block and receipt of the ACK or NAK. Clearly the quality of the channel determines the number of retransmissions required. The waiting time consists of the round-trip propagation time, plus the link-and-modem turnaround times (if the channel operates in the half-duplex

mode), plus the time duration of the forward (data) and return (ACK and NAK) signals. Waiting time can be so excessive on channels such as satellite channels that more time is spent waiting than in delivering good data to the receiver. Consequently throughput efficiency may be poor when using the stop-and-wait ARQ scheme even on good-quality channels, although this efficiency can be optimized by proper blocklength choice, as shown in later sections.

Variation 1 of the Stop-and-Wait ARQ Scheme

In the basic scheme, whenever the receiver requests that a block be repeated, the block is repeated one time. In a variation [6], referred to herein as variation 1, whenever the transmitter receives a repeat-request, m duplicates of the repeat-requested block are successively transmitted before the transmitter stops and waits for the receiver acknowledgment. With this technique a repeat-request will be indicated again only when none of the m successively transmitted duplicates are received correctly.

This technique spreads one normal waiting-time delay for stop-and-wait systems over m duplicates, with m being the expected number of duplicates that would ordinarily be required for correct receipt in a given poor-quality channel. Thus improvement in throughput performance depends on the expected number of retransmissions per block and the system round-trip delay.

Variation 2 of the Stop-and-Wait ARQ Scheme

In another variation on the basic stop-and-wait ARQ scheme, referred to herein as variation 2, a sequence of M blocks (a superblock) is transmitted before the transmitter stops and waits for receiver acknowledgment. When the receiver desires a repeat of some of the M blocks, it identifies their positions within the received sequence. The transmitter responds by placing these repeat-requested blocks at the beginning of the next sequence (superblock). Only negative (NAK) acknowledgments by the receiver are needed once communication is established; thus the feedback channel is used more efficiently.

This technique spreads the normal waiting-time delay (more accurately, the propagation delay plus the average acknowledgment-message time) for the basic stop-and-wait system over M blocks. In other words, the time intervals are longer between the stop-and-wait points, but only those individual blocks within the received superblock which are found in error are retransmitted. This variation of the technique should yield improved throughput performance over the basic stop-and-wait scheme for channels of all qualities. However the cost to implement this scheme is greater, due to increased storage and block-reordering requirements.

Continuous-ARQ Schemes

The continuous-ARQ schemes reduce the waiting time between blocks and consequently are more efficient. In exchange for the increased efficiency these schemes require

a simultaneous two-way data link, a more complex implementation, and larger temporary storage for several transmitted but unacknowledged blocks of data. The number of temporarily stored unacknowledged blocks N is the number of blocks in transit or received but unacknowledged at the transmitter. The two basic continuous-ARQ schemes are the selective-repeat scheme and the go-back-N scheme. Both schemes require the transmitter to continue transmitting subsequent blocks while awaiting receiver acknowledgment (ACK/NAK) for an unconfirmed block. Thus both schemes productively use their acknowledgment waiting time. These two schemes and variations of them differ in what is retransmitted.

Go-Back-N Continuous-ARQ Scheme

The go-back-N ARQ scheme [1-6,8] is the easiest continuous-ARQ scheme to implement but also, the least efficient. The transmitter sends blocks of data in sequence without waiting for an acknowledgment (ACK or NAK). This "continuous" transmission is stopped only when a NAK is received at the transmitter. The NAK indicates that a block was received in error and a retransmission is required. In response the transmitter stops sending new blocks and backs up to retransmit the NAK'ed block and all the following transmitted but unacknowledged blocks.

The go-back-N ARQ scheme is efficient for good-quality channels. It requires no block reordering or resequencing logic at the receiver, but it retransmits $N - 1$ extra blocks for each block found in error. Generally either these extra blocks or, usually, the originals are discarded without using them to improve error-control efficiency. However this may be a good error-control technique for burst error environments. This scheme suffers from the same type of delay or waiting time as the stop-and-wait ARQ when a block is received in error. Thus the average waiting time depends directly on the block error rate (which is a function of blocklength), and the throughput efficiency can be optimized for specific channels by proper choice of blocklength.

Variation of the Go-Back-N Continuous-ARQ Scheme

A variation is a modified go-back-N ARQ scheme [6] that differs from the basic go-back-N ARQ scheme in the content of the N retransmitted blocks. Instead of retransmitting the last N transmitted blocks, in this modified scheme the repeat-requested block is continuously retransmitted at least N times until received correctly (until *accepted*). Once the acceptance acknowledgment is received at the transmitter, the original block sequence transmission picks up from the block following the last accepted block. Thus $N - 1$ duplicates of the repeat-requested block are received but ignored before the receiver sees continuation of sequence.

The improvement in throughput provided by this variation of the go-back-N ARQ scheme is gained by the reduction of time required for the total number of retransmissions for a repeat-requested block. In this instance it is not necessary to waste time by transmitting the $N - 1$ subsequent blocks, since they get ignored with each repeat-request

retransmission. The amount of throughput performance improvement depends on the expected number of retransmissions and is more significant in poor-quality, burst error channels, in which a large number of block retransmissions are likely to be required.

Selective-Repeat Continuous-ARQ Scheme

The selective-repeat scheme [1-4] is a more complex continuous-ARQ scheme with respect to implementation but also is the most efficient in throughput efficiency. The transmitter sends blocks of data in sequence without waiting for an acknowledgment (ACK or NAK). This "continuous" transmission is interrupted only when a NAK is received at the transmitter indicating that a block was received in error and a retransmission is required. In response the transmitter stops the normal block-sequence flow long enough to transmit the requested block. The selective-repeat ARQ scheme is efficient for good- and moderate-quality channels. The additional implementation complexity is due to the block reordering or resequencing logic required, since some accepted blocks will be out of their original order.

Hybrid ARQ Schemes

In the hybrid ARQ schemes [7-9] forward-error-correction (FEC) techniques are used in conjunction with a basic ARQ technique to improve the throughput performance on poor-quality channels. The purpose of the FEC component is to improve the channel quality (reduce error rate) experienced by the basic ARQ error-detecting-code component, thus reducing the high expected number of block retransmissions in high-error-rate channels.

There are two methods of incorporating the FEC. The first is to use an error-correcting code to encode blocks which have been encoded with an error-detection code [8,9]. Thus this method uses a two-step coding procedure. The second method uses a single code for both error detection and error correction [7]. This dual capability is exploited at the receiver by the decoders. The error-correcting component can be selected for independent bit errors, burst errors, or a combination of both conditions. This capability may in some instances be used or made operative only when the channel conditions warrant it.

THROUGHPUT PERFORMANCE OF THE VARIOUS ARQ SCHEMES

In this section throughput efficiency equations for the various ARQ schemes discussed in the preceding section are developed and evaluated. Throughput performance curves for some of these ARQ schemes are presented. Consequently a data-transmission-system design engineer is provided with information to help in choosing error-control techniques or in assessing the performance of proposed ARQ designs.

Preliminary Notation, Definitions, and Assumptions

The throughput equations presented in this section for the various ARQ schemes are expressed using common notation defined as follows:

η — Throughput efficiency expressed as a percent of the symbol transmission rate;

R — Symbol transmission (signaling) rate of a system in symbols per second;

D — System delay: the roundtrip propagation delays plus the transmitter and receiver response times for receipt of blocks or acknowledgement of messages;

B_I — Number of user information symbols per message block;

B_S — Number of system information symbols per transmitted message block: B_I plus the number of protocol or reader information symbols per message block;

B — Number of total symbols per transmitted message block after encoding: B_S plus the number of error-control redundancy symbols per message block;

$E = E\{B\}$ — Expected number of transmissions of a given block with blocklength B before correct reception;

$P_e = P_e\{B\}$ — Probability of a given block with blocklength B being in error

α — Transmission efficiency of a system: B divided by the average number of symbols used for correct reception;

β — Background efficiency of a system: the throughput efficiency that a particular system would have if the channel were error-free.

The derivations of throughput equations in this section are based on some common simplifying assumptions:

- The acknowledgment channel is error-free or is *acceptable* due to use of forward-error-correction (FEC) coding. Equivalently the probability of receiving an incorrect acknowledgment message is negligible.

- The error-detection-coding component of the various ARQ schemes essentially detects all errors. Equivalently the probability of not detecting errors in a received block is negligible.

- System delay time is constant for a given connection, whether an ACK or NAK acknowledgment is required.

- Steady-state operation of the transmission system is the only mode considered. Block synchronization failure or acquisition are not considered, or they are no more damaging to throughput than normal block errors.

- In most cases considered, except where noted, the block errors are independent from block to block, because the channel produces independent error bursts and because the error bursts do not appreciably overlap adjacent blocks.

- In most cases considered, except where noted, each block consists of *real* information; that is, each message length is exactly an integral number of blocklengths.

With these assumptions, the expected value $E = E\{B\}$ of the number of transmissions n is

$$E\{B\} = \sum_{n=1}^{\infty} n(P_e\{B\})^{n-1}(1 - P_e\{B\}) = (1 - P_e\{B\})^{-1} = (P_a\{B\})^{-1}, \tag{1}$$

where n is the number of transmissions required for acceptance of an arbitrary block, and $1 - P_e\{B\} = P_a\{B\}$ is the probability of no errors detected in the block: the block's *acceptability*.

Throughput of the Basic Stop-and-Wait ARQ Scheme

In the stop-and-wait ARQ scheme, for each correct reception of B_I information symbols, $B + DR$ symbol periods must be expended E times on the average. Thus, the throughput [1-6] efficiency to the demand-access user is

$$\eta_{SW} = \frac{B_I}{(B + DR)E}. \tag{2}$$

The expression $DR + (B + DR)(E - 1)$ can be interpreted as the expected *wasted time*, expressed as symbol periods, due to system delays and required block retransmissions.

Curves of η_{SW} for the aforementioned assumptions are exhibited in Fig. 1 for the case of $B_I = 200$ bits and $B = 255$ bits. Throughput efficiency essentially attains its maximum value given by $\beta_{SW} = B_I/(B + DR)$ when $P_e \leqslant 10^{-2}$. Also, $\alpha_{SW} = E^{-1}$. Observe the dependence of β_{SW} on DR for a given block length (B_I and B), or, conversely, observe the dependence of β_{SW} on blocklength (B_I and B) for given DR. The choice of optimal or "good" blocklengths with respect to throughput efficiency will be explored in the next main section.

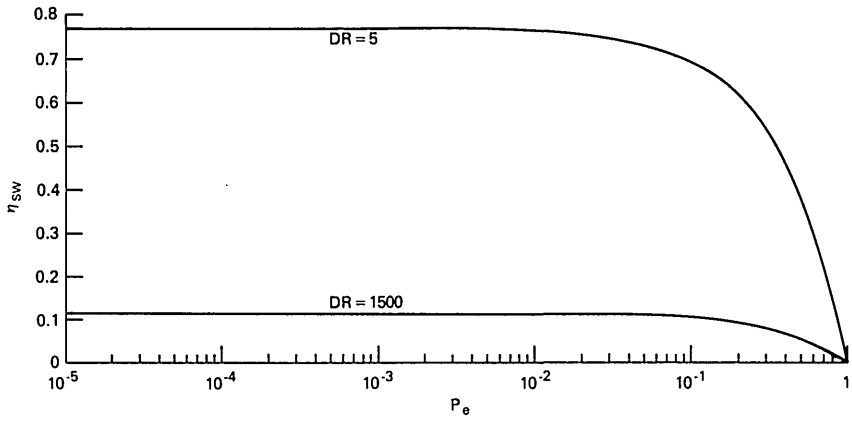


Fig. 1 — Stop-and-wait (SW) throughput efficiency versus block error rate for the case of B_I = 200 bits and B = 255 bits, calculated using Eqs. (2) and (1)

Throughput of Variation 1 of the Stop-and-Wait ARQ Scheme

In variation 1 on the basic stop-and-wait ARQ scheme, for each correct reception of B_I information symbols, $[(B + DR) + (mB + DR)P_e/(1 - P_e^m)]$ symbol periods on the average must be expended. This expression is obtained as follows. For each set of m duplicate blocks B (each containing the same B_I information symbols) the probability that each set is the last is $(1 - P_e^m)$ and the probability that the nth set is required is $P^{(n-1)m+1}$. The expected number of sets (each of m duplicate blocks) transmitted is then

$$E_1 = E_1\{B\} = \sum_{n=1}^{\infty} n(1 - P_e^m)P_e^{(n-1)m+1}$$

$$= P_e(1 - P_e^m) \sum_{n=1}^{\infty} n(P_e^m)^{n-1} = P_e/(1 - P_e^m).$$

Thus, $B + DR$ symbol periods are expended on the first transmission of B_I, and $(mB + DR)E_1$ additional symbol periods on the average are required for receiver acceptance of B_I. The throughput efficiency [6] for this scheme is then

$$\eta_{SW1} = \frac{B_I}{(B + DR) + (mB + DR)E_1}.$$

In this case $DR + (mB + DR)E_1$ expresses the expected *wasted time* in symbol periods due to system delays and required block retransmissions; also,

$$\alpha_{SW1} = \left[1 + \frac{m + (DR/B)}{1 + (DR/B)} E_1\right]^{-1} \quad \text{and } \beta_{SW1} = B_I/(B + DR).$$

For $m = 1$ we have the basic stop-and-wait throughput efficiency η_{SW}.

This variation yields better performance than the basic scheme ($\eta_{SW1} > \eta_{SW}$) if m satisfies the inequality

$$\frac{m + (DR/B)}{1 - P_e^m} < \frac{1 + (DR/B)}{1 - P_e}. \tag{3}$$

The optimum value m_o for given P_e, D, R, and B is obtained by minimizing the term $(mB + DR)P_e/(1 - P_e^m)$ with respect to m. This results in

$$B = \frac{DR \ \ell n \ P_e}{1 + \ell n \ P_e^{-m_o} - P_e^{-m_o}}, \tag{4}$$

which defines the minimum over $m \geqslant 1$ as long as

$$\frac{-DR}{B} \ \ell n \ P_e \geqslant 2\left(\frac{1 - P_e^{m_o}}{1 + P_e^{m_o}}\right) + \ell n \ P_e^{m_o}.$$

But the left-hand side is positive and the right-hand side is negative for all positive B and m_o, since $P_e \leqslant 1$; therefore (4) does define the minimum. The scheme generally yields effective improvement over the basic stop-and-wait ARQ scheme for DR/B much larger than m and for $P_e \geqslant 0.1$. This implies relatively short blocklengths in systems characterized by long system delays and high block error rates. Thus situations may exist in which η_{SW} becomes so low that, even with the improvement gained by this suggested variation, implementation may not be cost effective. Table 1 illustrates the evaluation of the left-hand side of the inequality (3) for DR = 1500 and B = 255.

Figure 2 illustrates the relationship between m_o and P_e for B = 255 and DR = 1500. The corresponding throughput efficiency η_{SW1} is shown in Fig. 3. Figure 4 is a family of curves for B versus -log $(1 - P_e)$ parameterized by m. Note that -log $(1 - P_e)$ increases with P_e.

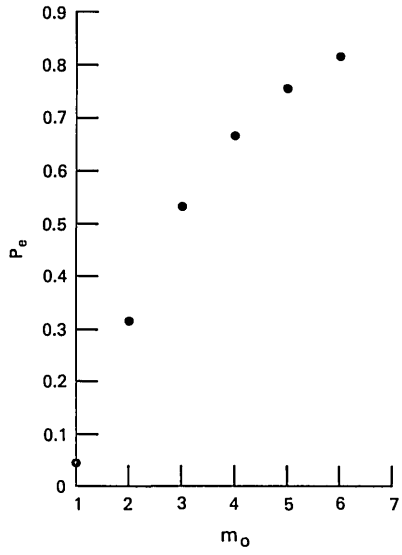


Fig. 2 — Block error rate versus optimal number of duplicate blocks for stop-and-wait variation 1 for the case of B_I = 200 bits, B = 255 bits, and DR = 1500

Table 1 — Values for the Left-Hand Side of Eq. (3) with DR = 1500 and B = 255

m	$[m + (DR/B)]/(1 - P_e^m)$				
	$P_e = 0.1$	$P_e = 0.2$	$P_e = 0.3$	$P_e = 0.4$	$P_e = 0.5$
1	7.647	8.603	9.832	11.471	13.765
2	7.962	8.211	8.662	9.384	10.51
3	—	8.954	9.129	9.49	10.151
4	—	—	9.963	10.142	10.541
5	—	—	—	10.995	11.233
6	—	—	—	11.931	12.071
7	—	—	—	—	12.984
8	—	—	—	—	13.937

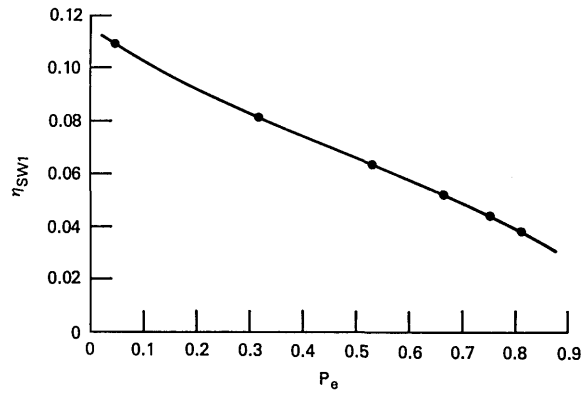


Fig. 3 — Throughput efficiency of stop-and-wait variation 1 versus block error rate using m_o for the case of B_I = 200 bits, B = 255 bits, DR = 1500

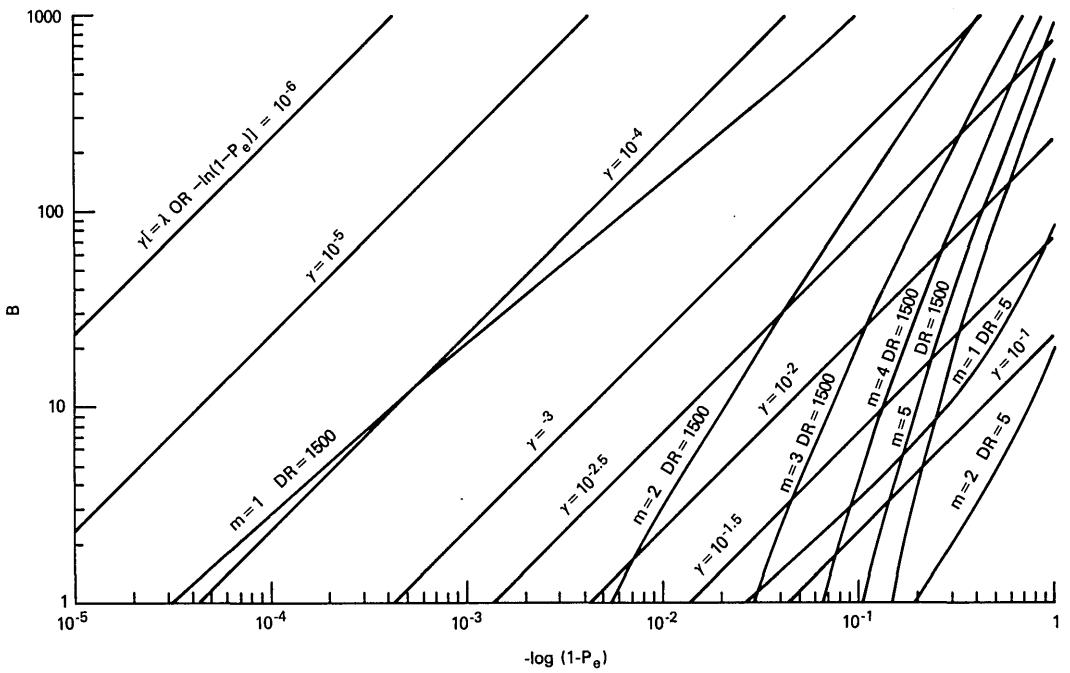


Fig. 4 — Plots of blocklength versus -log(1 - P_e) relating the families of curves $B = DR\ell nP_e / 1 + \ell nP_e^{-m_o} - P_e^{-m_o}$ and $P_e = 1 - e^{-\gamma B}$ for stop-and-wait variation 1

Throughput of Variation 2 of the Stop-and-Wait ARQ Scheme

In variation 2 on the basic stop-and-wait ARQ scheme, for each correct reception of B_I information symbols, $[B + (DR/M)]E$ symbol periods on the average must be expended. The term (DR/M) is the average fraction of total system delay assigned to an individual block. Thus the throughput efficiency for this scheme is

$$\eta_{SW2} = \frac{B_I}{[B + (DR/M)]E} = \frac{MB_I}{(MB + DR)E}. \tag{5}$$

In this case $(DR/M) + [B + (DR/M)](E - 1)$ expresses the expected wasted time in symbol periods due to system delays and required block retransmissions. Also, $\alpha_{SW2} = E^{-1}$ and $\beta_{SW2} = MB_I/(MB + DR)$. For $M = 1$ we have the basic stop-and-wait throughput efficiency. This variation yields better performance than the basic scheme if $M > 1$, as is clear from Eq. (5). Performance curves of η_{SW2} are the same shape as for η_{SW}. Their correct interpretation in this case requires multiplication by the factor $(B + DR)/(B + DR/M)$. Thus observations made on the η_{SW} performance curves are directly applicable to η_{SW2} performance. For instance, if $DR = 1500$ and $M = 300$, this variation would yield a performance displayed by the $DR = 5$ curve of Fig. 1 instead of the $DR = 1500$ curve.

Throughput of the Go-Back-*N* Continuous-ARQ Scheme

In the go back-*N* continuous-ARQ scheme $(B + DRP_e)(1 - P_e)^{-1}$ symbol periods must be expended on the average ($B + DRP_e$ symbol periods E times on the average) for each correct reception of B_I information symbols. The DRP_e term is the average delay experienced by a given transmitted block B: the delay incurred when a block error occurs times the probability of block error. Thus the throughput [1-6, 8, 9] efficiency for this scheme is

$$\eta_G = \frac{B_I}{(B + DRP_e)E}, \tag{6a}$$

which can also be expressed as

$$\eta_G = \frac{B_I}{B}\left(1 + \frac{NP_e}{1 - P_e}\right)^{-1} \tag{6b}$$

$$\approx \frac{B_I}{B}(1 - NP_e) \tag{6c}$$

for $NP_e/(1 - P_e) \approx NP_e << 1$, where $N = 1 + (DR/B)$. The scheme receives its name from the fact that B/R is usually chosen so as to make N have integer values. For example, if $D = kB/R$, then the transmitter goes back by $N = k + 1$ blocks when a repeat-request is received. Either the system delay must be essentially constant or N is variable and thus the transmitter must know its current value (then the average N is used in Eqs. (6). For this scheme the expected wasted time is $DRP_e + (B + DRP_e)(E - 1)$ or equivalently $BNP_e (1 - P_e)^{-1} = BNP_eE = BN(E - 1)$. In addition $\beta_G = B_I/B$ and $\alpha_G = \{1 + [NP_e/(1 - P_e)]\}$.

Figure 5 displays η_G as a function of $P_e\{B\}$ for our representative examples. The curves have the same value at low P_e, but the performance for $N = 7$ falls off sooner than for $N = 1$. This reflects the degrading effects of long delays.

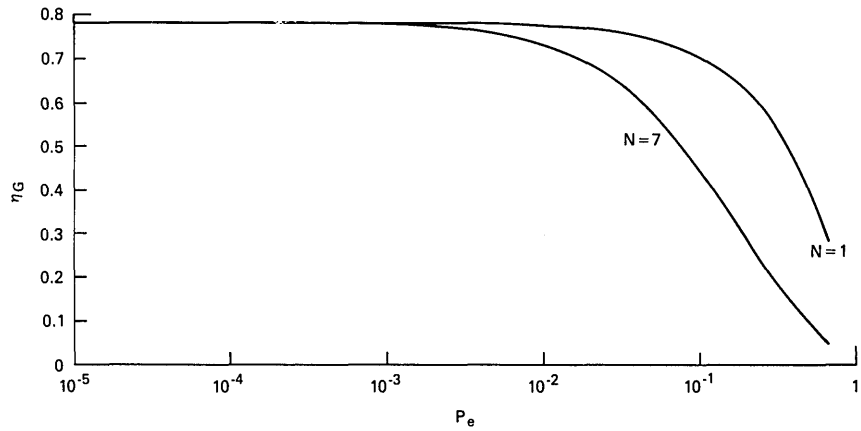


Fig. 5 — Throughput efficiency of the go-back-N scheme versus block error rate for $B_I = 200$ bits and $B = 255$

Throughput of the Variation of the Go-Back-N Continuous-ARQ Scheme

The variation on the basic go-back-N scheme improves the throughput performance by reducing the wasted time involved in retransmission of the repeat-requested block plus $N - 1$ subsequent blocks. Here $B\{1 + P_e\ [E + 2(N - 1)]\}$ symbol periods must be expended on the average for each correct reception of B_I information symbols. Thus the throughput efficiency [6] in this instance is

$$\eta_{G1} = \frac{B_I}{B\{1 + P_e[E + k(N - 1)]\}}, \text{ for } k = 2.$$

The term $2(N - 1)$ is the number of blocks transmitted but ignored either before the first NAK or after the first ACK is received at the transmitter. The expression $P_eB[E + 2(N - 1)] = B(E - 1) + 2(N - 1)BP_e$ is the expected wasted time due to system delays and required block retransmissions. Transmission and background efficiencies for this scheme are $\alpha_{G1} = \{1 + P_e[E + 2(N - 1)]\}^{-1}$ and $\beta_{G1} = B_I/B$.

This variation yields better throughput performance than the basic go-back-N scheme ($\eta_{G1} > \eta_G$) if

$$1 + \frac{NP_e}{1 - P_e} > 1 + \frac{P_e}{1 - P_e} + 2(N - 1)P_e$$

or

$$P_e = P_e\{B\} > \frac{1}{2}.$$

The implication of this requirement is that the basic go-back-N scheme is better than this variation under most practical conditions ($P_e \leqslant 1/2$). If the term $2(N - 1)$ is reduced by a factor of 2, then the resulting equations are

$$\eta_{G1} = \frac{B_I}{B[1 + P_e[E + k(N - 1)]]}, \text{ for } k = 1,$$

$$1 + \frac{NP_e}{1 - P_e} > 1 + \frac{P_e}{1 - P_e} + (N - 1)P_e,$$

and

$$P_e = P_e\{B\} > 0.$$

Thus, if the number of blocks transmitted but-ignored, $2(N - 1)$, can be reduced to $N - 1$, then this modified variation yields improved performance for any block error rate. However reducing the number of transmitted-but-ignored blocks requires additional processing and storage for the $N - 1$ blocks subsequent to the block received in error. For example, by the time the transmitter receives the first NAK for a repeat-requested block, the receiver could have processed (detected errors in) the next $N - 1$ or less blocks from the original sequence. Thus, when the repeat-requested block is received correctly, there is no need to wait an additional $N - 1$ symbol periods before processing and acknowledging the subsequent block in the original sequence. In addition, after each received ACK, transmission is initiated for the next block in the original sequence which has not as yet been transmitted.

Throughput efficiencies η_{G1} and η'_{G1} are shown in Fig. 6 for the example of B_I = 200 bits, B = 255 bits, and N = 1 and 7. Throughput for k = 1 is better or as good as the throughput for k = 2 in this example. The gain in performance is more pronounced at high block error rates and long delays.

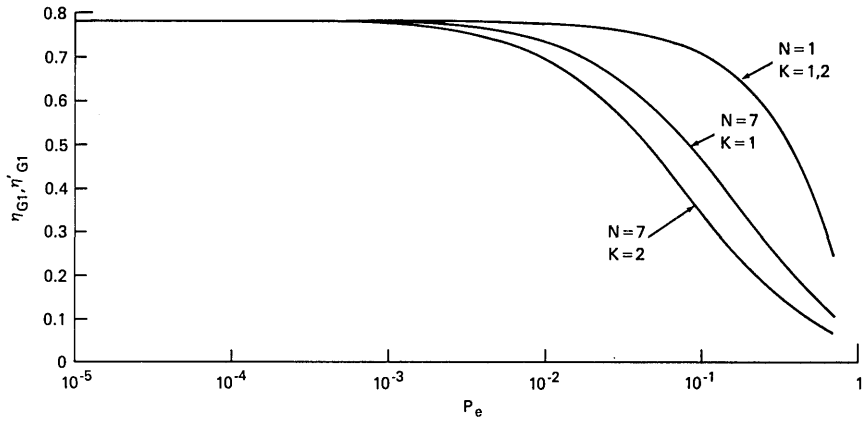


Fig. 6 — Throughput efficiency of the variation of the go-back-N scheme versus block error rate for B_I = 200 bits and B = 255

Throughput of the Selective-Repeat Continuous-ARQ Scheme

In the selective-repeat continuous-ARQ scheme, which is the most efficient of the various ARQ schemes, the stop-and-wait-for-acknowledgment period and other system delays are essentially eliminated. Consequently, for each correct reception of B_I information symbols, B symbol periods must be expended E times on the average. The throughput [1-4] efficiency to the demand-access user in this case is

$$\eta_{SR} = \frac{B_I}{BE} .$$

The expression $B(E - 1)$ is the expected wasted time due to required block retransmissions. Transmission efficiency and background efficiency for this scheme are $\alpha_{SR} = E^{-1}$ and $\beta_{SR} = B_I/B$ respectively. From the form of η_{SR}, throughput efficiency for the selective-repeat scheme is proportional to $\alpha_{SR} = E^{-1}$, the same as for the stop-and-wait scheme. Figure 7 depicts η_{SR} as a function of P_e for our representative example with B_I = 200 bits and B = 255 bits. Comparison of this curve with Fig. 1 shows that as expected the stop-and-wait ARQ performance approaches the selective-repeat ARQ performance as the delays decrease.

Throughput Error of the Hybrid ARQ Scheme with Separate Error Detection and Correction Encoding

In the first hybrid scheme, the concern again is for the average number of symbol periods expended in order for B_I information symbols to be accepted at the receiver (correctly received). The error-correction capability of an error-correcting code is determined by a measure of the distance between the code words. For block codes and convolutional codes this distance measure is the minimum Hamming distance $d_{\min}$. (According to Ref. 10, p. 104, the minimum distance for convolutional codes tend to be slightly higher than the corresponding results for block codes for given values of n and k. Thus the minimum-distance bound for block codes will be used for both block and convolutional codes in this analysis.)

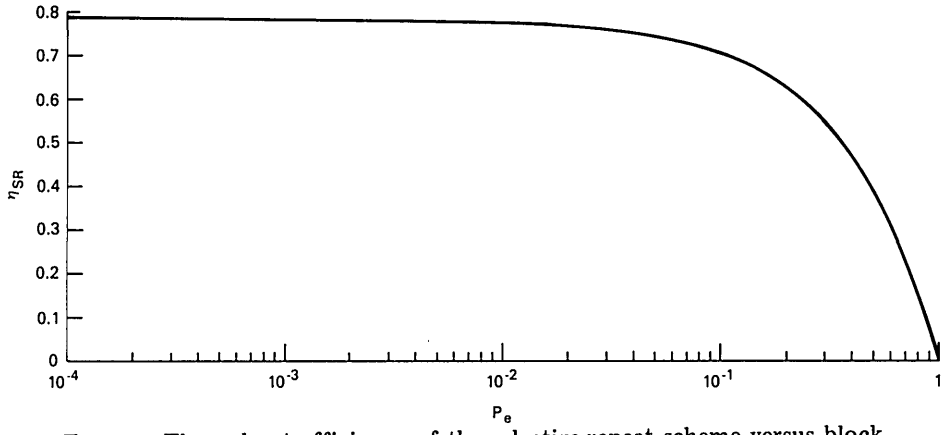


Fig. 7 — Throughput efficiency of the selective-repeat scheme versus block error rate for B_I = 200 bits and B = 255

The error-correction capability of a block code is given as

$$t = (d_{\min} - 1)/2;$$

that is, a code with minimum distance $d_{\min}$ can correct up to t symbol errors per codeword. Thus the maximum bit error rate per codeword for which the code provides complete protection is

$$\rho_{\max} = t/B_c\,,$$

where $B_c = B$ plus error-correcting-code redundancy symbols. When independent symbol errors are assumed, the probability P_e that a received B_c-tuple is not decoded into an error-detecting-code codeword (thus requiring a repeat-request) satisfies

$$P_e \leqslant \sum_{j=t+1}^{B_c} \binom{B_c}{j} p_e^j (1 - p_e)^{B_c - j} = 1 - \sum_{j=0}^{t} \binom{B_c}{j} p_e^j (1 - p_e)^{B_c - j}$$

where $B_c = B/r_c$, p_e is the bit error rate, and r_c is the least code rate for the best t-error correcting block code given by the Varsharmov-Gilbert lower bound [10, section 4.1] on minimum distance. Thus for a given B and P_e the optimum t and r_c can be determined which minimizes this upper bound on P_e. More precisely, the term

$$\sum_{j=t+1}^{(B/r_c)} \binom{B/r_c}{j} p_e^j (1 - p_e)^{(B/r_c)-j} = 1 - \sum_{j=0}^{t} \binom{B/r_c}{j} p_e^j (1 - p_e)^{(B/r_c)-j}$$

is minimized, subject to the constraint on t and r_c

$$\sum_{i=0}^{2t-1} \binom{B/r_c}{i} \geqslant 2^{(1/r_c-1)B}.$$

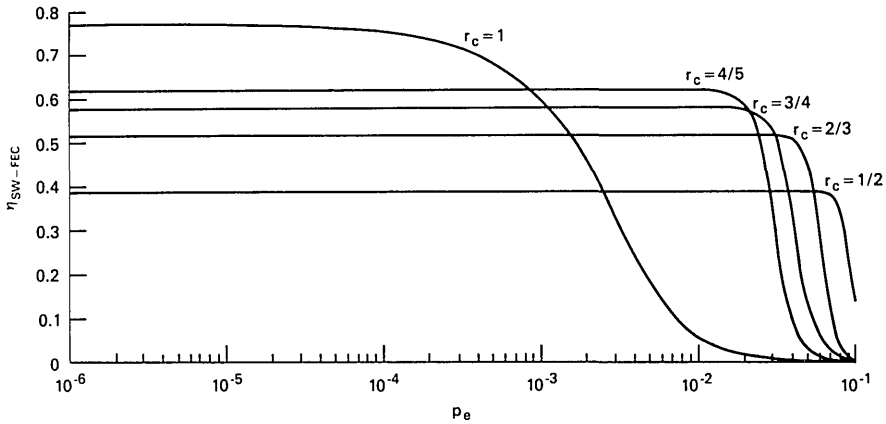
This inequality is a form of the Varsharmov-Gilbert lower bound [theorem 4.7 on page 87 in Ref. 10]. The minimum upper-bound-block-error probability P_e^o and the corresponding code rate r_c^o (hence $B_c^o = B/r_c^o$) can be substituted in any of the throughput-efficiency equations to determine a lower bound on the throughput efficiency of the various ARQ schemes when the best error-correcting code is added. Figures 8a and 8b display a lower bound on the stop-and-wait efficiency η_{SW} versus p_e (BER) for $DR = 5$ and $DR = 1500$ respectively. The five curves correspond to the code rates 1, 4/5, 3/4, 2/3, and 1/2. In addition the block error rate P_e versus p_e is depicted in Fig. 9 for the five code rates. As the code rate decreases, protection for high-error-rate conditions (high values of p_e) is gained at the expense of lower low-error-rate performance. This is to be expected, since low code rates require additional redundant bits per transmitted information block. In addition the points of intersection in Fig. 8 for the $r_c = 1$ curve and any of the other curves indicate at what value of p_e error-correction coding should be added or dropped.

Since the minimum-distance lower bounds for block codes can be used for convolutional codes as well, the probability of erroneous decoding P_e for the best (mB_c, mB) convolutional code is bounded above by

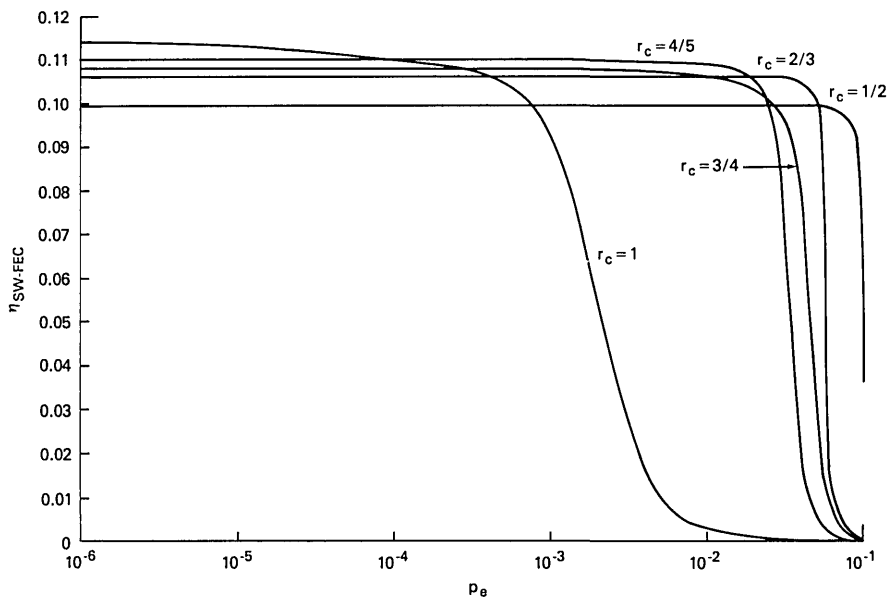
$$P_e \leqslant \sum_{j=[d/2]}^{d-1} \sum_{i=d-j}^{j} \sum_{h=d}^{j+i} \binom{mB_c-j}{\frac{h+i-j}{2}} \binom{j}{\frac{h+j-i}{2}} \frac{\binom{mB_c}{j} p_e^j (1-p_e)^{mB_c-j}}{2mB_c - mB - 1}$$

$$+ \sum_{j=d}^{mB_c} \binom{mB_c}{j} p_e^j (1-p_e)^{mB_c-j},$$

where $B_c = B/r_c$, $[d/2]$ denotes the integer part of $d/2$, and m is the constraint length of the code. Other bounds on P_e have been derived which could be used here. For example Viterbi [11] has derived an upper bound on P_e when Viterbi decoding is used. Again these bounds can be used in the aforementioned throughput-efficiency equations to determine performance with respect to the added convolutional coding.



(a) DR = 5



(b) DR = 1500

Fig. 8 — Stop-and-wait-plus-FEC throughput-efficiency lower bound versus bit error rate for various FEC rates r_c with B_I = 200 bits and B = 255.

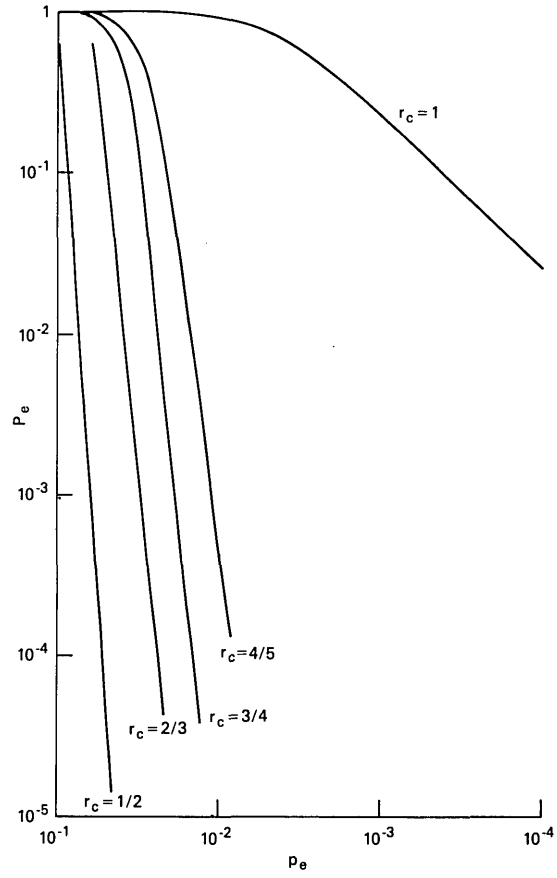


Fig. 9 — Block-error-rate upper bound versus bit error rate of the best error-correcting-code for the different code rates in Fig. 8 and B_I = 200 bits and B = 255

OPTIMAL BLOCKLENGTHS FOR ARQ SYSTEMS

As evident from the throughput and wasted-time expressions that were derived, the message blocklength B is a prevalent factor of throughput efficiency. Consequently the knowledge of optimal blocklengths which maximize throughput efficiency (or conversely minimize the wasted time) is important in the efficient design of any of the ARQ systems discussed earlier [5, 12, 13].

The message length of any message source generally varies from one message to the next and can best be described by a probability distribution. Usually, for efficient processing and transmission, the random-length messages are partitioned into several fixed-length blocks. Admittedly the last partitioned block usually cannot be entirely filled by the message and is either filled with dummy information or is terminated by an end-of-message indication. Thus the message blocklength must be chosen with these facts in mind [12, 13]. In addition there is a tradeoff in selecting the optimal blocklength. On the one hand it is desirable to select the largest blocklength to minimize the number of acknowledgments and

the length of retransmissions. On the other hand it is desirable to select the smallest possible blocklength to minimize the block-error probability and to minimize the wasted time due to the last unfilled partitioned block (when an end-of-message indicator is not used).

The goal of this section is to optimize the throughput efficiency η with respect to the blocklength B. The standard procedure is to obtain the derivative of η with respect to B, set it to zero, and determine if this leads to the global maximum; otherwise other techniques must be applied. However, if the expected wasted-time expressions for the various ARQ schemes are normalized by the blocklength B, it is clear from the form of η for all schemes that it is sufficient to minimize this normalized expected-wasted-time expression with respect to B. This procedure ignores however the situation in which the last block of the message may not be completely filled with useful data. Nevertheless a good approximation is sought without including that facet of the problem. The description of block error rate $P_e = P_e\{B\}$ of particular concern is

$$P_e = 1 - e^{-\gamma B}, \tag{7}$$

where

$$\gamma = -\ln(1 - p_e)$$

for the independent bit error environment, with p_e being the bit error rate, and where

$$\gamma = \lambda ,$$

the mean block error rate for the Poisson-distributed block error environment.

Optimal Blocklength for the Basic Stop-and-Wait ARQ Scheme

The expected-wasted-time expression $DR + (B + DR)(E - 1)$ for the basic stop-and-wait ARQ scheme has the following form when normalized by B:

$$\begin{aligned} f(B) &= (E - 1) + DRB^{-1}E \\ &= \frac{P_e}{1 - P_e} + \frac{DR}{B(1 - P_e)}, \end{aligned} \tag{8}$$

where $E = (1 - P_e)^{-1}$. Differentiation of this expression with respect to B yields the stationary-point equation

$$\frac{dP_e}{dB} = \frac{DR(1 - P_e)}{B(B + DR)},$$

from which the minimum can be determined. Under the assumption of Eq. (7) the stationary-point equation is

$$\gamma = \frac{DR}{B(B + DR)} .$$

This leads to the stationary-point expression [5]

$$B^o = DR\left[\left(\frac{1}{4} + \frac{1}{\gamma DR}\right)^{1/2} - \frac{1}{2}\right]. \tag{9}$$

To determine whether or not the stationary point is the global minimum for this characterization, the behavior of $f'(B)$ must be better understood. If we substitute say Eq. (7) in (8) and differentiate, we find

$$f'(B) = \frac{\gamma B^2 + \gamma DRB - DR}{B^2 e^{-\gamma B}} .$$

The denominator on the right-hand side is positive. The numerator is quadratic in B with positive leading term and two real roots: one at B^o and the other obtained by taking the square root in Eq. (9) with a negative sign. Therefore f' is negative between the roots of the numerator and positive elsewhere. But the second root is negative; therefore f' is negative on the interval $(0, B^o)$ and positive on (B^o, ∞). It follows that for positive blocklengths (the only ones of interest) f has its minimum at B^o. Figure 10 displays $f(B)$ for λ, $p_e = 10^{-1}$, 10^{-3}, and 10^{-6}, and $DR = 5$ and 1500, where $B_I = 200$ and $B = 255$. Note that B^o is highly sensitive to variations in λ and p_e; thus the choice of a single blocklength for all error rates will not yield good uniform performance.

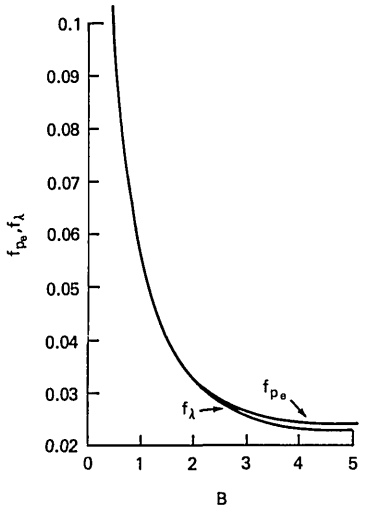
Optimal Blocklength for Variation 1 of the Stop-and-Wait ARQ Scheme

The normalized-wasted-time expression to be minimized in the case of variation 1 of the stop-and-wait ARQ scheme is

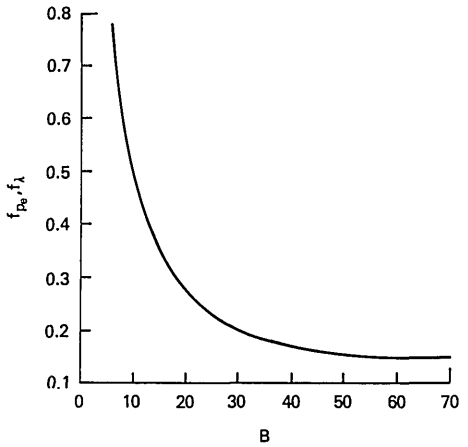
$$f(B) = DRB^{-1}(E_1 + 1) + mE_1,$$

where m is the number of duplicate blocks per set transmitted (which may have been chosen to satisfy Eq. (4)) and $E_1 = P_e/(1 - P_e^m)$ is the expected number of sets transmitted. Thus

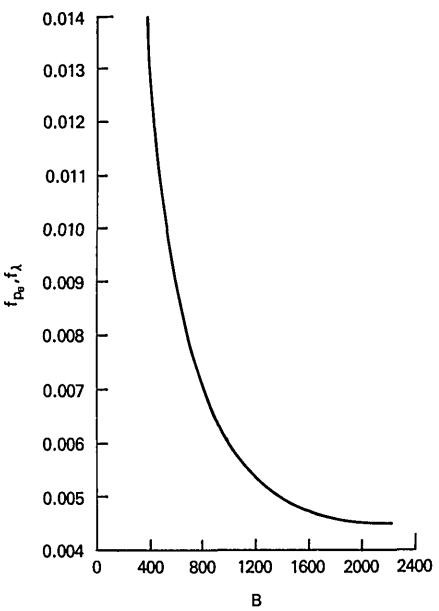
$$f(B) = \frac{DR}{B}\left(1 + \frac{P_e}{1 - P_e^m}\right) + \frac{mP_e}{1 - P_e^m} .$$



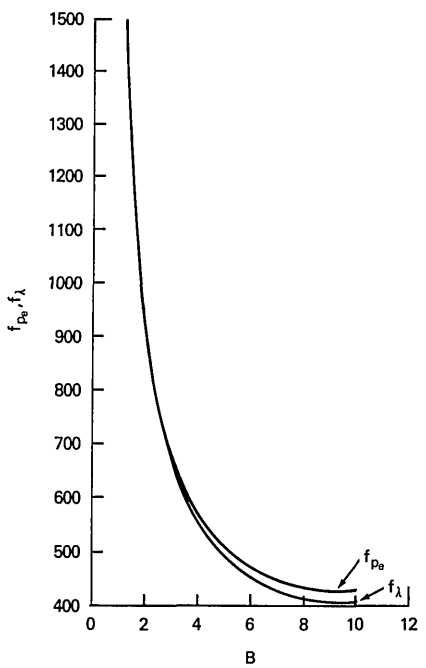
(a) $DR = 5$, $f_{p_e}(B)$ at $p_e = 10^{-1}$, $B^o_{p_e} = 4.8$, and $f_\lambda(B)$ at $\lambda = 10^{-1}$, $B^o_\lambda = 5.0$



(b) $DR = 5$, $f_{p_e}(B)$ at $p_e = 10^{-3}$, $B^o_{p_e} = 68.2$, and $f_\lambda(B)$ at $\lambda = 10^{-3}$, $B^o_\lambda = 68.3$

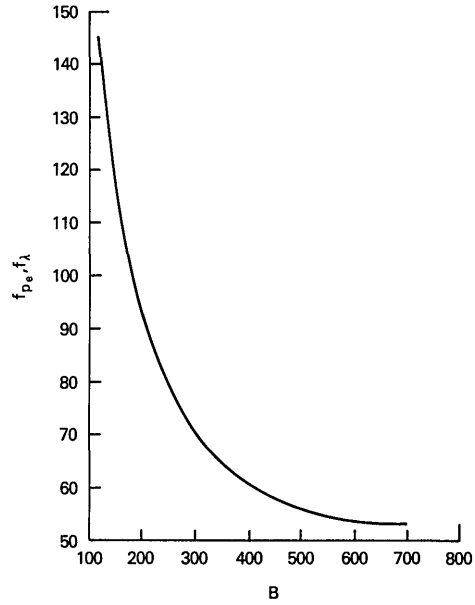


(c) $DR = 5$, $f_{p_e}(B)$ at $p_e = 10^{-6}$, $B^o_{p_e} = 2219$, and $f_\lambda(B)$ at $\lambda = 10^{-6}$, $B^o_\lambda = 2234$

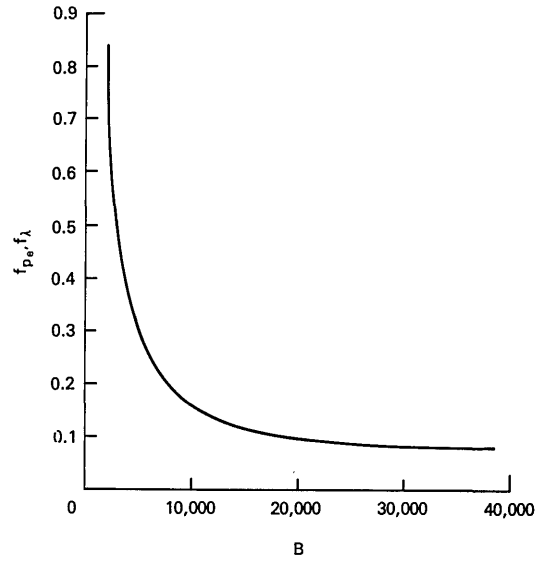


(d) $DR = 1500$, $f_{p_e}(B)$ at $p_e = 10^{-1}$, $B^o_{p_e} = 9.4$, and $f_\lambda(B)$ at $\lambda = 10^{-1}$, $B^o_\lambda = 9.9$

Fig. 10 — Stop-and-wait wasted-time functions versus blocklength for $B_I = 200$ and $B = 255$



(e) $DR = 1500$, $f_{p_e}(B)$ at $p_e = 10^{-3}$, $B^o_{p_e} = 685.9$ and $f_\lambda(B)$ at $\lambda = 10^{-3}$, $B^o_\lambda = 686.1$



(f) $DR = 1500$, $f_{p_e}(B)$ at $p_e = 10^{-6}$, $B^o_{p_e} = 37732.5$, and $f_\lambda(B)$ at $\lambda = 10^{-6}$, $B^o_\lambda = 37987.1$

Fig. 10 (Continued) — Stop-and-wait wasted-time functions versus blocklength for $B_I = 200$ and $B = 255$

The stationary-point equation is then

$$\frac{dP_e}{dB} = \frac{1 - P_e^m + P_e}{\left[B + \dfrac{mB^2}{DR(1 - P_e^m)}\right]\left[1 + (m - 1)P_e^m\right]}.$$

Under the assumption of Eq. (7) the stationary-point equation becomes

$$\gamma = \frac{2 - e^{-\gamma B} - (1 - e^{-\gamma B})^m}{e^{-\gamma B}\left\{B + \dfrac{mB^2}{DR[1 - (1 - e^{-\gamma B})^m]}\right\}\left[1 + (m - 1)(1 - e^{-\gamma B})^m\right]}. \tag{10}$$

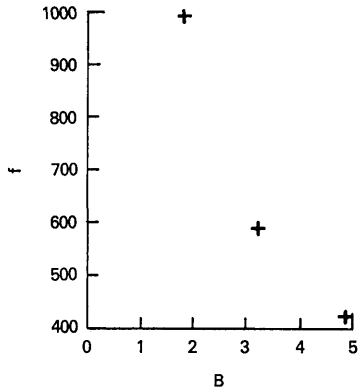
This nontrivial equation is only an implicit form of the stationary-point equation. Thus obtaining the solution with this scheme requires much more computation than with the basic scheme. Although

$$\frac{dP_e}{dB} \geqslant \frac{1 - P_e^m + P_e}{\left[B + \dfrac{mB^2}{DR(1 - P_e^m)}\right]\left[1 + (m - 1)P_e^m\right]}$$

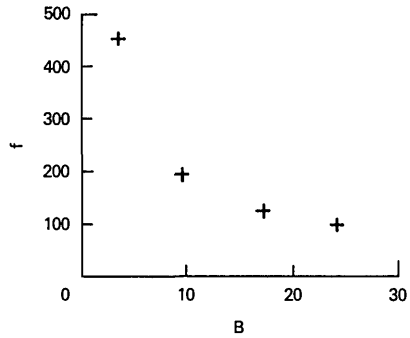
is a necessary and sufficient condition for $f(B)$ to be monotonically nondecreasing, note that

$$\frac{dP_e}{dB} \geqslant \frac{1 - P_e^m + P_e}{B\left[1 + (m - 1)P_e^m\right]} \geqslant \frac{1 - P_e^m + P_e}{\left[B + (mB^2/DR)\right]\left[1 + (m - 1)P_e^m\right]}$$

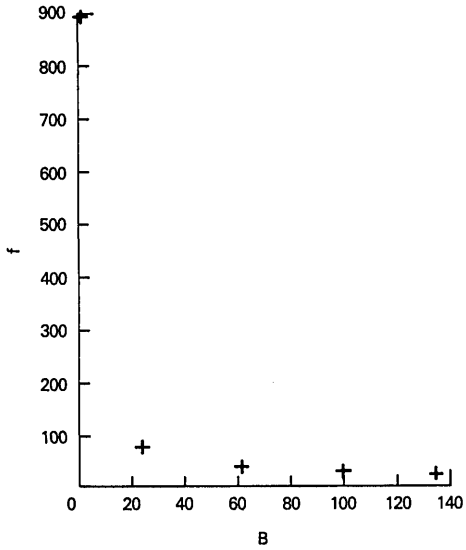
is a sufficient condition for $f(B)$'s monotonicity, which may be exploited in determining the global minimum. The function $f(B)$ from Eq. (10) is displayed in Fig. 11 for $DR = 1500$ and $\gamma = \lambda$ or $-\ln(1 - p_e)$. These points are computed using the triple of points (m_o, P_e, B) which are the simultaneous solutions to Eq. (4) and $P_e = 1 - e^{-\gamma B}$ (Fig. 4). In only one instance (Fig. 11d with $\gamma = 10^{-2.5}$) does the minimum of $f(B)$ not occur for the largest values of intersection points m_o or B obtained from Fig. 4 (for a given value of γ). However B_{min} in Fig. 11d is much larger than for any of the other examples. Again B_{min} appears highly sensitive to variations in the error rate γ.



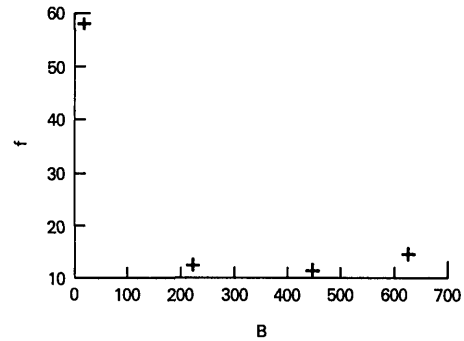
(a) $\gamma = 10^{-1}$



(b) $\gamma = 10^{-1.5}$



(c) $\gamma = 10^{-2}$



(d) $\gamma = 10^{-2.5}$

Fig. 11 — Stop-and-wait variation-1 wasted-time function versus blocklength for DR = 1500

Optimal Blocklength for Variation 2 of the Stop-and-Wait ARQ Scheme

For variation 2 of the stop-and-wait ARQ scheme the normalized-wasted-time expression to be minimized is

$$f(B) = (E - 1) + DR(MB)^{-1}E$$

$$= \frac{P_e}{1 - P_e} + \frac{DR}{MB(1 - P_e)},$$

where M is the number of blocks transmitted in sequence before the transmitter stops and waits for an acknowledgment. Differentiation of this expression with respect to B yields the stationary-point equation

$$\frac{dP_e}{dB} = \frac{DR(1 - P_e)}{B(MB + DR)}.$$

For $M = 1$, corresponding to the basic stop-and-wait scheme, we get the stationary-point equation for that scheme as expected. This stationary-point equation becomes $\gamma = DR/B(MB + DR)$, yielding

$$B^o = DR\left[\left(\frac{1}{4M^2} + \frac{1}{\gamma MDR}\right)^{1/2} - \frac{1}{2M}\right]. \tag{11}$$

As stated previously, the results for the basic stop-and-wait scheme are directly applicable here, with the substitution of DR/M for DR in the appropriate equations. This yields

$$f'(B) = \frac{\gamma B^2 + \gamma(DR/M)B - (DR/M)}{B^2 e^{-\gamma B}},$$

and B^o, given by Eq. (11), is the minimum by the argument used for the basic scheme.

Optimal Blocklength for the Go-Back-N Continuous-ARQ Scheme

The normalized wasted-time expression to be minimized for the go-back-N continuous-ARQ scheme is

$$f(B) = \frac{DR}{B}P_e + \left(1 + \frac{DR}{B}P_e\right)(E - 1)$$

$$= \frac{\left(1 + \frac{DR}{B}\right)P_e}{1 - P_e},$$

where $1 + (DR/B) = N$ is the number of blocks the transmitter goes back when a repeat-request is received. Note that $f(B) \to P_e(1 - P_e)^{-1} \to \infty$ as $B \to \infty$; thus large values of B are detrimental to throughput efficiency for this scheme.

The stationary-point equation is

$$\frac{dP_e}{dB} = \frac{DR(1 - P_e)}{B(B + DR)} P_e,$$

which is the same as for the basic stop-and-wait scheme except for the factor P_e. By use of Eq. (7) the stationary-point equation becomes

$$\gamma = \frac{DR(1 - e^{-\gamma B})}{B(B + DR)}.$$

For any block error distribution,

$$\frac{dP_e}{dB} \geqslant \frac{DR(1 - P_e)}{B(B + DR)} P_e$$

insures that $f(B)$ is monotonically nondecreasing. When this inequality is not satisfied for all B and practical considerations dictate the minimum size of B, the behavior of $f(B)$ should be examined in the region up to and including the last interval in which its derivative is negative. However this inequality is satisfied on the interval $[1, \infty]$ for the error case $P_e = 1 - e^{-\gamma B}$ considered here, since for all $B \geqslant 1$ and $\gamma > 0$

$$B^2 + BDR \geqslant \frac{DR(1 - e^{-\gamma B})}{\gamma},$$

because the equality is satisfied at $B = 0$ and

$$\frac{d}{dB}(B^2 + BDR) = 2B + DR \geqslant \left[\frac{DR(1 - e^{-\gamma B})}{\gamma}\right] = DRe^{-\gamma B}.$$

Thus $B = 1$ is the minimum point for $f(B)$. In application the smallest B allowed under practical considerations is the best choice for minimizing the expected wasted time.

Optimal Blocklength for the Variation of the Go-Back-N Continuous-ARQ Scheme

The variation of the basic go-back-N scheme and the modified variation are similar to the basic scheme and are treated together here. The variation (corresponding to η_{G1}) has the normalized expected wasted time

$$f_1(B) = (E - 1) + k(N - 1)P_e \ , \ k = 2,$$

$$= \frac{P_e}{1 - P_e} + 2\left(\frac{DR}{B}\right)P_e,$$

and the modified variation (corresponding to η'_{G1}) has

$$f_2(B) = \frac{P_e}{1 - P_e} + \frac{DR}{B} P_e.$$

It is sufficient to consider the behavior of

$$f(B) = P_e\left(\frac{1}{1 - P_e} + k\frac{DR}{B}\right).$$

The stationary-point equation is

$$\frac{dP_e}{dB} = \frac{kDR(1 - P_e)^2 P_e}{B[B + kDR(1 - P_e)^2]} ,$$

which yields the implicit-solution equation

$$\gamma = \frac{kDR(e^{-\gamma B})\ (1 - e^{-\gamma B})}{B\ [B + kDR(e^{-2\gamma B})]}.$$

Now $f(B)$ is monotonically nondecreasing as long as

$$\frac{dP_e}{dB} \geqslant \frac{kDR(1 - P_e)^2 P_e}{B\ [B + kDR\ (1 - P_e)^2]}.$$

This inequality is valid for the semi-infinite interval $[B^o, \infty)$, where

$$B^o = \left\{kDR\left[\frac{e^{-\gamma\sigma}}{\gamma} - e^{-2\gamma\sigma}\left(\sigma + \frac{1}{\gamma}\right)\right]\right\}^{1/2} ,$$

with σ satisfying

$$\gamma\sigma = \ell n(1 + 2\gamma\sigma). \tag{12}$$

Thus for this case it is sufficient to evaluate $f(B)$ on the interval $[1, B^o]$ to obtain the desired minimum. In general the complexity of the stationary-point equation may require the plotting of $f(B)$ directly or the use of an optimization algorithm. Figure 12 indicates the behavior of $f(B)$, where the minimizing value of B does not occur at the endpoints of the intervals under investigation. The solution for Eq. (12) is $\sigma_{p_e} = 1.2564247/\ell n(1 - p_e)$ and $\sigma_\lambda = 1.256231765/\lambda$ respectively. Table 2 lists the minimizing values of B for error rates λ and $p_e = 10^{-n}$, where $n = 1, ..., 6$. Only for $DR = 1500$ do we have $B_{min} \neq 1$ for all λ and p_e considered. As a test consider

$$C(B) = \frac{e^{\gamma B} - 1}{2\gamma B} + \frac{e^{2\gamma B}}{\gamma KDR} .$$

If $C(1) \geqslant 1$, then $B_{min} = 1$. If $C(1) < 1$ and $C(B) \geqslant 1$ for $B = B^o$, then $B_{min} \in (1, B^o]$ As before, B_{min} is highly sensitive to the error rate λ or p_e.

Optimal Blocklength for the Selective-Repeat Continuous-ARQ Scheme

In the selective-repeat continuous-ARQ scheme the normalized expected-wasted-time expression to be minimized is

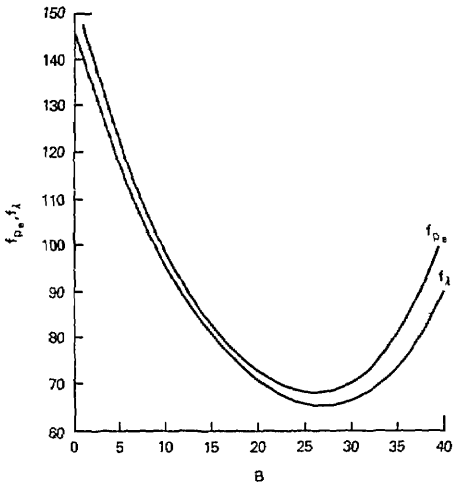
$$f(B) = E - 1 = \frac{P_e}{1 - P_e} .$$

Note that $f(B)$ is monotonically increasing in B, since P_e is monotonically increasing in B (that is, as the blocks get longer, the probability of a block error increases). Thus the smallest B allowed under practical considerations is the best choice for minimizing the expected wasted time.

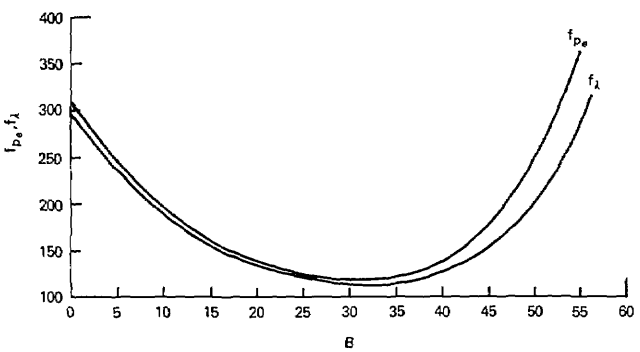
SUMMARY AND CONCLUSIONS

This report provides the data-communication design engineers with information to facilitate their choice of error control techniques or to assess the performance of their proposed ARQ or hybrid designs.

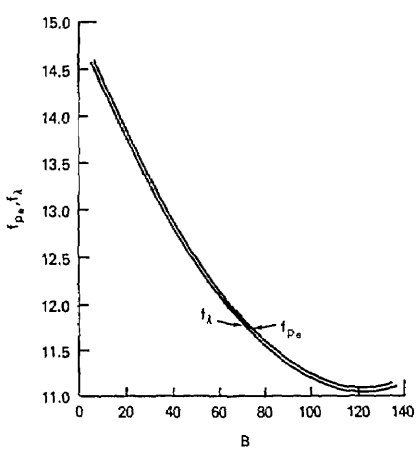
In this report the performance of various ARQ schemes used for error control in data-transmission systems was described. The basic schemes, such as stop-and-wait, go-back-N, and selective-repeat, were described from a heuristic as well as a theoretical viewpoint. Recently reported variations and a hybrid (FEC plus ARQ) were also discussed. Specifically, the throughput performance of these ARQ schemes under certain assumptions and their optimal blocklengths were sought.



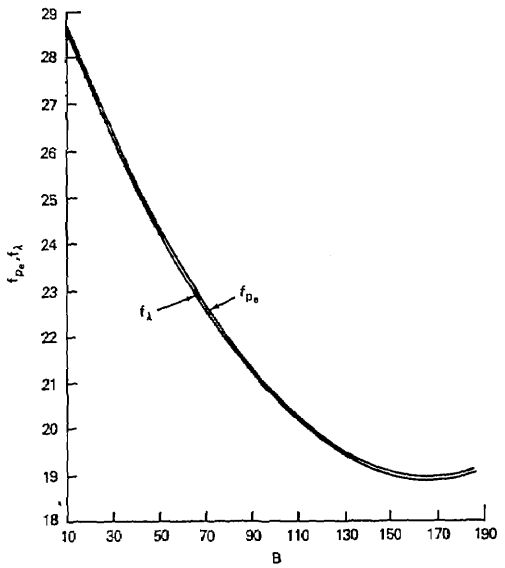
(a) $k = 1, \lambda = p_e = 10^{-1}$, where $f_{p_e}(B)$ for $B^o_{p_e} = 38$, $\sigma_{p_e} = 12$, and $f_\lambda(B)$ for $B^o_\lambda = 39$, $\sigma_\lambda = 13$



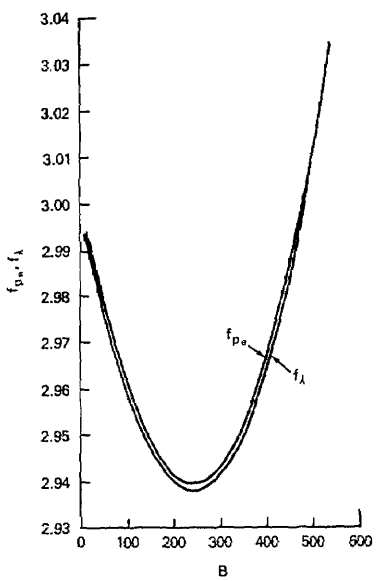
(b) $k = 2, \lambda = p_e = 10^{-1}$, where $f_{p_e}(B)$ for $B^o_{p_e} = 54$, $\sigma_{p_e} = 12$, and $f_\lambda(B)$ for $B^o_\lambda = 55$, $\sigma_\lambda = 13$



(c) $k = 1, \lambda = p_e = 10^{-2}$, $\sigma_{p_e} = \sigma_\lambda = 126$, where $f_{p_e}(B)$ for $B^o_{p_e} = 123$, and $f_\lambda(B)$ for $B^o_\lambda = 124$



(d) $k = 2, \lambda = p_e = 10^{-2}$, $\sigma_{p_e} = \sigma_\lambda = 126$, where $B^o_{p_e} = 174$, and $B^o_{p_e} = 175$



(e) $k = 2, \lambda = p_e = 10^{-3}$, $\sigma_{p_e} = \sigma_\lambda = 1256$, and $B^o_{p_e} = B^o_\lambda = 553$

Fig. 12 — Go-back-N-variation wasted-time function versus blocklength for $DR = 1500$

Table 2 — Optimal Blocklengths B for the Variation of the Go-Back-N Continuous-ARQ Scheme

λ, p_e	k	DR	σ_{p_e}	σ_λ	$B^o_{p_e}$	B^o_λ	$B_{p_e,\,\min}$	$B_{\lambda,\,\min}$	
10^{-1}	1	5	12	13	2.2	2.26	1	1	C(1)≥1
10^{-1}	1	1500	12	13	38.1	39.1	≈ 26.5	≈ 27.3	
10^{-1}	2	5	12	13	3.11	3.19	1	1	C(1)≥1
10^{-1}	2	1500	12	13	53.84	55.24	≈ 30.7	≈ 32	
10^{-2}	1	5	126	126	7.12	7.13	1	1	C(1)≥1
10^{-2}	1	1500	126	126	123.3	123.6	≈123.3	≈123.6	
10^{-2}	2	5	126	126	10.06	10.09	1	1	C(1)≥1
10^{-2}	2	1500	126	126	174.33	174.8	167.35	167.78	
10^{-3}	1	5	1256	1256	22.6	22.6	1	1	C(1)≥1
10^{-3}	1	1500	1256	1256	390.7	390.8	1	1	C(1)≥1
10^{-3}	2	5	1256	1256	31.9	31.9	1	1	C(1)≥1
10^{-3}	2	1500	1256	1256	552.5	552.7	≈245	≈248	
10^{-4}	1	5	12564	12562	71.3	71.4	1	1	C(1)≥1
10^{-4}	1	1500	12564	12562	1235.7	1235.8	1	1	C(1)≥1
10^{-4}	2	5	12564	12562	100.9	100.9	1	1	C(1)≥1
10^{-4}	2	1500	12564	12562	1747.6	1747.7	1	1	C(1)≥1
10^{-5}	1	5	125642	125623	225.8	225.6	1	1	C(1)≥1
10^{-5}	1	1500	125642	125623	3911.2	3908	1	1	C(1)≥1
10^{-5}	2	5	125642	125623	319.4	319.1	1	1	C(1)≥1
10^{-5}	2	1500	125642	125623	5531.2	5526.7	1	1	C(1)≥1
10^{-6}	1	5	1256425	1256232	708.8	713.5	1	1	C(1)≥1
10^{-6}	1	1500	1256425	1256232	12275.9	12358.2	1	1	C(1)≥1
10^{-6}	2	5	1256425	1256232	1002.3	1009	1	1	C(1)≥1
10^{-6}	2	1500	1256425	1256232	17360.7	17477.1	1	1	C(1)≥1

The results obtained are in the form of equations and algorithms for computing throughput efficiency and optimal blocklengths. The more detailed results are for independent-bit-error and Poisson-distributed block error channels.

In all cases the design engineer's choice is based on the common tradeoff between acceptable performance and implementation or complexity costs. The better performing schemes, such as selective-repeat ARQ, require additional storage and intelligence within the transmitter-receiver pair (not to mention separate feedback channel), which are not necessary in the lesser performing stop-and-wait ARQ scheme. But microprocessor and memory costs are dropping, so that the cost of additional control logic is becoming less dominant in the choice of schemes. In general the optimal blocklength for a specific ARQ scheme is sensitive to the error rate. Thus different acceptable blocklengths may be required, for several corresponding ranges of error rate. Adaptivity is also implied for some hybrid ARQ designs in which error-correction coding (FEC) is added at some predetermined level of error rate to improive the throughput performance at high error rates.

Other conclusions can be drawn from the data in this report. If the variations of implementation and complexity cost among the ARQ and hybrid schemes are marginal or of minor importance, then the selective-repeat ARQ is the best choice at low-to-moderate block error rates. Moreover performance falls off smoothly for high error rates ($P_e \geqslant 10^{-2}$). This is true for all block sizes. For small delays DR, the stop-and-wait, stop-and-wait variation-2, and go-back-N schemes yield less but comparable throughput efficiency. The difference becomes quite pronounced at moderate to high delays. From the example in which error-correction coding (FEC) was added, substantial improvement is possible over all the schemes here at moderate to high bit error rates ($10^{-3} \leqslant P_e$). However, these gains may be at the expense of only low to moderate performance when used at low bit error rates. The questions of interest then become: should FEC be switched on at some predetermined error rate or should it be applied at all error rates, and what improvements are possible for FEC when bit errors are not independent.

Under the assumption of Poisson block errors or independent bit errors we found that the go-back-N and selective-repeat ARQ schemes had optimal blocklengths of one bit (throughput efficiency improves as blocklength decreases) and that the stop-and-wait, stop-and-wait variation-2, and go-back-N-variation ARQ schemes have optimal throughput efficiency on their respective intervals $[1, B^o]$, where B^o is the solution to the corresponding throughput-efficiency stationary-point equation. However, for the go-back-N and selective-repeat schemes, practicality dictates some minimum blocklength to include block overhead information, such as redundancy bits for error detection or other block processing data. In addition the original assumption of all message blocks being the same length and containing no filler bits is thus justified for the go-back-N and selective-repeat cases. The unsuitability of the assumption for the other ARQ schemes, for which the optimal blocklengths may be rather large, will have to be determined from further investigation.

The throughput results of this report are in terms of throughput efficiency η for a fixed transmission rate R. In some applications the choice of transmission rates may be flexible. Thus it may be required to optimize overall throughput ηR with repect to R.

But because η is also a function of R, both explicitly and implicitly (via the optimal blocklength dependence on R), determining the optimal transmission rate and the optimal blocklength to maximize the throughput (in bits per second) is in general not trivial. It is one of several topics for further investigation.

We briefly recommend some other extensions of this report:

- A comparison of optimal performances of the various ARQ schemes. These performances could be composed for both fixed and optimal transmission rates.

- An evaluation of throughput performance, when various popular error-correcting codes (FEC) are used at high error rates to augment the various ARQ schemes. The evaluation should extend to dependent-bit errors as well. Likewise additional desirable information would include the error levels at which FEC is added, whether FEC should be applied for the entire range of error rates (not just switched on at high error rates), and the relationship between FEC and desired optimal blocklengths.

- A determination of performance degradation due to the deviation of an actual application from the preliminary assumptions of this report. For example the assumption of messages consisting of only an integral number of equal-length blocks is surely not realistic. Yet how much this assumption improves, or possibly degrades, the throughput performance is not known.

- An evaluation of the sensitivity of the optimal blocklengths for specific ARQ schemes to the error rate environment. In addition the plausibility of optimal blocklengths for several error-rate ranges should be investigated.

REFERENCES

1. R.J. Benice and A.H. Frey, Jr., "An Analysis of Retransmission Systems," IEEE Trans. Commun. Technol. **COM-12** (No. 4), 135-145 (Dec. 1964).

2. R.J. Benice and A.H. Frey, Jr., "Comparisons of Error Control Techniques," IEEE Trans. Commun. Technol. **COM-12** (No. 4), 146-154 (Dec. 1964).

3. E.R. Cacciamani and K.S. Kim, "Circumventing the Problem of Propagation Delay on Satellite Data Channels," Data Communications, July/Aug. 1975, pp. 19-24.

4. D.R. Doll, "Calculating Throughput on Full-Duplex Data-Link Controls," Data Communications, Jan./Feb. 1976, pp. 25-28.

5. W.G. McGruther, "Thruput of High Speed Data Transmission Systems Using Block Retransmission Error Control Schemes Over Voicebandwidth Channels," Int'l Commun. Conf. Rec., 1972, pp. 15-19 through 15-24.

6. A.R.K. Sastry, "Improving Automatic Repeat-Request (ARQ) Performance on Satellite Channels Under High Error Rate Conditions," IEEE Trans. Commun. COM-23 (No. 4), 436-439 (Apr. 1975).

7. A.J. Viterbi, "Error Control for Data Communication," Computer Communication Review, Jan. 1976, pp. 27-37.

8. A.R.K. Sastry, "Performance of Hybrid Error Control Schemes on Satellite Channels," IEEE Trans. Commun. COM-23 (No. 7), 689-694 (July 1975).

9. A.R.K. Sastry and L.N. Kanal, "Hybrid Error Control Using Retransmission and Generalized Burst-Trapping Codes," IEEE Trans. Commun. COM-24 (No. 4), 385-393 (Apr. 1976).

10. W.W. Peterson and E.J. Weldon, Jr., *Error-Correcting Codes*, 2nd edition, Cambridge, Mass., MIT Press, 1972.

11. A.J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," IEEE Trans. Inform. Theory IT-13, 260-269 (Apr. 1967).

12. W.W. Chu, "Optimal Message Block Size for Computer Communications with Error Detection and Retransmission Strategies," IEEE Trans. Commun. COM-22 (No. 10), 1516-1525 (Oct. 1974).

13. A. Shaheen and J.C. Majithia, "Comparison of Overhead Factors in Message-Switched and Packet-Switched Networks," Electron. Lett. 12 (No. 1), 20-21 (8 Jan. 1976).