Transactions Papers

An Adaptive Error Control System Using Hybrid ARQ Schemes

Maan A. Kousa and Mushfiqur Rahman, *Member, IEEE*

***Abstract*— Reliability and throughput are two important aspects of data communication systems. Various error control strategies such as FEC, ARQ, and hybrid ARQ schemes can be used to enhance system performance. For time-varying channels, the optimum solution is to adaptively match the rate of the error-correcting code to the prevailing channel conditions. A simple and efficient system utilizing the well-known class of Hamming codes in a cascaded manner is proposed to provide high throughput over a wide range of channel bit-error probability. Comparisons with other adaptive schemes indicate that the proposed system is superior from the point of view of throughput, while still providing the same order of reliability as an ARQ system. The main feature of this system is that the receiver uses the same decoder for decoding the received information after each transmission while the error-correcting capability of the code increases. As a result, the system is kept to the minimum complexity and the system performance is improved.**

I. Introduction

Basically, there are two fundamental techniques for error control; automatic repeat request (ARQ) and forward-error control (FEC) [1], [2]. The former employs pure error detection, while the latter employs error correction only. To evaluate their performance, two important parameters are defined: throughput and reliability.

Throughput is defined as the ratio of the average number of information bits successfully accepted by the receiver per unit time, to the total number of bits that could be transmitted per unit time [2].

Reliability is a measure of the correctness of the decoded data.

An FEC system provides a constant throughput efficiency, set by the code rate, regardless of the channel conditions. However, as the decoded word must be delivered to the user regardless of whether it is correct or not, the system reliability falls as the channel degrades. On the contrary, an ARQ system, incorporating powerful error-detection codes, can provide high system reliability, reasonably independent of the channel quality, but the throughput depends strongly on the number of requested retransmissions, i.e., on channel quality, and therefore falls rapidly with increasing channel error rate. A proposed system which can have the advantages of both and disadvantages of neither is termed "hybrid ARQ." Such schemes can be broadly classified as follows [3].

1) Type-I hybrid ARQ scheme.
2) Type-II hybrid ARQ scheme.

Type-I ARQ scheme employs one code for error correction (FEC) and another code for error-detection (ARQ). At the receiver, the decoder first attempts to correct any error in the received block, then the decoded block is tested for error detection. If no errors are detected the message is delivered to the user, otherwise the receiver requests a retransmission of the same block. As error correction is always based on the same code, this scheme may be referred to as "fixed-rate hybrid ARQ."

Fixing the rate of the error-correcting code has drawbacks for channels having nonstationary behavior. If the channel is fairly good, the redundancy in the system may be more than the optimum value. On the other hand, if the channel is bad, more errors are expected to occur than those which can be handled by the capability of the error-correcting code. Consequently, too many retransmissions are requested, and the throughput falls down. Thus, in order to achieve optimum performance, the rate of the error-correcting code should be matched to the prevailing channel conditions. This forms the basis of Type II hybrid ARQ [3], or adaptive ARQ schemes. The first Type-II hybrid ARQ using a parity retransmission strategy was proposed by Metzner [4]. Among all the Type-II hybrid ARQ schemes that have been reported, the scheme proposed by Lin and Yu [3], [5] is the most analyzed.

Krishna and Morgera [6] were able to generalize the system proposed by Lin and Yu by discovering one class of linear codes, referred to as KM codes, after their names. The overall system is referred to as the "generalized hybrid ARQ scheme (GH-ARQ)."

GH-ARQ Scheme: The GH-ARQ scheme also uses two codes; one is a high rate (n, k) code C_0 which is designed for error detection only, and the second is the code C_1 which is used adaptively for error correction. The code C_1 is an (mn, n) KM code having distance d, and the integer m is referred to

Paper approved by the Editor for Coding Theory and Applications of the IEEE Communications Society. Manuscript received December 14, 1988; revised May 28, 1990.

M. A. Kousa is with the Department of Electrical Engineering, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia.

M. Rahman is with Telecom Research Laboratories, Clayton, Victoria, 3168, Australia.

IEEE Log Number 9100337.

0090-6778/91/0700-1049$01.00 © 1991 IEEE

as the depth of the code. This code is designed to possess the following property: the code

$$C_1^{(i)}, \qquad 2 \le i \le m$$

is an (in, n) invertible error-correcting code of distance d_i where

$$d_i < d_j \qquad \text{for } 2 \le i < j \le m.$$

A code is said to be invertible if, knowing only the parity bits of a codeword, the associated information bits can be uniquely determined. Note that the code $C_1^{(m)}$ is the code C_1.

The receiver performs error correction on the basis of the codes $C_1^{(2)}, C_1^{(3)}, \cdots, C_1^{(m-1)}, C_1, C_1, \cdots$, having distances $d_2, d_3, \cdots, d_{m-1}, d, d, \cdots$. Thus, with each retransmission, a code having a larger distance, i.e., a greater error-correcting capability, is used until the code C_1 is reached. The reader is referred to [6] for a detailed description and analysis of the GH-ARQ system.

The generalized scheme achieves adaptation by severely penalizing the code rate. The code rate is reduced from R to $R/2$ in only one step of adaptation. The newly discovered KM codes, due to their nature of construction, cannot provide smoother adaptation.

This paper introduces the use of the well-known Hamming codes in a *cascaded fashion* to perform adaptation. The proposed system possesses two main advantages.

1) The decoding strategy of the receiver is able to use the essential simplicity of decoding Hamming codes.

2) In general, the proposed system can employ any Hamming code, thus reduction in rate required for adaptation is more gradual.

II. Cascaded Hamming Codes

Consider a sequence of bits of length n. We refer to those n bits as being of zeroth level of cascading, in the sense that they are as yet uncoded for error correction. These bits may now go through one or more stages of encoding as discussed below.

A. Encoder

A particular (N, K) Hamming code is adopted for error correction. The n information bits are divided into groups of K bits. Each set of K bits is then mapped into N bits by means of an (N, K) Hamming encoder. The total number of bits at this stage is $(n/K)N$. This stage is referred to as the first level of cascading. The second level of cascading is formed by sorting the $(n/K)N$ available bits into groups of K bits again. These bits should constitute $(n/K^2)N$ such groups. Each group is again encoded into N bits by the same encoder. At the end of this stage, there should be $(n/K^2)N^2 = n(N/K)^2$ bits. Continuing this process, any level of cascading can be reached. Fig. 1 illustrates this process for two levels of cascading. To obtain an integer number of subblocks after each division by K at any stage, the initial sequence should be of length K^m or any integer multiple of K^m, where m is the maximum intended level of cascading.

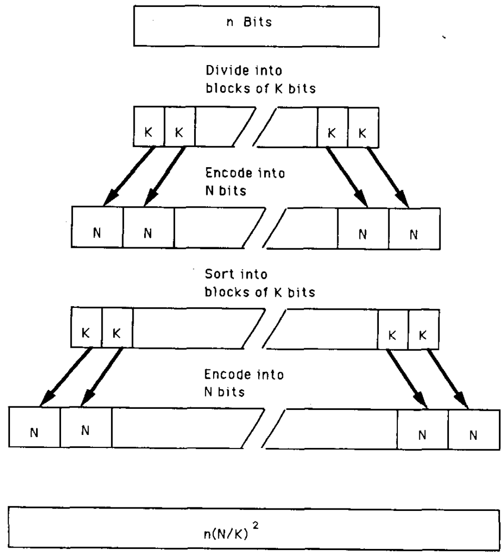


Fig. 1. Encoding for two levels of cascading.

B. Decoder

Now suppose that the receiver receives a sequence of bits which are known to have been encoded to the mth level. Let the channel bit error rate be ε. The receiver considers the received bits as blocks of N bits, and computes the syndrome of each block to perform error correction. After completing the first-order check, the receiver discards the check bits leaving the K information, but now with a different bit error probability, say ε'. It can be shown that the bit error probability after decoding is the same for all N positions in the block and discarding the check bits does not alter the error probability per position for the information bits [7]. The remaining bits correspond to the $(m-1)$th level of cascading. The receiver again divides these bits into blocks of N bits, checks for errors in each block, and discards the check bits. In this way, another level of cascading is decoded, where each decoded bit has a bit error probability ε''. This process is continued until the receiver decodes the original n information bits. Fig. 2 illustrates the decoding process for two levels of cascading.

C. Decoded Bit Error Probability

When a received word is decoded incorrectly, it does not mean that every decoded bit is erroneous. A decoded codeword is wrong whenever at least one bit is wrong. A useful quantity to calculate is the bit error rate after decoding ε', when the channel bit error rate is ε. For an (N, K) Hamming code, it is shown in Appendix A that

$$\varepsilon' = \frac{1}{N} \sum_{i=0}^{N} \left[(i+1) \binom{N}{i} - A_i - 2A_{i-1}(N-i+1) \right] \cdot \varepsilon^i (1-\varepsilon)^{N-i} \qquad (1)$$

where A_i is the number of codewords of weight i. Fig. 3 shows ε' versus ε for the first three Hamming codes (7,4), (15,11),

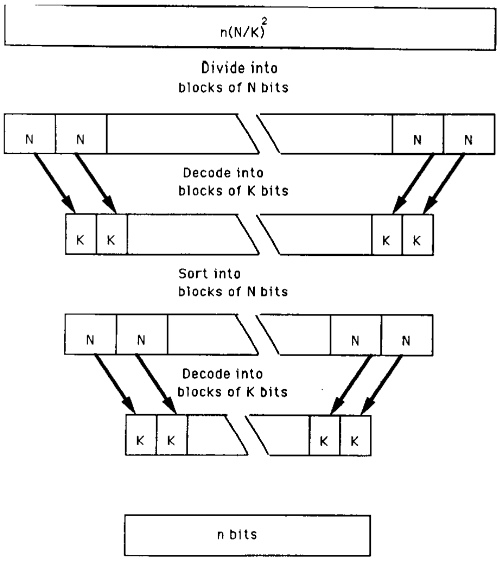


Fig. 2. Decoding of two levels of cascading.

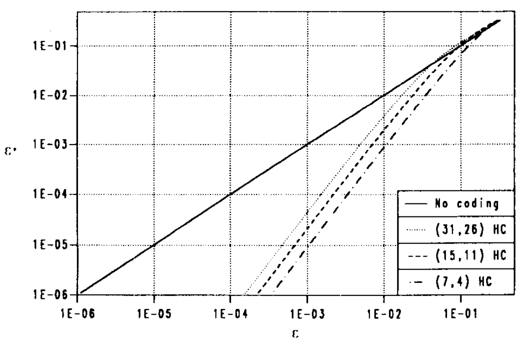


Fig. 3. Decoded bit error rate compared to channel bit error rate for three Hamming codes.

and (31, 26), which shows a considerable reduction in the bit error rate due to coding.

D. Interleaving

We claim that if the selection of K bits for a subblock is done carefully at each stage, then any further cascading leads to a higher error-correction capability of the system.

The simplest choice for the second level of cascading would be to take adjacent K bits at a time as a block. This is guaranteed not to work because the second-level checking (decoding) can never give correct decision unless the received sequence is error-free or contains exactly one error per block; in both such cases, there is no need for further cascading at all. On the other hand, if there are errors left in this group of bits after the first-order check (which is always the case when the channel introduces more than one error per block), there will be at least three errors per block. Most of the errors left will be confined to the information section, and the second-order check can never correct them. The solution to this problem is to spread or *interleave* the first-order checked bits of the same block. The process is equivalent to randomizing the errors after any decoding process. At the transmitter side, to perform a second level of encoding, each K information bit should be chosen from K different blocks of the first level of coding. This processing is necessary in order to keep the information bits in the same order of significance.

If the cascaded encoding process is organized by such suitable interleaving, then the bit errors in each block after deinterleaving are guaranteed to be statistically independent, because each decoded bit comes from a different block. Therefore, the decoded bit error probability with two levels of decoding is given by

$$\varepsilon'' = \frac{1}{N}\sum_{i=1}^{N}\left[(i+1)\binom{N}{i} - A_i - 2A_{i-1}(N-i+1)\right] \cdot \varepsilon'^{i}(1-\varepsilon')^{N-i} \qquad (2)$$

where ε' is given by (1). Thus, (1) can be used recursively, where ε is the bit error probability after decoding from the previous level. The required interleaving for three levels of cascading is established in Fig. 4 for the (7, 4) Hamming code. At the receiver side, a corresponding deinterleaving operation must be done.

E. Probability of Correct Decoding of a Cascaded System

If K bits are to be transmitted through a BSC having a bit error rate ε, then the probability of receiving them correctly is equal to

$$(1-\varepsilon)^K. \qquad (3)$$

If, on the other hand, an (N, K) Hamming code is employed, then the probability becomes

$$(1-\varepsilon)^N + N\varepsilon(1-\varepsilon)^{N-1}. \qquad (4)$$

Now suppose that we use two levels of cascading. At the receiver, after the first decoding step, the bit error rate will reduce to ε'. Due to interleaving, we obtain N such independent bits. The probability of correct decoding of a block after the second decoding step is then given as

$$(1-\varepsilon')^N + N\varepsilon'(1-\varepsilon')^{N-1}. \qquad (5)$$

If three levels of cascading are in use, then the probability of correct decoding of one block of N bits is given by

$$(1-\varepsilon'')^N + N\varepsilon''(1-\varepsilon'')^{N-1} \qquad (6)$$

where ε'' is obtained from (2). This technique applies to any level of cascading under the condition of perfect interleaving.

The previous discussion takes one block of N bits into consideration. In principle, for any level of cascading, if one starts with n bits, then just before the final step of decoding there should be n/K blocks of N bits to be decoded.

Let $p_{c,i}$ denote the probability of correct decoding of the superblock (the block of n/K subblocks), when the bits are encoded for level i.

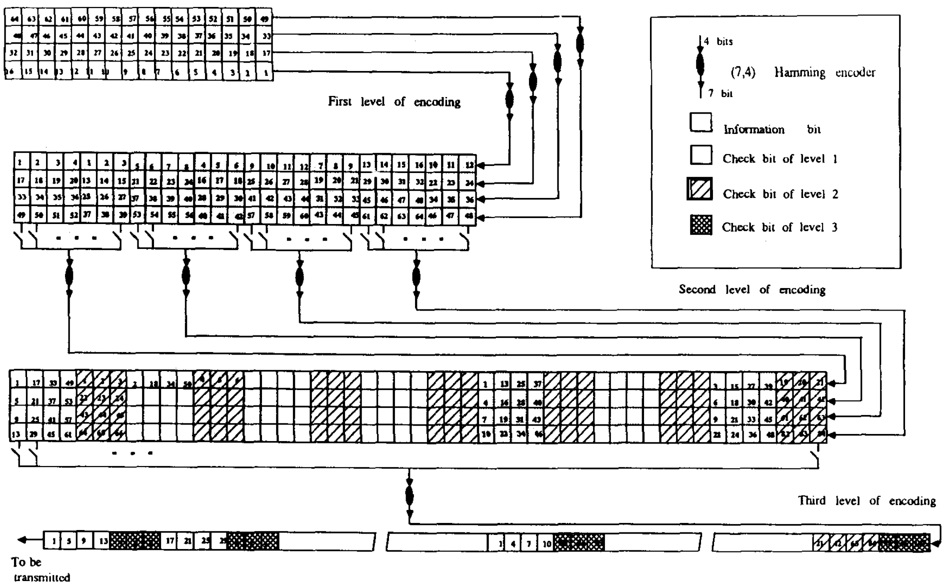


Fig. 4. Interleaving for three levels of cascading using the (7, 4) Hamming code.

Then, for $m = 1$

$$p_{c,1} = \left[(1-\varepsilon)^N + N\varepsilon(1-\varepsilon)^{N-1}\right]^{n/K}. \quad (7)$$

This follows from the fact that all n/K subblocks are completely independent. Unfortunately, this is not the case for $m > 1$ because although all N bits in one block are independent, bits in different blocks are dependent on each other due to past decoding steps. The probability of correct decoding of the superblock cannot be obtained simply by multiplying the individual probabilities of subblocks. One solution to this problem is to use the following union bound.

If

$p_c \equiv$ probability of correct decoding of a subblock
$P_C \equiv$ probability of correct decoding of the superblock
$p_e = 1 - p_c \equiv$ probability of wrong decoding of a subblock

then

$P_E \equiv$ probability of wrong decision of the superblock

$$\leq p_e + p_e \cdots + p_e = (n/K)p_e$$

and $P_C \geq 1 - P_E$.

Applying this bound to the case $m = 2$ yields

$$p_{c,2} \geq 1 - \frac{n}{K}\left\{1 - \left[(1-\varepsilon')^N + N\varepsilon'(1-\varepsilon')^{N-1}\right]\right\}. \quad (8)$$

The tightness of the union bound depends on the value of p_e and the number of subblocks n/K. In so far as this investigation is concerned, the union bound was found to be satisfactory for most cases. It should be emphasized that the value of P_C obtained by the union bound as above will be a conservative estimate and the actual value will always be superior.

III. Adaptive ARQ Scheme Based on Cascaded Hamming Codes

If the cascaded system is to be employed in a hybrid ARQ scheme, the initial bit sequence (n) should have some redundant bits for error detection based on an (n, k) error-detection code C_0. This is the only requirement for a message in its first transmission. When the receiver detects the presence of errors in a received word, it saves the erroneous word in a buffer, and at the same time requests a retransmission. The retransmission consists of a superblock of parity-check-bit subblocks, which is based on the original message and a Hamming error-correcting code at the first level of encoding. When this superblock of parity-check bits is received, it is used to correct the errors in the erroneous word stored in the receiver buffer. The decoded block is checked for errors by the error-detecting code. If error correction is not successful, the receiver stores the first-level parity-check bits also and requests a second retransmission. The second retransmission is another superblock of parity-check bit subblocks based on the original message, the first-level check bits already stored at the transmitter, and the same error-correcting code. Appropriate interleaving before encoding and deinterleaving after decoding are assumed. When this block of second-level check bits is received, it is again used to correct the erroneous message stored at the receiver. This process is repeated if necessary until the mth level parity-check bits are transmitted. If NAK is still sent back, the next transmission should be the initial information sequence again. The receiver will automatically replace the old information bits by the new ones and try to decode the message with the help of all parity checks already available. If an NAK is again received, the next transmission is the same as the first retransmission, containing the first level of parity checks, and so on. Doing this, the sequence of retransmissions for successive NAK's will be $I, P_1, P_2, \cdots, P_m, I, P_1$ where I is the original n-bit message and P_i is the ith level parity subblock. After the $(2m)$th retransmission, the whole set will have been renewed. With the reception of parity bits at the ith level, $i = 1, 2, \cdots m$, the decoder is required to rearrange the information sequence and the parity bits up to the $(i-1)$th level, in order to be consistent with the interleaving operation by the encoder.

The cascaded coding scheme described here has much in common with the notion of multidimensional product codes. In general, $IP_1P_2 \cdot \cdot P_m$ would represent a codeword in an m-dimensional product code, where P_i represents parity bits on $IP_1P_2 \cdot \cdot P_{i-1}$. However, the number of these parity subblocks to be transmitted depends on transmission errors. Under the best circumstances, none are transmitted and only I (which includes the error-detecting code) is transmitted. Thus, the basic difference between the proposed coding scheme and a conventional product code lies in the variability of the dimension of the product code, ranging from zero (no FEC coding) to some value m.

It should be noted that this system transmits blocks of different lengths. This may introduce some complexity increase relative to transmission of equal block lengths.

A. Reliability

Reliability, as defined earlier, is a measure of the correctness of the received data. Let $P(E)$ denote the probability that the received word is accepted and it is in error. To have a highly reliable system, the error-detection code incorporated should be properly chosen to make $P(E)$ very small. In any ARQ system, $P(E)$ is given by [2]

$$P(E) = \frac{P_{\text{ud}}}{P_{\text{ud}} + P_f}$$

where

$P_{\text{ud}} \equiv$ probability of an undetected error in the received/decoded n-bit word error pattern

$P_f \equiv$ probability that the received/decoded n-bit word is error free.

To make $P(E)$ very small, P_{ud} should be very small compared to P_f. A simple, yet tight, bound on P_{ud} which is obeyed by many classes of codes is given as [8]

$$P_{\text{ud}} \le 2^{-(n-k)}. \tag{9}$$

This bound is a simplified version of a bound by Korzhik [9]. Some classes of codes satisfy the even tighter bound [10]

$$P_{\text{ud}} \le 2^{-(n-k)}[1 + (1-2\varepsilon)^n - 2(1-\varepsilon)^n]. \tag{10}$$

If a code satisfying one of the above bounds is used for error detection, the probability of an undetected error can be made very small by using a moderate number of parity bits, say 20.

For the proposed cascaded system, requiring an initial stream of K^m bits or any multiple of this length puts a constraint on the length of the (n,k) error-detecting code. To satisfy this constraint, an (aK^m, k) code must exist a being an integer. This condition can be satisfied almost always, and here is the justification: for most practical channels two levels of cascading is adequate, and further cascading does not result in any significant improvement. If the codeword length of the chosen (N, K) Hamming code is not too long, K^m will be small compared with the length of an efficient error-detecting code. To satisfy the above condition, take an integer multiple of K^m such that an (aK^m, k) error-detecting code can be constructed. Due to the various families of already existing error-detecting codes, this condition can be sometimes satisfied, but not for the majority of cases. For these cases, the condition can be satisfied by making use of the theory of extended or shortened codes [11]. An error-detection code of a suitable length can be constructed as follows: Choose a such that aK^m is nearest to n, the length of an error-detecting code C_0. One of following two cases arises.

1) $aK^m < n$: For this situation, shorten the code C_0 by omitting $(n - aK^m)$ of its message digits. This way, we modify the code (n,k) to the code $(aK^m, k - n + aK^m)$.
2) $aK^m > n$: For this case, we extend the code C_0 by annexing $(aK^m - n)$ additional check digits and the modified code will be an (aK^m, k) code.

In the first case, no general statement can be made regarding the capability of the error-detecting code. If the number of deleted message bits is small, P_{ud} of the shortened code may be expected to be of the same order as the original code. However, an investigation of the P_{ud} of the shortened code will have to be made before its selection as the error-detecting code. In the second case, the error-detection capability is enhanced, and the extended code is guaranteed to be at least as good as the original code. It was possible to modify the code at the cost of reducing its rate. If n is sufficiently large, which is usually the case, and if aK^m is not that far from n, the reduction in the rate is negligible.

Here are some possible choices of the error-detection code for two levels of cascading $(m = 2)$:

1) $(7,4)$ Hamming code, $a = 128$. All extended primitive BCH codes of lengths 2^ℓ, $\ell \ge 4$, are suitable. Two good examples are the distance-8 (2048, 2014) and the distance-6 (2048, 2025) BCH codes. Both codes satisfy the bound given by (10) [11]. $P(E)$ is kept below 5.8×10^{-11} and 1.2×10^{-7} for the two codes, respectively. Keeping in mind that this is the case of the worst possible channels ($\varepsilon = 0.5$), high system reliability can be achieved.
2) $(15,11)$ Hamming code, $a = 17$. The suitable length of the code is 2057. Any one of the BCH codes introduced in the last paragraph can be used by annexing nine more check bits. $P(E)$ will be even smaller than the above calculated figures as a result of adding more check bits.
3) $(31,26)$ Hamming code, $a = 6$. Then n should be 4056. Distance-6 (4096, 4072) BCH code can be shortened to the (4056, 4032) code by deleting 40 message bits. The upper bound of $P(E)$ is expected to be of the order of $2^{-24} = 6 \times 10^{-10}$. The reduction in the rate of the code is negligible.

It should be emphasized that the overall system reliability depends on the undetected error probabilities of both the cascaded coding scheme and the embedded error-detection code. While both contribute to the overall reliability, it is the error-detecting code which has the primary role of the ensuring acceptable reliability. Therefore, a proper choice of this code is crucial for the success of the proposed hybrid ARQ scheme.

B. Throughput Analysis

We assume that the forward channel is a random error channel with bit error rate ϵ, and that the feedback channel is noiseless. Since selective repeat (SR) ARQ is the most efficient ARQ scheme, we consider the throughput of cascaded systems in the SR mode. The throughput of such a scheme depends on the receiver buffer size and, in this regard, we restrict our attention to the infinite buffer case. In order to calculate the throughput, we determine the average number of bits needed to be transmitted before k information bits are successfully accepted by the receiver. If we denote this number by T, then the throughput η is

$$\eta = \frac{k}{T}. \tag{11}$$

Upon the ith retransmission, $i = 0, 1, \cdots$ (where $i = 0$ refers to the first transmission), let $E_{c,i}$ be the event that the receiver recovers the block correctly; $E_{d,i}$ will be the event that the receiver detects the presence of errors, and requests the next retransmission; $E_{e,i}$ will be the event that the receiver cannot detect the presence of errors. Clearly,

$$\Pr(E_{c,i}) + \Pr(E_{d,i}) + \Pr(E_{e,i}) = 1, \quad i = 0, 1, 2, \cdots. \tag{12}$$

We will assume that $\Pr(E_{e,i}) = 0$. Then, we can write

$$\begin{aligned} T = {} & [n(0)] \Pr(E_{c,0}) + [n(0) + n(1)] \Pr(E_{d,0}E_{c,1}) + \cdots \\ & + [n(0) + n(1) + \cdots n(i)] \\ & \cdot \Pr(E_{d,0}E_{d,1} \cdots E_{d,i-1}E_{c,i}) + \cdots \end{aligned} \tag{13}$$

where $n(i)$, $i = 0, 1, \cdots$ is the length, in bits, of the ith retransmission.

This expression involves the probability of joint events. The probability of the event $E_{c,i}$, for example, depends on the probability of all the events $E_{d,0}E_{d,1} \cdots E_{d,i-1}$. This form of unlimited dependency on past events is extremely difficult to analyze. We, therefore, adopt the following inferior system approach for further analysis.

C. Definitions of Inferior System

The system for analysis is defined in the following way. Up to the mth retransmission, the system is exactly the same as we have defined before. If an $(m+1)$th retransmission is still required, the receiver discards all bits available to it, and sends a NAK. Due to the reception of this message, the transmitter will send the complete set already stored in the buffer. This set contains the information data plus the parity-check bits of all m levels. Subsequent retransmissions also constitute the entire message and parity bits up to the mth level. The sequence of retransmissions will be $I, P_1, P_2, \cdots, P_m$, $(I + P_1 + P_2 + \cdots + P_m)$, $(I + P_1 + P_2 + \cdots + P_m), \cdots$, assuming successive NAK's.

It is clear that this system is inferior to the original system. In simple words, after the mth retransmission the inferior system renews the whole set without performing any intermediate decodings, whereas the original system renews the data part by part, and performs decoding at each stage. Obviously, we need not renew the whole set to have a successful decoding. The ease of analysis of the inferior system is evident. The most tedious calculation involved in this case will be that of the evaluation of $\Pr(E_{c,m})$ and this event depends on the previous m events only. The throughput of the inferior system is shown in Appendix B.

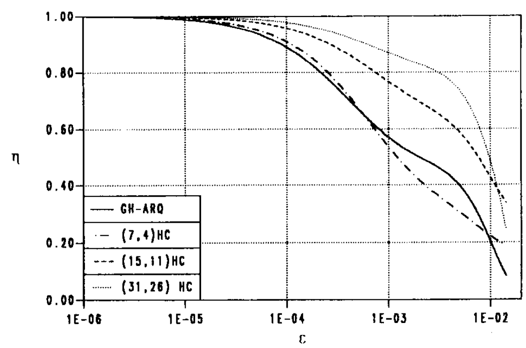


Fig. 5. Throughput of the GH-ARQ system compared with the cascaded system for three Hamming codes.

D. Comparison Between the Proposed System and the GH-ARQ System

Comparison is made with the GH-ARQ system because of its superiority over other ARQ schemes. We will make the comparison up to two levels of adaptation, i.e., two other retransmissions (of parity bits) will be used to enhance a first transmission, before the transmission of data is repeated again. In our notation, this corresponds to $m = 2$. For the notation used by Krishna and Morgera, it corresponds to $m = 3$. For a valid comparison, the error-detection codes used in the two systems should be of the same length and rate. The rate of the error-detecting code, being close to unity, is ignored in the derivation. The length of the error-correction code used imposes a constraint on the length of the error-detecting code, which makes it not possible to assume the same length of the latter, but the lengths are chosen to be nearly the same in all cases so that the comparisons are meaningful. The chosen lengths n satisfying the constraints are 1344, 1331, and 1352 b for the $(7,4)$, $(15,11)$, and $(31,26)$ Hamming codes, respectively. If the $(15,5,5)$ KM code is to be used in the GH-ARQ system, a convenient length of the error-detection code is 1340 b, which is close to the length chosen for the proposed scheme.

Fig. 5. compares the GH-ARQ system to the proposed system for three Hamming codes. It is evident that the proposed system allows higher throughput than the other system, especially for $(15,11)$ and $(31,26)$ codes. It should be remembered

that the comparison is being made not on the basis of the actual throughput, which is very difficult to evaluate in both cases, but on the lower bounds on the throughput. As the derivations of the throughput for each system are subject to different kinds of bounding arguments, one cannot guarantee that the bounds on the two systems are of the same degree of tightness. However, the throughput curves for the (31, 26) and (15, 11) codes lie so far above the curve for the GH-ARQ system that there is every reason to believe that the system proposed here is superior to the GH-ARQ system from the point of view of throughput.

IV. Conclusion

In a GH-ARQ system, whenever a retransmission is required the transmitter sends a block of parity checks *of the same length* as the block of message bits. It follows that the first retransmission causes the rate of error correction code to drop to 1/2. The second retransmission makes the rate 1/3, and so on. This is considered a drastic drop in the rate and a request for a retransmission may not justify such a drastic drop in the rate. On the other hand, after the first retransmission in the proposed system, the rate drops to (K/N), the rate of Hamming code in use. The second retransmission reduces the rate to $(K/N)^2$, and so on. If the moderate rate (31, 26) Hamming code is used, three levels of cascading can be reached before reducing the rate to 1/2. From our previous observations, we can say that even two steps of adaptation is adequate to cover the worst practical interval of ε. The superiority of the proposed system lies in its ability to adapt its rate much more smoothly compared to the GH-ARQ system.

The above comments are valid under the conditions stated in the derivation, i.e., SR mode and infinite buffer size. For these conditions, a gradual change does not introduce any further delay over a sharp one. For other conditions, such as different retransmission model and/or finite buffer size, the throughput comparison given here is not valid and requires a detailed analysis.

There is one more advantage which the proposed system can provide. Since there are many choices of Hamming codes with different rates, an optimum one may be chosen to match a certain channel to provide the best adaptation. Also, decoders for Hamming codes are trivially simple.

On the other hand, the proposed system introduces some complexity to the transmitter and receiver. It needs a circuitry for the interleaver at the transmitter and deinterleaver at the receiver. These two blocks are not required for the GH-ARQ system. In view of the advances in solid-state technology which are taking place, this may not be a severe drawback.

Another drawback arises from observing that the transmissions in the proposed system are of unequal blocks lengths as compared to equal block lengths in a GH-ARQ system. Equal block length transmissions may be more convenient and may require less overhead control bits. In addition, a GH-ARQ system has more flexibility in the choice of the error-detection code.

Appendix A
Derivation of a General Expression For the Decoded Bit Probability of Any (N,K) Hamming Code

The approach adopted here is to find the average number of errors in the decoded codeword subject to i errors in the received word, which is denoted by x_i. As all N bits of the decoded word have the same probability to be in error, the bit error rate can be found by dividing x_i by N. ε' is then defined as

$$\varepsilon' = \frac{1}{N}\sum_{i=2}^{N} x_i P_i \tag{A-1}$$

where

$x_i \equiv$ the number of errors in the decoded word due to i error patterns

$P_i \equiv$ probability of exactly i errors in a block of N bits, $= \varepsilon^i(1-\varepsilon)^{N-i}$.

The summation starts from $i = 2$ because $x_0 = x_1 = 0$ for single error-correcting codes.

Let A_i represent the number of codewords of weight i. If the received vector contains i errors, then one of the following three cases occur.

i) It is the same as one of the A_i codewords of weight i; in this case, no additional errors are introduced,

ii) It is at a distance 1 from one of the A_{i-1} codewords of weight $i-1$; there are $(N-i+1)A_{i-1}$ such received vectors and the decoded vector contains $(i-1)$ errors,

iii) It is at a distance 1 from one of the A_{i+1} codewords of weight $i+1$; there are $(i+1)A_{i+1}$ such received vectors and the decoded vector contains $(i+1)$ errors. Thus, we get

$$x_i = iA_i + (i-1)(N-i+1)A_{i-1} + (i+1)^2 A_{i+1}. \tag{A-2}$$

Recalling that the weight distribution of Hamming codes satisfy the recurrence [12] $A_0 = 1$, $A_1 = 0$,

$$(i+1)A_{i+1} + A_i + (N-i+1)A_{i-1} = \binom{N}{i} \qquad i = 1, 2, \cdots, N-1. \tag{A-3}$$

We can simplify (A-2) as

$$x_i = (i+1)\binom{N}{i} - A_i - 2A_{i-1}(N-i+1). \tag{A-4}$$

Substitution of (A-4) in (A-1) yields the required bit error probability ε' after decoding

$$\varepsilon' = \frac{1}{N}\sum_{i=2}^{N}\left[(i+1)\binom{N}{i} - A_i - 2A_{i-1}(N-i+1)\right] \cdot \varepsilon^i(1-\varepsilon)^{N-i}. \tag{A-5}$$

Appendix B
Throughput Analysis of the Inferior System For Two Levels of Cascading

For $m = 2$, the following remarks can be made.

1) $\Pr(E_{c,3}) = \Pr(E_{c,4}) = \cdots = P_{c,2}$, and, consequently,

$$\Pr(E_{d,3}) = \Pr(E_{d,4}) = \cdots = 1 - P_{c,2} = P_{d,2}.$$

2) These events are independent of each other and of the previous events. It follows that the average number of transmissions, V, is

$$\begin{aligned} V = {} & 1 \cdot \Pr(E_{c,0}) + 2 \cdot \Pr(E_{d,0}E_{c,1}) + 3 \cdot \Pr(E_{d,0}E_{d,1}E_{c,2}) \\ & + 4 \cdot \Pr(E_{d,0}E_{d,1}E_{d,2})P_{c,2} \\ & + 5 \cdot \Pr(E_{d,0}E_{d,1}E_{d,2})P_{d,2}P_{c,2} \\ & + 6 \cdot \Pr(E_{d,0}E_{d,1}E_{d,2})P_{d,2}^2 P_{c,2} + \cdots. \end{aligned} \quad \text{(B-1)}$$

Let us evaluate the different probability terms in the equation.

i) $\Pr(E_{c,0})$ is simply $P_{c,0}$ given by the equation

$$P_{c,0} = (1 - \varepsilon)^n. \quad \text{(B-2)}$$

ii) $\Pr(E_{d,0}E_{c,1}) = \Pr(E_{c,1} \mid E_{d,0}) \Pr(E_{d,0})$.

$\Pr(E_{c,1} \mid E_{d,0})$ is the probability of correct decoding at the first level, given that an error was detected in the information data. This probability is calculated as

$\Pr(E_{c,1} \mid E_{d,0}) =$
[(probability of correct decoding of this stage)
− (probability of correctable patterns that could not happen because they should have been corrected in the in the previous transmission)]
/ (the probability of patternsthat can occur at this stage that can occur at this stage).

In numerical form, for an (N, K) Hamming code, it reads as follows below in (B-3). If B blocks of K^2 bits are available as information bits at the start, they will be independent after decoding. Then, see (B-4) below.

iii) The second joint probability term is

$$\Pr(E_{d,0}E_{d,1}E_{c,2}) = \Pr(E_{d,0}) \Pr(E_{d,1} \mid E_{d,0}) \cdot \Pr(E_{c,2} \mid E_{d,0}E_{d,1}). \quad \text{(B-5)}$$

The first term on the right side is $\{1 - P_{c,0}\}$. The second term is $1 - \Pr(E_{c,1} \mid E_{d,0})$, which can be obtained from (B-4). The third term is quite involved, and it is sufficient for our purpose to use the bound

$$\Pr(E_{c,2} | E_{d,0}E_{d,1}) > \left\{ \frac{P_{c,2} - P_{c,1}}{1 - P_{c,1}} \right\}^B. \quad \text{(B-6)}$$

This bound is valid as $P_{c,1}$ is more than what should be subtracted from $P_{c,2}$.

iv) The last joint probability term is

$$\Pr(E_{d,0}E_{d,1}E_{d,2}) = \Pr(E_{d,0}) \Pr(E_{d,1} \mid E_{d,0}) \cdot \Pr(E_{d,2} \mid E_{d,0}E_{d,1})$$

where

$$\begin{aligned} \Pr(E_{d,2} \mid E_{d,0}E_{d,1}) &= 1 - \Pr(E_{c,2} \mid E_{d,0}E_{d,1}) \\ &< 1 - \left\{ \frac{P_{c,2} - P_{c,1}}{1 - P_{c,1}} \right\}^B. \end{aligned} \quad \text{(B-7)}$$

To evaluate the throughput of the system, we should calculate the length of each transmission. In any case, the first transmission is of length n. When an (N, K) Hamming code is used, with $x = N - K$ parity check bits, the length of the ith retransmission, ℓ_i, in bits is given as

$$\ell_i = n \left(\frac{N}{K} \right)^{i-1} \left(\frac{x}{K} \right) \qquad i \le m.$$

For $i > m$, the transmitter will transmit (in he case of the inferior system) all stored check bits in addition to information bits. Thus,

$$\ell_i = n \left(\frac{N}{K} \right)^m \qquad i > m.$$

In this case, the receiver will depend only on the instantaneous received block, and ignore all previous bits. Then, for $m = 2$

$$\Pr(E_{c,1} \mid E_{d,0}) = \frac{[(1-\varepsilon)^N + N\varepsilon(1-\varepsilon)^{N-1}]^K - [(1-\varepsilon)^N + (N-K)\varepsilon(1-\varepsilon)^{N-1}]^K}{1 - (1-\varepsilon)^{K^2}}. \quad \text{(B-3)}$$

$$\Pr(E_{c,1} \mid E_{d,0}) = \left\{ \frac{\left[(1-\varepsilon)^N + N\varepsilon(1-\varepsilon)^{N-1}\right]^K - \left[(1-\varepsilon)^N + (N-K)\varepsilon(1-\varepsilon)^{N-1}\right]^K}{1 - (1-\varepsilon)^{K^2}} \right\}^B. \quad \text{(B-4)}$$

$$
\begin{aligned}
T = nP_{c,0} &+ n\left[1+\frac{x}{K}\right]\Pr(E_{d,0}E_{c,1}) \\
&+ n\left[1+\frac{x}{K}+\frac{x}{K}\left(\frac{N}{K}\right)\right]\Pr(E_{d,0}E_{d,1}E_{c,2}) \\
&+ n\left[1+\frac{x}{K}+\frac{x}{K}\left(\frac{N}{K}\right)+\left(\frac{N}{K}\right)^2\right] \\
&\cdot \Pr(E_{d,0}E_{d,1}E_{d,2})P_{c,2} \\
&+ n\left[1+\frac{x}{K}+\frac{x}{K}\left(\frac{N}{K}\right)+2\left(\frac{N}{K}\right)^2\right] \\
&\cdot \Pr(E_{d,0}E_{d,1}E_{d,2})P_{d,2}P_{c,2} \\
&+ \cdots
\end{aligned}
$$

$$
\begin{aligned}
T = n\Bigg\{P_{c,0} &+ \left[1+\frac{x}{K}\right]\Pr(E_{d,0}E_{c,1}) \\
&+ \left[1+\frac{x}{K}+\frac{x}{K}\left(\frac{N}{K}\right)\right]\Pr(E_{d,0}E_{d,1}E_{c,2}) \\
&+ \left[1+\frac{x}{K}+\frac{x}{K}\left(\frac{N}{K}\right)\right]\Pr(E_{d,0}E_{d,1}E_{d,2}) \\
&+ \left(\frac{N}{K}\right)^2\Pr(E_{d,0}E_{d,1}E_{d,2})\left(\frac{1}{P_{c,2}}\right)\Bigg\}. \qquad \text{(B-8)}
\end{aligned}
$$

Then the throughput η is simply k/T.

Acknowledgment

The authors would like to acknowledge King Fahd University of Petroleum and Minerals for its support. They would also like to express their thanks to the editor and the reviewers for their helpful suggestions. They are particularly grateful to one reviewer for suggesting the more compact derivation of Appendix A.

References

[1] W. W. Peterson and E. J. Weldon, Jr., *Error Correcting Codes.* Cambridge, MA: M.I.T. Press, 1978.

[2] S. Lin, D. J. Costello, Jr., and M. J. Miller, "Automatic repeat request error control schemes," *IEEE Trans. Commun.*, vol. 22, pp. 5–17, Dec. 1984.

[3] S. Lin and P. S. Yu, "A hybrid ARQ scheme with parity retransmission for error control of satellite channels," *IEEE Trans. Commun.*, vol. COM-30, pp. 1701–1719, July 1982.

[4] J. J. Metzner, "Improvements in block-retransmission schemes," *IEEE Trans. Commun.*, vol. COM-27, pp. 525–532, Feb. 1979.

[5] Y. M. Wang and S. Lin, "A modified selective-repeat type II hybrid ARQ system and its performance analysis," *IEEE Trans. Commun.*, vol. COM-31, pp. 593–607, May 1983.

[6] H. Krishna and S. D. Morgera, "A new error control scheme for hybrid ARQ systems," *IEEE Trans. Commun.*, vol. COM-35, pp. 981–989, Oct. 1987.

[7] P. Elias, "Error-free coding," *IEEE Trans. Inform. Theory*, vol. IT-4, pp. 29–37, 1954.

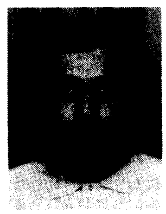
[8] J. K. Wolf, A. M. Michelson, and A. H. Levesque, "On the probability of undetected error for linear block codes," *IEEE Trans. Commun.*, vol. COM-30, pp. 317–324, Feb. 1982.

[9] V. I. Korzhik, "Bounds on undetected error probability and optimum group codes in a channel with feedback," *Radiotecknika*, vol. 1, no. 20, pp. 27–33, 1965.

[10] T. Kasami, T. Klove, and S. Lin, "Error detection with linear block codes," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 131–136, Jan. 1983.

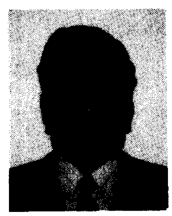
[11] E. R. Berlekamp, *Algebraic Coding Theory.* New York: McGraw-Hill, 1968.

[12] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes.* New York: North-Holland, 1977.



Maan A. Kousa was born in Aleppo, Syria, on October 30, 1963. He received the B.Sc. degree in physics and the B.S. and M.S. degrees in electrical engineering (all with highest honors) from King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, in 1984, 1985, and 1988, respectively.

From 1986 to 1988, he was a Research Assistant in the Department of Electrical Engineering at KFUPM. Since September 1988, he has been a Lecturer in the same department. His research interests include digital communication, coding theory, and network communications.



Mushfiqur Rahman (S'74–M'78) received the B.Sc. degree in electrical engineering from the Bangladesh University of Engineering and Technology, and the M.ASc. and Ph.D. degrees in electrical engineering from the University of Waterloo, Canada, in 1969, 1972, and 1975, respectively.

From 1975 to 1976, he worked at the Brazilian Space Research Institute, San Paulo. From 1976 to 1989, he was a member of the faculty in the Department of Electrical Engineering at King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia. Since July 1989, he has been with Telecom Research Laboratories, Melbourne, Australia, where he is a Senior Engineer. His areas of interests are error-correcting codes and communication theory and systems.