Reliability–throughput optimisation for adaptive forward error correction systems

M. Kousa
L. Turner

Indexing terms: Adaptive forward error correction systems, Performance optimisation

Abstract: Adaptive error control systems are very efficient for digital transmission over time-varying channels. An adaptive forward error correction system based on code selection is considered. An estimation of the channel quality is utilised by the transmitter to adaptively encode the information bits to be sent over the channel. The proper code is selected to optimise the system performance in terms of the two most important evaluation measures: throughput and reliability. The optimisation technique is applied to the case of an adaptive system using BCH codes over Rayleigh fading channels. The achievement of the adaptive system over the non-adaptive system is studied in detail.

1 Introduction

Error control coding is an efficient technique for improving digital transmission over most real channels. Basically there are two fundamental schemes for error control: automatic repeat request (ARQ) and forward error correction (FEC). The former employs error detection while the latter employs error correction. The performance of those schemes is measured in terms of their throughput and reliability. Throughput is defined as the ratio of the average number of information bits delivered to the user per unit time, to the total number of bits that could be transmitted per unit time. Reliability is a measure of the correctness of the decoded data, which is evaluated from the probability of undetected error.

Since there are no retransmissions in an FEC system, the throughput is constant, set by the code rate, regardless of the channel conditions, but the reliability of the system falls down as the channel degrades. On the contrary, an ARQ scheme, 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, and therefore falls rapidly with increasing channel error rate. As a result, ARQ schemes are best suited for good-quality channels which are subject to infrequent bursts of noise; whereas FEC schemes are more suitable for time-invariant noisy channels, where the error-correcting code can be designed to combat the most probable error patterns [1].

For noisy channels that are mostly time-invariant but are subject to infrequent noise impulses, both schemes can be incorporated as in type-I hybrid ARQ systems. But the performance of such systems over time-varying channels, for example fading channels, would not be satisfactory; when the channel is fairly good, the redundancy in the system would be unnecessary and represents a waste of throughput. On the other hand, when the channel is subject to a deep fade the amount of redundancy would not be sufficient to combat the errors. Consequently too many retransmissions are requested, and the throughput falls down. This argument leads to the conclusion that in order to achieve optimum performance over time-varying channels the error correction capability of the system should be adapted to match the prevailing channel conditions.

Adaptive coding techniques can be classified into two groups, incremental redundancy and variable redundancy [2]. In incremental redundancy systems additional check digits are sent on request. The best examples are type-II hybrid ARQ [3], and generalised type-II hybrid ARQ [4] systems. In variable-redundancy systems error-correction codes of different rates are selected according to the channel conditions. A number of different variable redundancy systems have been proposed and analysed [5–8]. These systems are also ARQ/FEC hybrid systems, i.e. they employ both error correction and error detection with retransmissions.

The objective of a system designer is to optimise the performance of the system in terms of the achievable reliability and throughput. In ARQ/FEC hybrid systems the reliability and throughput are determined independently. The reliability is related to the undetected error probability of the error detection code and is (essentially) independent of the error correcting code in use. The undetected error probability of most classes of codes is upper bounded by $P_u \leq 2^{-d}$ where d is the number of check digits [9]. Therefore, the required level of reliability is guaranteed by employing the proper error detection code with sufficient number of check digits. The throughput is then maximised by employing the error correction code with the redundancy that best matches the channel quality. When the redundancy of

IEE Proceedings online no. 19960953

Paper first received 21st December 1995 and in revised form 25th September 1996
M.A. Kousa is with the Electrical Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia

L.F. Turner is with the Department of Electrical Engineering, Imperial College of Science Technology and Medicine, London SW7 2BT, UK

the error correcting code is high, the average number of retransmissions will generally be low, but when the redundancy is low there will be more requests for retransmission. It follows that there is an optimum redundancy associated with the error-correcting code, i.e. a redundancy that maximises the throughput for a specific channel quality (measured in SNR or channel bit error probability).

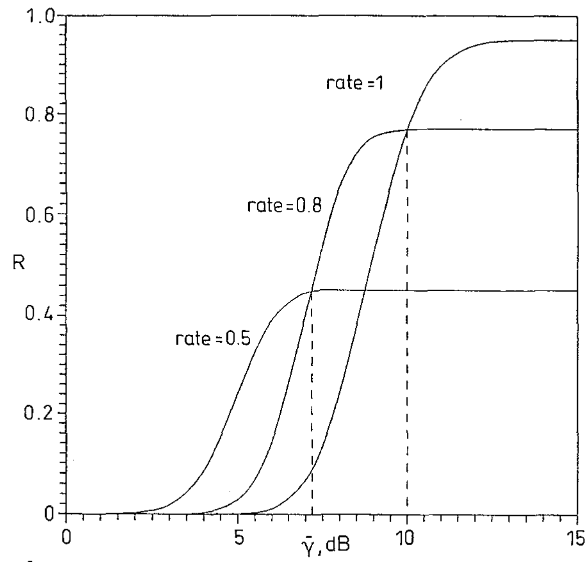

Fig. 1 *Throughput performance of type-I hybrid ARQ system for error-correction codes of different rates*

Fig. 1 shows the throughput curves for type-I hybrid ARQ system using error correction codes of different rates. The intersections between these curves define the regions over which a given code is optimum, i.e. the region over which it provides the maximum throughput (compared to the other codes in the set). If each code is used over the region in which it is optimum, the average throughput of the system is maximised.

In some communication systems, retransmissions are simply not allowed. In such applications only FEC can be used. For FEC schemes, both reliability and throughput are functions of the error correction code, and are thus (inversely) related; the lower the throughput the higher the reliability and vice versa. This paper addresses the issue of reliability–throughput optimisation for adaptive forward error correction (AFEC) systems.

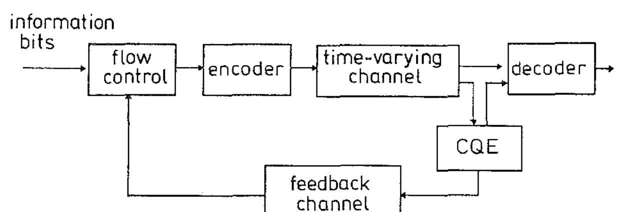

Fig. 2 *Block diagram of AFEC system*

2 Reliability–throughput optimisation

Consider a system such as that illustrated in Fig. 2 in which a transmitter and a receiver are connected by a channel of fixed transmission rate C bits/sec. It is assumed that an estimate of the signal-to-noise (SNR) ratio is obtained by the use of a channel quality estimate (CQE) unit, and is updated periodically. The rate of flow of information bits into the channel is then varied in accordance with the information fed back from the CQE unit. When the channel is good, the rate of flow of information bits is increased and only few (or no) check bits are added by the encoder. On the other hand, when the channel is noisy, the rate of flow of information bits is reduced in order that more check bits can be added for the purpose of carrying out error correction. The information flow and associated choice of error-correction code are to be carried out subject to the constraint of either a specified average (decoded) bit error probability, or a specified average throughput.

The decoded bit-error probability, P, is a function of the code-weight structure and the decoding algorithm, and is unknown for most codes. Hard-decision decoding is assumed here. For this type of decoding many approximations are used in the literature. The calculations carried out in this paper are based on the approximation derived in [10]

$$P \approx \frac{d}{n}\sum_{i=t+1}^{d}\binom{n}{i}\varepsilon^{i}(1-\varepsilon)^{n-i} + \frac{1}{n}\sum_{i=d+1}^{n}\binom{n}{i}\varepsilon^{i}(1-\varepsilon)^{n-i} \tag{1}$$

where n, d and t are respectively the length, the minimum distance and the error-correction capability of the code, and ε is the channel bit-error rate. It is shown in [10], by numerical simulation of few examples, that this approximation is very close to the exact value. The approximations used in [11] were also tested, and the differences in the results obtained were insignificant.

It should be noted that ε and t are related to the SNR, γ, and to the rate, r, of the code, once the modulation scheme and the code are specified. Thus $P(\varepsilon, t)$ can be expressed in terms of $P(\gamma, r)$. it is important to note that in the present analysis the rate, r, of the code is also a function of γ. So, in fact, P is a function of γ only. However, to keep the analysis more tractable, the post-decoding bit-error probability will be denoted by $P(\gamma, r(\gamma))$.

For a time-varying channel which is described by the probability density function $f(\gamma)$ of the received SNR, the average post-decoding bit error probability is evaluated as:

$$\bar{P} = \frac{1}{R}\int_{0}^{\infty} r(\gamma)P(\gamma, r(\gamma))f(\gamma)d\gamma \tag{2}$$

where

$r(\gamma) \equiv$ the function describing the change in the code rate in accordance with the measured γ,

$f(\gamma) \equiv$ p.d.f of the SNR of the time varying channel, and

$R \equiv$ the average throughput of the system.

Although the post-decoding bit error probability is the same for both information bits and check bits, more information bits are transmitted with a high-rate code than with a low-rate code. The factor $r(\gamma)/R$ is a weighting factor which is introduced to make up for the different contributions, in terms of number of information bits, of the different codes. The average throughput, R, at which the system is operating is given by:

$$R = \int_{0}^{\infty} r(\gamma)f(\gamma)d\gamma \tag{3}$$

In arriving at eqns. 2 and 3, it was assumed that the channel is slowly varying such that γ remains constant over the duration of at least one codeword.

The aim of the analysis is to solve for the rate function $r(\gamma)$ which achieves the optimum performance. We define the optimum performance as that of:

(1) minimising the post-decoding bit error probability subject to the constraint of having to operate at a specified average throughput, or

(2) maximising the average throughput subject to the constraint of having to operate at a specified post-decoding bit error probability.

2.1 System performance under throughput constraints

In this case eqn. 2 is to be minimised while satisfying the constraint in eqn. 3 for a specified value of R, say the value ρ. Such constraint problems can be solved by the Lagrange multiplier technique. The objective function and the constraint are combined to form the new function

$$F = \bar{P} + \omega(R - \rho) \tag{4}$$

where ω is the Lagrange multiplier. The technique requires the differentiation of the composite function P with respect to $r(\gamma)$ and ω. Unfortunately, the dependency of $P(\gamma, r(\gamma))$ on $r(\gamma)$ cannot be put in a form which renders differentiation possible; so the Lagrange multiplier approach is not applicable to eqn. 2 in its present form.

From a practical point of view, the process of varying the rate of the code reduces to selecting one code from a set of available codes for an interval (region) of the estimated γ. The optimisation task may then be simplified, without any practical loss, by partitioning the interval into m sub-intervals in each of which a selected fixed-rate code is employed. In this way, the dependence of r on γ within the region is removed. Eqns. 2 and 3 can then be rewritten as

$$\bar{P} = \frac{1}{R}\sum_{i=0}^{m-1} r_{i+1} \int_{a_i}^{a_{i+1}} P\left(\gamma, r_{i+1}\right) f(\gamma) d\gamma \tag{5}$$

$$R = \sum_{i=0}^{m-1} r_{i+1} \int_{a_i}^{a_{i+1}} f\left(\gamma, r_{i+1}\right) f(\gamma) d\gamma \tag{6}$$

where $a_0 = 0$, $a_m = \infty$ and $r_1, r_2, \ldots, r_m$ are the rates of the error correcting codes. Clearly, $0 \le r_1 < r_2 < \ldots r_m \le 1$. The value $r_1 = 0$ refers to the case where the transmitted block contains no information bits. (However it would be necessary to keep on transmitting redundant bits to maintain the CQE unit operational). The value $r_m = 1$ refers to the case where the transmitted block contains all information bits and no check bits. The optimisation problem reduces to solving for the breakpoints $a_1, a_2, \ldots, a_{m-1}$ at which adaptation (changing from one code to another) should take place.

For the sake of simplicity of notation, let $P_i(\gamma)$ denote $P(\gamma, r_i)$ which represents the bit error probability when the code (n, k_i) of rate $r_i = k_i/n$ is used. By differentiating eqn. 4 with respect to each of the breakpoints a_i the following set of equations are obtained:

$$\frac{\partial F}{\partial a_i} = \frac{f(a_i)}{R}\Big\{ r_i P_i(a_i) - r_{i+1} P_{i+1}(a_i) - (r_i - r_{i+1})(\bar{P} - \omega R) \Big\},\ i = 1, \ldots, m-1 \tag{7}$$

Also, on differentiating F with respect to the Lagrange multiplier, ω, the following equation is obtained:

$$\frac{\partial F}{\partial \omega} = \left\{ \sum_{i=0}^{m-1} r_{i+1} \sum_{a_i}^{a_i+1} f(\gamma) d\gamma \right\} - \rho \tag{8}$$

The desired breakpoints are obtained by equating the m derivatives in eqns. 7 and 8 to zero and solving the set of the resulting m equations. The appropriate code rate is selected by comparing the measured γ with the set of thresholds.

2.2 System performance under bit-error rate constraints

The method of analysis developed earlier can be applied to the case where it is desired to obtain maximum average throughput subject to the constraint that the average bit-error probability, $\bar{P}$, is maintained at a prescribed value, say δ. In this case eqn. 3 is taken as the objective function to be maximised and eqn. 2 is the constraint. The composite function is then

$$F = R + \omega\left(\bar{P} - \delta\right) \tag{9}$$

The threshold values are found by differentiating eqn. 9 with respect to the a_is and with respect to ω, then equating the resulting derivatives to zero, as outlined above.

Clearly, in order for the system to operate properly, both the transmitter and the receiver must know the code in use. This implies either that the feedback channel is noise-free, or, more realistically, that the information sent back via the feedback channel can be adequately protected by coding. Since the information that has to be fed back is small (simply the code number), such protection, and hence the reliability of the system, can be made as high as necessary, and, for all practical purposes, the feedback channel can be rendered error-free.

3 System performance over a Rayleigh fading channel using BCH codes

3.1 Fixed average throughput

The optimisation technique developed above has been applied to transmission over a Rayleigh fading channel. The received signal-to-noise ratio, γ, for a Rayleigh fading channel has an exponential probability density function, i.e.

$$f(\gamma) = \frac{1}{\bar{\gamma}} \exp\left(-\frac{\gamma}{\bar{\gamma}}\right), \quad \gamma \ge 0 \tag{10}$$

where $\bar{\gamma}$ is the average received SNR.

The class of BCH codes is used as an example for the ideas developed above. BCH codes constitute a powerful class of multiple error-correcting codes, especially at moderate lengths. More importantly from the point of view of the present work, they posses the flexibility that for a given codeword length, n, codes having a large range of information bits, k, can be constructed. This property allows smooth adaptation. Moreover, BCH codes are cyclic and therefore enjoy the simplicity of the decoding algorithms of cyclic codes. The decoders may be implemented in software or hardware, although the former is more appropriate for the application presented here. Since the length of the code is kept fixed, the adaptation process results in minor modifications of the decoders.

Two sets of codes were examined; one of length 31 bits, and the other of length 63 bits. The first set is composed of seven codes and the second set is composed of 15 codes, as given in Table 1, which is extracted from Appendix D in [12]. The notation (n, k, t) is the familiar notation in which t denotes the guaranteed error correction capability of the code.

Table 1: BCH codes of lengths 31 and 63

Code	Length–31 codes		Length–63 codes	
	(n,k,t)	r	(n,k,t)	r
1	(31, 00, –)	0.0	(63, 00, –)	0.0
2	(31, 06, 7)	0.19	(63, 07, 15)	0.11
3	(31, 11, 5)	0.36	(63, 10, 13)	0.16
4	(31, 16, 3)	0.52	(63, 16, 11)	0.25
5	(31, 21, 2)	0.68	(63, 18, 10)	0.28
6	(31, 26, 1)	0.84	(63, 19, 09)	0.30
7	(31, 31, 0)	1.0	(63, 21, 08)	0.33
8	–	–	(63, 28, 07)	0.44
9	–	–	(63, 30, 06)	0.48
10	–	–	(63, 36, 05)	0.57
11	–	–	(63, 39, 04)	0.62
12	–	–	(63, 46, 03)	0.73
13	–	–	(63, 51, 02)	0.81
14	–	–	(63, 57, 01)	0.90
15	–	–	(63, 63, 00)	1.0

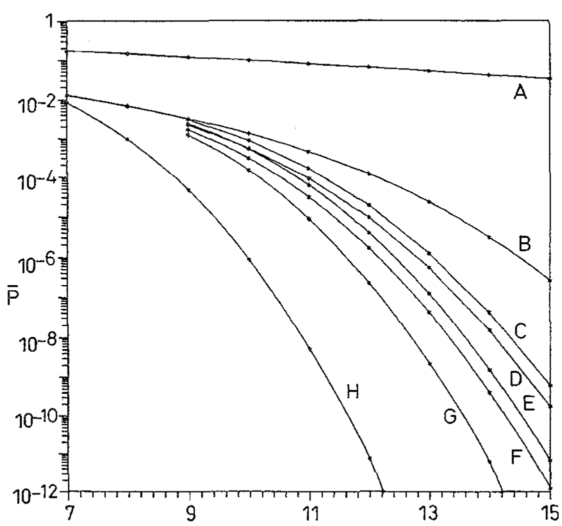

Fig. 3 *Reliability of AFEC compared to the non-adaptive and non-fading length-31 BCH codes ($R = 0.5$)*
Curve A: code 4 (over fading channel)
Curve B: codes 1, 7
Curve C: codes 1, 6, 7
Curve D: codes 1, 4, 7
Curve E: codes 1, 5, 6, 7
Curve F: codes 1, 3, 6, 7
Curve G: all seven codes of Table 1
Curve H: code 4 (over non-fading channel)

In the example considered it was assumed that the average throughput, ρ, is equal to 0.5. The results obtained for the length-31 codes are shown in Fig. 3. Curve A shows the situation when a single fixed-rate code, the (31, 16, 3) code, of rate 16/31 ≅ 0.5, is used in a non-adaptive manner over the fading channel. Curve H shows the performance of the same code operating over a non-fading channel. These curves provide limits of the achievable performance. Curve B refers to the simplest adaptive system which operates in one of two modes, based on one threshold value; if the estimated γ is below the threshold, no information bits are transmitted (OFF mode), and if it is above the threshold, only information bits are transmitted (ON mode). In both modes no coding is involved. Curves C–G refer to the various modes of the AFEC system, as explained in the following paragraph.

Curves C and D are obtained when the region $0 \leq \gamma < \infty$ is partitioned into three intervals only, with one code from the set being allocated to each region. The results of the analysis showed that larger improvement in performance is achieved if the system is allowed to use the (31, 0, -) code, i.e no information is transmitted, for the region having the lowest SNR, and the (31, 31, 0) code, i.e. all 31 transmitted bits are information bits, for the region of highest SNR. The two curves are based on the use of these two codes. They differ only in respect of the code used for the intermediate interval. Curves C and D relate to the situation in which the codes (31, 26, 1) and (31, 16, 3), respectively, are used for the intermediate SNR region. Curves F and F show the results for a four-interval adaptation. In this case, the curves differ in respect of the two codes used for the two intermediate regions. Curve G represents the system performance for seven-interval adaptation. The table provided below Fig. 3 lists the codes associated with each of the curves A to H.

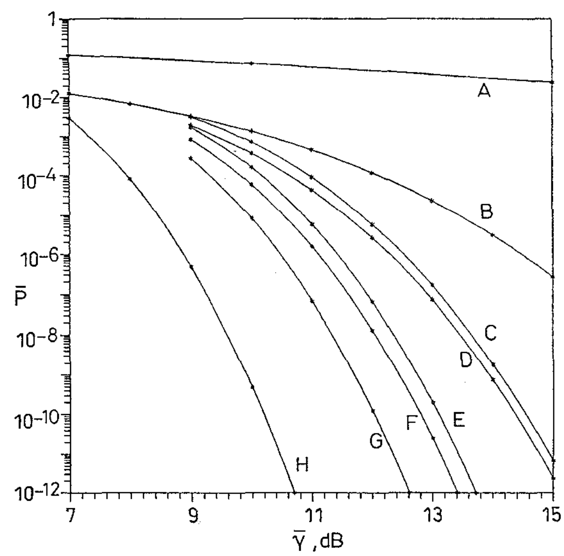

Fig. 4 *Reliability of AFEC compared to the non-adaptive and non-fading length-63 BCH codes ($R = 0.5$)*
Curve A: code 9 (over fading channel)
Curve B: codes 1, 15
Curve C: codes 1, 13, 15
Curve D: codes 1, 10, 15
Curve E: codes 1, 10, 11, 12, 13, 14, 15
Curve F: codes 1, 3, 5, 8, 11, 13, 15
Curve G: all fifteen codes of Table 1
Curve H: code 9 (over non-fading channel)

Fig. 4 is similar to Fig. 3, but is obtained using the set of codes of length 63 bits. The list provided below Fig. 4 indicates the number of intervals and the associated codes used when arriving at curves A to H in the figure. By examining Figs. 3 and 4 the following observations can be made;

(i) for the values of γ covered in the two figures, the performance of the fixed-rate system is not acceptable for almost all types of communication.

(ii) By using a simple ON–OFF scheme (curve B in both figures) a very significant improvement can be obtained in comparison with the non adaptive system. The reason underlying this improvement is that such a system avoids transmitting information during periods of deep fade. Such a transmission would be very unreliable and would badly degrade the overall average performance. The price to be paid for not transmitting information during the periods of deep fades is that unprotected data has to be transmitted when the channel is good. In general, however, the effect of this is minimal, because the probability of error occurring during these periods is small. The net average improvement is shown as the difference between curves A and B in Figs. 3 and 4.

(iii) Further improvement can be obtained by increasing the number of intervals, since this results in better matching to the channel conditions. However, the improvement is of diminishing order; the largest relative improvement is achieved from increasing the number of intervals from one to two.

(iv) The choice of the codes for the different intervals has a significant effect on performance. This can be seen by comparing curves C and D, or E and F in either of the two figures. By examining the code sets of these curves, it can be seen that an improved performance is obtained when using a set of codes that are more evenly distributed. In fact, a poor assignment of codes to the intervals can waste most of the advantage of adaptation.

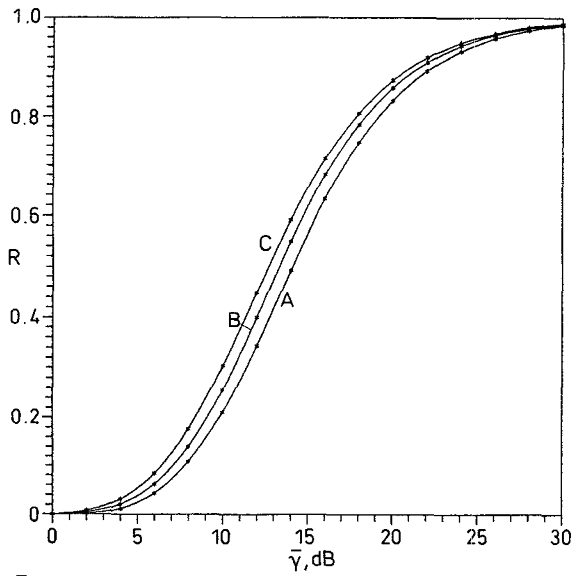

Fig. 5 *Throughput of AFEC system, using length-31 BCH codes ($\delta = 10^{-8}$)*
Curve A: codes 1, 4, 7
Curve B: codes 1, 3, 6, 7
Curve C: all seven codes of Table 1

3.2. Fixed average bit-error probability

Fig. 5 illustrates some results obtained with the AFEC system using length-31 BCH codes, operating subject to the constraint that $\delta = 10^{-8}$. Curves A, B and C are obtained for three, four and seven steps of adaptation, respectively. The list provided with Fig. 5 indicates the code sets associated with each of the three curves. Similarly, curves A, B and C in Fig. 6 show the results for three, seven and fifteen intervals, respectively, when using the length-63 BCH code sets as indicated in the list. The following points are worth noting;

(i) Increasing the number of intervals has a significant effect on the achievable throughput rate. For example, at $\bar{\gamma}$ = 12dB, a throughput gain of approximately 10% can be achieved when using seven intervals rather than three intervals. This applies to both figures. However, the improvement in throughput is of diminishing nature; increasing the number of intervals to 15 in Fig. 6 has a much smaller impact (3).

(ii) The greatest benefit to be had from the adaptation process is obtained for channels of moderate quality ($6 < \bar{\gamma} < 22$dB). For poor channels, although a full range of code rates might be available, the system has to operate, most of the time, with the lowest-rate codes in order to achieve a BER of 10^{-8}. In fact, for a considerable percentage of time the system will be in the OFF mode.

(iii) Presenting the results in the form shown in Figs. 5 and 6 is very useful as practical system design is concerned. In most data communication systems, a certain level of reliability is required for the system to be useful. Failure to attain this specified level of reliability may render the system unacceptable even though the throughput may be satisfactory. The above analysis provides a method for designing a forward error correction system at a desired degree of reliability, and determines the maximum average throughput that can be obtained, and how it could be obtained. To the knowledge of the authors, this form of presentation for FEC schemes is reported for the first time. That also provides a basis for meaningful companson with ARQ schemes quantitatively, which is currently investigated by the authors.

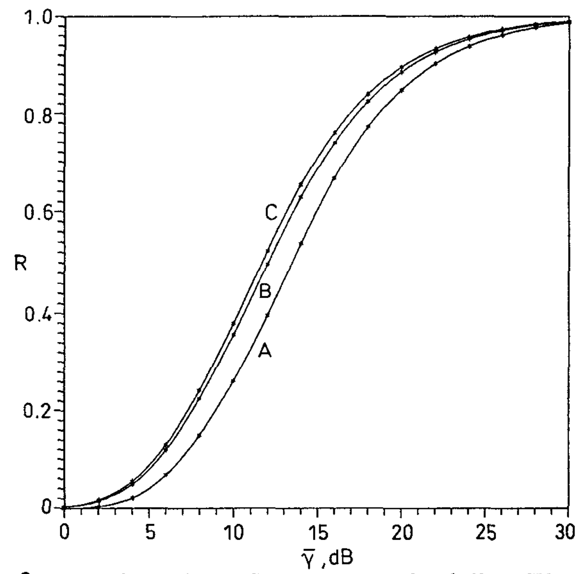

Fig. 6 *Throughput of AFEC system, using length-63 BCH codes ($\delta = 10^{-8}$)*
Curve A: codes 1, 10, 15
Curve B: codes 1, 3, 4, 8, 11, 13, 15
Curve C: all fifteen codes of Table 1

4 Conclusions

In this paper, it has been shown how forward error correction systems can be used in an adaptive way to improve the performance of communication transmission. Two important design questions were considered, namely (i) subject to the constraint of having to operate at a given average error probability, what is the maximum throughput that can be achieved? and (ii) subject to the constraint of having to operate at a given throughput, what is the minimum average error probability that can be achieved? For each case, the breakpoints that lead to the optimum performance are obtained. The adaptation process reduces, then, to the selection of the right code based on comparing the estimated SNR of the channel with the breakpoints. it was found that an adaptive forward error correction can be made significantly more reliable than non-adaptive fixed-rate forward error correction systems, with greater enhancement achieved when using more codes (i.e. more adaptation intervals).

5 Acknowledgment

King Fahd University of Petroleum and Minerals (KFUPM) is acknowledged for its support in carrying out this work.

6 References

1 BERLEKAMP, E.R., PEILE, R.E., and POPE, S.P.: 'The application of error control to communication', *IEEE Commun. Mag.*, April 1987, **25**, (4), pp. 44–57

2 BATE, S.D., HONARY, B., and FARRELL, P.G.: 'Error control techniques applicable to HF channels', *IEE Proc. I*, Feb. 1989, **136**, (1), pp. 57–63

3 LIN, S., CASTELLO, D., and MILLER, M.: 'Automatic repeat request error control schemes', *IEEE Commun. Mag.*, Dec. 1984, **22**, pp. 5–17

4 KRISHNA, H., and MORGERA, S.D.: 'A new error control scheme for hybrid ARQ systems', *IEEE Trans. Commun.*, Oct. 1987, **COM-35**, pp. 981–989

5 GOODMAN, R.M., and FARRELL, P.G.: 'Data transmission with variable redundancy error control over a high frequency channel', *IEE Proc.*, Feb. 1975, **122**, (2), pp. 113–118

6 VUCETIC, B., DRAJIC, D., and PERISIC, D.: 'Algorithm for adaptive error control system synthesis', *IEE Proc. F*, Feb. 1988, pp. 85–94

7 VUCETIC, B.: 'An adaptive coding scheme for time-varying channels', *IEEE Trans. Commun.*, May 1991, **39**, (5), pp. 653–663

8 PURSLEY, M.B., and SANDBERG, S.D.: 'Variable-rate hybrid ARQ for meteorburst communications', *IEEE Trans. Commun.*, Jan. 1992, **40**, (1), pp. 60–73

9 WOLF, J.K., MICHELSON, A.M., and LEVESQUE, A.H.: 'On the probability of undetected error for linear block codes', *IEEE Trans. Commun.*, Feb. 1982, **COM-30**, (2), pp. 317–324

10 TORRIERIE, D.J.: 'The information-bit error rate for block codes', *IEEE Trans.*, April 1984, **COM-32**, (4), pp. 475–476

11 MICHELSON, A.M., and LEVESQUE, A.H.: 'Error-control techniques for digital communication' (John Wiley & Sons, New York, 1985)

12 PETERSON, W.W., and WELDON, E.J.: 'Error correcting codes' (MIT Press, Cambridge, 1972)