A Bound on the Reliability of Block Coding with Feedback

By J. E. SAVAGE

(Manuscript received March 25, 1966)

A conceptually simple, block-coding feedback strategy which applies to all time-discrete, memoryless channels is introduced and examined. This strategy provides the first conclusive evidence of an improvement at nonzero rates in the reliability of block coding with feedback on the additive Gaussian noise channel. This result was previously observed by Berlekamp for the binary symmetric channel.

I. INTRODUCTION

Coding with feedback has been considered by a number of authors.[1,2,3,4,5,6] principally because of the advantage it is expected to enjoy in rate, reliability, and equipment costs over coding without feedback. Some early feedback results were obtained by Shannon, who showed that channel capacity of a discrete memoryless channel (DMC) cannot be increased using feedback[1] and established that the sphere-packing (lower bound[7,8]) to the probability of error without feedback applies to *block coding* with feedback on the DMC uniform at the input (the sets of transition probabilities from each channel input letter are identical except for permutations[7]). He also conjectured that a sphere-packing bound applies to *block coding* with feedback on all discrete, memoryless channels.[2] (A conjecture supported recently by Berlekamp.[9])

It has been shown that the exponent on the sphere-packing bound agrees with the no-feedback, random-code exponent at rates above the critical rate R_{crit},[7,10] so that feedback cannot increase the reliability of block coding at rates greater than R_{crit}. Below R_{crit}, Berlekamp first showed that feedback will improve the reliability of block coding.[3] He showed that the zero rate exponent of the probability of error with block codes on the binary symmetric channel (BSC) and certain other binary channels having particular symmetries is larger with feedback than the best no-feedback exponent at zero rate. Also, he showed that the error-correction capability of block codes on the BSC at rates between zero and R_{crit} is improved with feedback.

Variable-block-length coding strategies, that is strategies where the code length is controlled through the feedback channel, have been proposed and in contrast with the block coding strategies have been shown to operate with a greater reliability than that given by the sphere-packing bound.[4,5] Hence, the best variable-length feedback strategy is superior to the best block-coding feedback strategy. The block-coding strategies, however, are interesting since any improvement in coding reliability observed with them can be ascribed directly to the effect of feedback on the choice of codewords representing messages. This is not true of the variable-length strategies since, in this case, some unknown fraction of the improvement in reliability is attributable to the variation of the code length with the level of channel noise.

In this paper, we shall be concerned only with *block-coding* with feedback and in particular with one particular block-coding strategy. This is a strategy which makes efficient but not complete use of the feedback channel. (This will be seen from the results.) It is recommended, however, by its simplicity, its easy evaluation in terms of known, (no-feedback) bounds, by its application to a large class of channels including all time-discrete, memoryless channels, by the substantial improvements it shows for many channels over the no-feedback, random-code exponents and the improvement it shows for some channels over the best upper bounds on no-feedback exponents. In particular, we see an improvement over the random-code exponents over a middle range of rates on the BSC with crossover probability less than 10^{-7} and on the average-power-limited, additive Gaussian noise channel (AGC) with signal-to-noise ratio (S/N) greater than 11.5 dB. This is interesting, since some believe that the random-code exponents are the best, no-feedback, block-coding exponents. Also, we see an improvement, again over a middle range of rates, over the Wyner upper bound[11] to the no-feedback exponent on the AGC with S/N $\geqq$ 22 dB.

It is clear that our strategy does not make full use of the feedback channel since the improvements noted apply only to cleaner channels and then only over a middle range of rates. Also, it can be shown that our strategy has an exponent which is smaller than the exponent implied by Berlekamp's results for the BSC.[3] It should be noted, however, that the BSC is the only channel for which it has been previously shown that feedback can improve the reliability of block coding.

II. THE CODING STRATEGY

The feedback, block-coding strategy which we present was suggested by arguments used by Shannon and Gallager to *underbound* the probability of error with block coding.[12] Our strategy, however, leads to an overbound to the probability of error with the best feedback strategy since it does not make the most efficient use of the return channel. We will describe this strategy without referring directly to the input and output alphabets of the forward channel. We assume only that the forward channel is time-discrete. With respect to the return channel, we assume that it is noiseless and of large but finite capacity. The forward and reverse channels are allowed to have a total delay of D channel symbols.

Our feedback, block-coding strategy is a 2-step procedure. For each of the two steps a block code is chosen and used without feedback. In the first step, a codeword from a code of M codewords each having length N_1, is chosen to transmit one of the M messages generated by the source. In the second step, a code of L codewords* where each word has length N_2 is used. Here N_1 and N_2 are chosen so that $N_1 + N_2 + D = N$, the number of channel symbol intervals allowed for the transmission of one of the M source messages.

The decoder receives a noisy version of the codeword chosen from the first code and he then makes a list of the L messages which are most likely given the received signal.† We assume that all messages are equally likely, *a priori*, so that the L messages on the list are those which have the largest likelihood probability. That is, if $p(\mathbf{v}_1 \mid m)$ is the probability (or probability density, if the channel output alphabet is continuous) of receiving the N_1 channel letters represented with $\mathbf{v}_1$ when message m is transmitted, then messages $m_1, m_2, \cdots, m_L$ are on the list if

$$p(\mathbf{v}_1 \mid m_i) \geqq p(\mathbf{v}_1 \mid m') \qquad \text{all } m' \neq m_i, \qquad 1 \leqq i \leqq L. \quad (1)$$

Once the list has been formed, the list and the order in which messages appear on the list is sent over the reverse channel to the transmitter. Since D time intervals will elapse before the list reaches the transmitter, the transmitter is ready to begin the second of the two steps after $N_1 + D$ intervals. In the second step, the transmitter chooses a codeword from the second code of length N_2 to indicate which message on

* L is arbitrary here but is fixed in later discussion.
† This decoding procedure is called "list decoding."[8,13]

the list of the L messages is the source output. If the source output is not on the list, the first codeword is transmitted. The received sequence is decoded using the max-likelihood decoding rule.

III. EVALUATION OF THE STRATEGY

Our coding strategy which first establishes a list of most probable transmitted messages and then resolves the ambiguity in the list can lead to a decoding error in either of two ways. First, the message delivered by the source may not be on the list because of excessive channel noise in the first N_1 transmissions or, second, it may be on the list but the list may be decoded in error during the last N_2 transmissions.

Now, the probability of a decoding error with the best feedback strategy, $P_{ef}(N,M1)$, is less than or equal to $P_f(N,M,1)$, the probability of error, with the feedback strategy given above. This, in turn, is less than or equal to the sum of the probability of list decoding error with the best list code (and no feedback), $P_e(N_1, M, L)$, and the probability of a max-likelihood decoding error with the best max-likelihood code, $P_e(N_2, L, 1)$. Thus, we have

$$P_F(N,M,1) \leqq P_f(N,M,1) \leqq P_e(N_1, M, L) + P_e(N_2, L, 1) \quad (2)$$

where

$$N = N_1 + N_2 + D \quad (3)$$

and D is the round-tip delay.

While our feedback strategy can be evaluated for any forward channel for which bounds to $P_e(N_1, M, L)$ and $P_e(N_2, L, 1)$ are known, we restrict our attention here to the discrete memoryless channel and to the time-discrete, average-power-limited, additive Gaussian noise channel. Bounds to these error probabilities for these two channels can be found in several places.[7,10,14,15] However, for easy reference we shall refer to Gallager.[10] Gallager does not bound $P_e(N_1, M, L)$ in his paper but the changes necessary in his analysis to bound it are relatively easy to effect and are outlined in the Appendix. These changes are due to unpublished results by Gallager* and are such that the random-code bound to $P_e(N_1, M, L)$ has the same form as Gallager's bound[10] to $P_e(N_1, M, 1)$ except that his parameter ρ is allowed to range between 0 and L rather than between 0 and 1.

The bounds to $P_e(N_1, M, L)$ and $P_e(N_2, L, 1)$ are given below where $R_1 = (\log_2 M)/N_1$, $R_2 = (\log_2 L)/N_2$ and $O_i(N_i)$, $1 \leqq i \leqq 2$, are quan-

* See also Ref. 16.

tities which approach zero faster than $1/N_i$:

$$P_e(N_1, M, L) \leqq 2^{-N_1[E_L(R_1)+O_1(N_1)]} \tag{4}$$

$$P_e(N_2, L, 1) \leqq 2^{-N_2[E_x(R_2)+O_2(N_2)]}. \tag{5}$$

Here $E_L(R_1)$ and $E_x(R_2)$ are the random-code bound and expurgated random-code bound, respectively, to the exponents on the two error probabilities. The formal statement of these two exponents for the DMC and the AGC is somewhat long. Consequently, we present Table I which lists the location of the two exponents by equation number in Ref. 10. We note again that $E_L(R_1)$ has the same form as $E_1(R_1)$, which is the random-code exponent given by Gallager, except that $0 \leqq \rho \leqq L$.

Equations (4) and (5) are used to bound (2). The block lengths

Table I — Location of Exponents in Ref. 10.

	DMC	AGC
$E_L(R_1)$	21,22	125,127,128
$E_x(R_2)$	86,87	133

N_1 and N_2 of the two codes are chosen before transmission to approximately optimize the bounds (4) and (5). That is, we set

$$N_1E_L(R_1) = N_2E_x(R_2). \tag{6}$$

Since $N_2 = (N - D) - N_1$ we have

$$N_1 = \frac{(N - D)E_x(R_2)}{E_x(R_2) + E_L(R_1)}. \tag{7}$$

Thus, the exponent $N_1E_L(R_1)$ becomes

$$N_1E_L(R_1) = \frac{(N - D)E_x(R_2)E_L(R_1)}{E_x(R_2) + E_L(R_1)}. \tag{8}$$

The signaling rate of the code is defined as $R = (\log_2 M)/N$ so that

$$R = \frac{N_1}{N} R_1 = \frac{\left(1 - \frac{D}{N}\right) R_1E_x(R_2)}{E_x(R_2) + E_L(R_1)}. \tag{9}$$

The exponent to the probability of error with our feedback strategy, $E_f(R)$, is defined as

$$E_f(R) = \lim_{N\to\infty} - \frac{\log_2 P_f}{N} \tag{10}$$

so that if we assume that $D/N \to 0$ *as* N *increases* and if L *is independent of* N ($R_2 \to 0$), then

$$E_f(R) = \frac{E_x(0)E_L(R_1)}{E_x(0) + E_L(R_1)} \tag{11}$$

where

$$R = \frac{R_1E_x(0)}{E_x(0) + E_L(R_1)} \cdot \tag{12}$$

Hence, $E_f(R)$ and R are parametrized by the rate R_1.

Let us now consider the list size L and show that it should be made independent of N. The list size appears in (11) and (12) through the list decoding exponent $E_L(R_1)$. As mentioned above, this exponent is parametrized with a parameter ρ, $0 \leqq \rho \leqq L$. Associated with $E_L(R_1)$ is a rate R_L^* such that if $R_1 < R_L^*$ then $E_L(R_1)$ is a straight line of slope $-L$. Also, if $R_1 > R_L^*$ then $E_L(R_1)$ has slope $-\rho$, $\rho < L$. Thus, $E_L(R_1)$ is largest for fixed R_1 if L is such that $R_1 > R_L^*$. For any R_1, the L satisfying this inequality is fixed and finite. Now, examination of (11) and (12) will show that $E_f(R)$ vs R is largest when $E_L(R_1)$ is largest so that $E_f(R)$ is maximized over L with a value of L which is independent of N. We choose to use that L for which $R_1 > R_L^*$ so that the arguments made in the previous paragraph hold.

It can be shown[7] that the exponent $E_L(R_1)$ is equal to the sphere-packing bound $E_{sp}(R_1)$ for $R_1 \geqq R_L^*$. Thus, we have, as our final result, the achievable (lower) bound, $E_f(R)$, to the *largest obtainable exponent with feedback*, $E_F(R)$, given below:

$$E_F(R) \geqq E_f(R) = \frac{E_x(0)E_{sp}(R_1)}{E_x(0) + E_{sp}(R_1)} \tag{13}$$

where

$$R = \frac{R_1E_x(0)}{E_x(0) + E_{sp}(R_1)} \cdot \tag{14}$$

A simple construction for $E_f(R)$ from $E_x(0)$ and the sphere-packing bound is shown in Fig. 1.

IV. SOME EXAMPLES

The exponent $E_f(R)$ on the BSC is significantly smaller than the exponent implied by Berlekamp's results and for this reason is not shown.

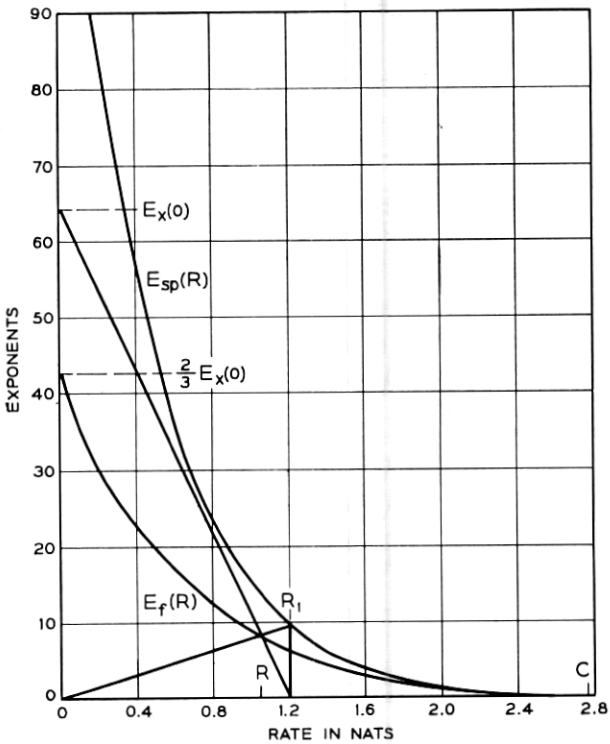

Fig. 1 — Construction for E_f (R) on AGC, S/N = 256.

A computer study of $E_f(R)$, however, shows that $E_f(R)$ is larger than the random code bounds with and without expurgation over a middle range of rates when the crossover probability $p \geqq 10^{-7}$. It is found that this range of rates increases with decreasing p.

The exponent $E_f(R)$ is shown in Figs. 2 and 3 for the time-discrete, average-power-limited, additive Gaussian noise channel with power signal-to-noise ratios of 64 (18.1 dB) and 256 (24.1 dB), respectively. In the first case, an improvement is seen over both the expurgated and unexpurgated random code exponents for a very large range of rates, namely, $0.15 \leqq R \leqq 1.6$, in nats, and in the second case $E_f(R)$ is larger than the Wyner *upper bound*[11] to the exponent without feedback, $E_w(R)$, for a substantial range of rates, namely $0.90 \leqq R \leqq 1.80$. Computer calculations have shown that if S/N < 160 (22 db), then $E_f(R) < E_w(R)$ and if S/N < 14 (11.5 dB), then $E_f(R) \leqq E_{rc}(R)$, which is the random-code exponent without feedback.

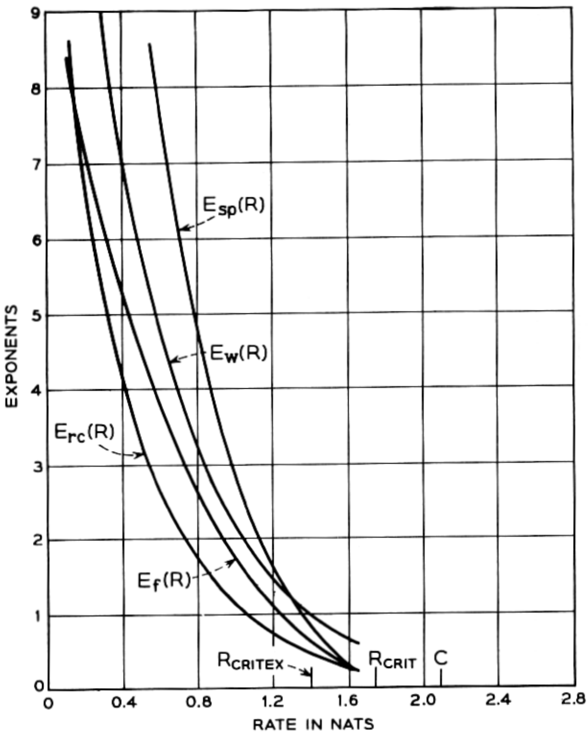

Fig. 2 — Exponents on AGC, S/N = 64.

V. CONCLUSIONS

We have introduced and examined a conceptually simple block-coding, feedback strategy. Using this strategy as an example of block-coding feedback strategies, we have found the first conclusive evidence of an improvement in the reliability of block coding with feedback at nonzero information rates on the additive Gaussian noise channel. This improvement is measured with the exponent on the probability of error and we have shown that an exponent larger than the random-code (lower bound) exponent can be obtained on many channels such as the relatively clean BSC and AGC and that on at least one channel, the AGC with S/N $\geqq$ 22 dB, an exponent which is superior to the best no-feedback exponent can be achieved. These results have been shown to hold when there is a nonzero channel delay as long as that delay does not grow as fast as linearly with block length.

The gain in reliability seen above can be translated into reduced

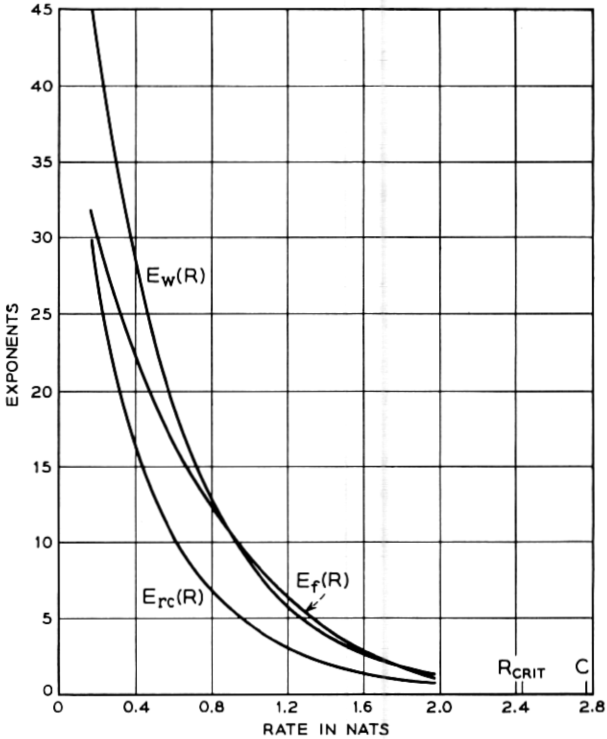

Fig. 3 — Exponents on AGC, S/N = 256.

equipment costs or increased rate. We have made our comparison of coding with and without feedback on the basis of fixed rate and block length.

It is important to emphasize that the improvement in the performance of block coding using feedback is due entirely to the fact that the codewords in the code are allowed to change with the channel noise. In our example, the list formed after the first step changes with the level of the channel noise so that the association of codewords in the second code with messages on the list becomes channel dependent.

VI. ACKNOWLEDGMENT

The author gratefully acknowledges many helpful conversations with P. M. Ebert during this work. He also acknowledges his help in the calculation of the Gaussian channel exponents.

APPENDIX

Equation (1) states the rule for choosing messages $m_1, m_2, \cdots, m_L$ to be placed on the list when list decoding of list size L is used and the channel sequence $\mathbf{v}_1$ is received. An error is made if the transmitted message m, say, is not on this list. Using an argument similar to that given in Ref. 10, (2) through (6), the probability of error with list decoding when message m is received can be stated formally for a particular code as

$$P_{em} = \sum_{\mathbf{v}_1 \in V_{N_1}} P(\mathbf{v}_1 \mid \mathbf{x}_m)\Phi_m(\mathbf{v}_1).$$

Here V_{N_1} is the set of channel sequences $\{\mathbf{v}_1\}$ of length N_1, $\mathbf{x}_m$ is the mth code word in the code and $\Phi_m(\mathbf{v}_1)$ is a characteristic function which is one if $\mathbf{v}_1$ results in a list which does not include m and is zero otherwise. Gallager, in unpublished work, has shown that the characteristic function $\Phi_m(\mathbf{v}_1)$ may be overbounded by the following

$$\Phi_m(\mathbf{v}_1) \leq \left\{ \sum_{\substack{m_1'}} \cdots \sum_{\substack{m_L' \\ m_i' \neq m \\ 1 \leqq i \leqq L}} \frac{1}{L!} \frac{P(\mathbf{v}_1 \mid \mathbf{x}_{m_1'})^{1/1+\rho} \cdots P(\mathbf{v}_1 \mid \mathbf{x}_{m_L'})^{1/1+\rho}}{P(v_1 \mid \mathbf{x}_m)^{1/1+\rho} \cdots P(\mathbf{v}_1 \mid \mathbf{x}_m)^{1/1+\rho}} \right\}^{\rho/L}$$

where $\rho \geqq 0$. Carrying out the random code arguments as given in Ref. 10 with $0 \leqq \rho \leqq L$, we have the desired result.

REFERENCES

1. Shannon, C. E., Zero Error Capacity of a Noisy Channel, IRE Trans, *IT-2,* September, 1956, pp. 8–19.
2. Shannon, C. E., unpublished seminar notes numbered 1p–3p, 1956.
3. Berlekamp, E. R., Block Coding with Noiseless Feedback, M.I.T. Thesis, August, 1964.
4. Horstein, M., Sequential Transmission Using Noiseless Feedback, IRE Trans, *IT-9,* 1963, pp. 136–143.
5. Weldon, E. J., Asymptotic Error Coding Bounds for the Binary Symmetric Channel with Feedback, Ph.D. Thesis, Dept. of Electrical Engineering, University of Florida, April, 1963.
6. Chang, S. S. L., Theory of Information Feedback Systems, IRE Trans, *IT-11,* 1956, pp. 29–40.
7. Fano, R. M., *Transmission of Information,* M.I.T. Press and John Wiley and Sons, New York, 1961.
8. Wozencraft, J. M., List Decoding, Quarterly Progress Report No. 18, Research Laboratory of Electronics, M.I.T., January 15, 1958.
9. Berlekamp, E. R., private communication.
10. Gallager, R. G., A Simple Derivation of the Coding Theorem and Some Applications, IEEE Trans, *IT-11,* January, 1965, pp. 3–17.
11. Wyner, A. D., An Improved Error Bound for Gaussian Channels, B.S.T.J., *63,* November, 1964, pp. 3070–3074.
12. Shannon, C. E., Gallager, R. G., and Berlekamp, E. R., Lower Bounds to Error Probability for Coding on Discrete Memoryless Channels, submitted to Information and Control.

13. Elias, P., List Decoding for Noisy Channels, Technical Report No. 335, Research Laboratory of Electronics, M.I.T., September 20, 1957.
14. Elias, P., Coding for Two Noisy Channels, *Information Theory* (C. Cherry, Editor), Butterworth, London, 1956.
15. Shannon, C. E., Probability of Error for Optimal Codes in a Gaussian Channel, B.S.T.J., *38*, May, 1959, pp. 611–656.
16. Ebert, P. M., Error Bounds for Parallel Communication Channels, M.I.T. Thesis, September, 1965.
17. Gallager, R. G., unpublished notes.