3

CONSTRUCTION OF CONVOLUTION CODES BY SUBOPTIMIZATION

MARVIN A. EPSTEIN

TECHNICAL REPORT 341

NOVEMBER 18, 1959

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
CAMBRIDGE, MASSACHUSETTS

The Research Laboratory of Electronics is an interdepartmental laboratory of the Department of Electrical Engineering and the Department of Physics.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the U. S. Army (Signal Corps), the U. S. Navy (Office of Naval Research), and the U.S. Air Force (Office of Scientific Research, Air Research and Development Command), under Signal Corps Contract DA36-039-sc-78108, Department of the Army Task 3-99-20-001 and Project 3-99-00-000.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 341 November 18, 1959

CONSTRUCTION OF CONVOLUTION CODES BY SUBOPTIMIZATION

Marvin A. Epstein

This report is based on a thesis submitted to the Department of Electrical Engineering, M. I. T., September 1958, in partial fulfillment of the requirements for the degree of Doctor of Science.

Abstract

A procedure for suboptimizing the choice of convolution codes is described. It is known that random convolution codes that have been passed through a binary symmetric channel, or a binary erasure channel, have a low probability of error and are easily decoded, but no practical procedure for finding the optimum convolution code for long code lengths is known. A convolution code is defined by its generator. It is proved that by sequentially choosing the generator digits, one can obtain a code whose probability of error decreases as fast as, or faster than, the usual upper bound for a random code. Little effort is required for suboptimizing the choice of the first few generator digits. This effort increases exponentially with the choice of successive generator digits. For a rate of transmission equal to 1/2, and the given procedure, a code of length 50 is the approximate limit, with the use of the digital computers that are available at present.

I. INTRODUCTION

It has been proved (1-3) that information can be transmitted through a noisy memoryless channel at any rate below capacity with an arbitrarily small error. These proofs also show that the probability of error, with optimum coding, decreases exponentially with code length. There are, however, no known practical procedures for finding such codes for arbitrary rates and arbitrary channels if the code length is large.

Golay (4), Slepian (5), and Fontaine and Peterson (6) have found specific optimum codes for the binary symmetric channel, but these codes are restricted to short code lengths. Hamming (7) and Reed (8) have found certain sets of codes of arbitrary length for the binary symmetric channel, but these codes do not provide for the transmission of information at a fixed nonzero rate when the probability of error must decrease indefinitely. Elias (9) has described an iterative code that will transmit at a positive rate with arbitrarily low probability of error. For such codes, however, the probability of error decreases much too slowly with increasing length. Thus our knowledge of specific good codes is very limited.

Elias (10) has shown that in a binary symmetric channel, or in a binary erasure channel, the probability of error averaged over all possible codes with a given length decreases exponentially with increasing code length. Hence, one can choose a long code at random and it is most likely to have a low probability of error.

Another possible approach to the problem of choosing a code, is to examine all the possible codes of a given length and find the code with the lowest probability of error. This approach will give specific codes, but for long codes more computation will be required than can be carried out by present or foreseeable computers. Slepian (5), and Fontaine and Peterson (6), have used this approach, but their exhaustive searches stopped at lengths 12 and 15, respectively.

In this report we treat an approach somewhere between these two extremes. We restrict the discussion to convolution codes and to either the binary symmetric channel or the binary erasure channel. A convolution code of length n can be described by a generator containing less than n binary digits. In the procedure that will be described here, the first few digits of the convolution-code generator are chosen sequentially to minimize the probability of error, and the remaining generator digits are chosen at random. This procedure results in good codes obtained with comparatively little effort.

We show that at rates of transmission of information that cannot approach capacity, the codes chosen by this method have a probability of error that decreases with the same exponent as Elias' bounds on a random code. In the example that will be given, the choice of the first 5 generator digits by the suboptimization procedure results in a probability of error that is 0.003 of the usual bound on the

probability of error. The choice of the first few generator digits requires little effort, and the effort required by the choice of succeeding digits increases exponentially with increasing length. For a code with a rate of transmission equal to 1/2, the limit for manual computation is approximately length 20, and the limit for present digital computers is approximately length 50.

II. CONVOLUTION CODES

We shall treat two channels: the binary symmetric channel, and the binary erasure channel. The binary symmetric channel has two input symbols, 0, 1, and the same two output symbols. If a zero or a one is transmitted, the probability that the same symbol is received is q, and the probability that the other symbol is received is p = 1−q (see Fig. 1a). The binary erasure channel has two input symbols, 0, 1, and three output symbols, 0, X, 1. If a zero or a one is transmitted, the probability that the same symbol is received is q, and the probability that an X is received is p = 1−q. There is zero probability that a zero is sent and a one received, or that a one is sent and a zero received (see Fig. 1b). For both channels, the output at a given time depends only upon the corresponding input, and not on other inputs or outputs. The capacity (10) of the binary symmetric channel is 1+p log p+q log q, and the capacity of the binary erasure channel is q.

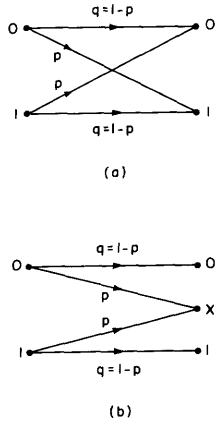

Fig. 1. Transition probabilities for: (a) the binary symmetric channel; and (b) the binary erasure channel.

We are interested in choosing the first few generator digits of a convolution code. In a convolution code the information digits are an infinite sequence of digits, and check digits are interleaved with the information digits in some alternating pattern, such as information digit, check digit, information digit, check digit, and so on. Each check digit in a convolution code of length n is determined by a parity check on a subset of the preceding n−1 digits. For rates of $R \geq 1/2$, each check digit checks a fixed pattern of the preceding information digits. For rates of $R < 1/2$, each information digit is checked by a fixed pattern of the succeeding check digits (see Fig. 2). This fixed pattern is called the generator. The value of each check digit is determined by the modulo 2 sums and matrix B of Eq. 1. Here each element of the matrix is either a one or a zero. The j^{th} transmitted digit is indicated by d_j, and the position of the i^{th} check digit by p(i). A modulo 2 sum is zero if the usual sum is even, and one if the usual sum is odd. Thus

$$\sum_{j=p(i)-n+1}^{p(i)} b_{ij}\, d_j = 0 \pmod 2 \qquad (1)$$

No more than n digits are needed for describing the parity-check matrix because one fixed pattern is used for describing the relation of the information digits to the check digits.

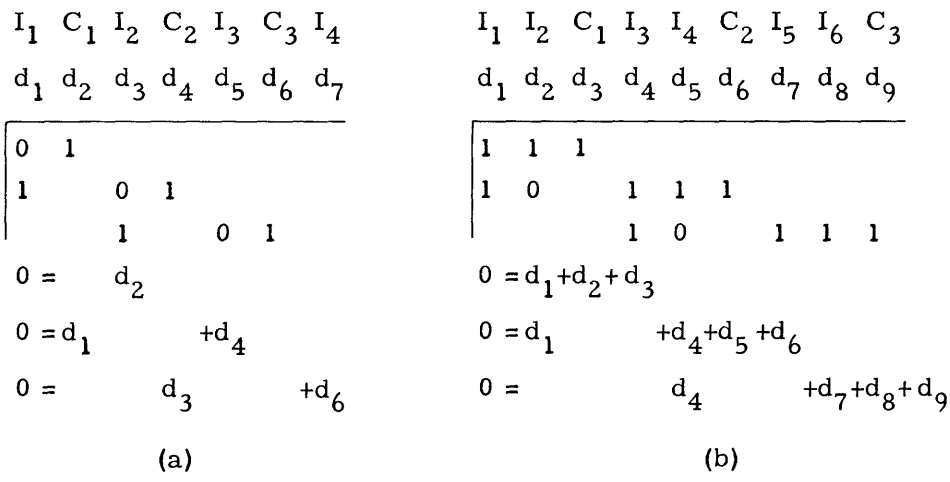

Fig. 2. Coding matrices and coding equations. All equations are modulo 2. (a) Convolution code, length = 4, rate = 1/2, generator = 10. (b) Convolution code, length = 6, rate = 2/3, generator = 1011.

If the information digits take on the values zero and one with probability 1/2 independently of the previous digits, each information digit contains 1 bit of information, and the rate of transmission of information, R, equals the fraction of the digits that are information digits. Elias (10) has found bounds on the probability of error per digit for a convolution code chosen at random from the set of all convolution codes of length n. These bounds are of the form $Kn^{-y} X^{-n}$, where K, y, and X are constants determined by the rate R and the channel capacity C. These bounds are given in Appendix A.

III. SUBOPTIMIZATION PROCEDURE

A convolution code is determined by the generator that describes the information digits checked by each check. The suboptimization procedure considered here is a procedure for sequentially choosing the first few digits of the generator, in order to obtain a code with a low probability of error. The procedure involves a number of steps. In each step we determine the value of one generator digit.

In the first step, we choose the value of the first generator digit in the following manner: We evaluate the set of codes whose first generator digit is a zero, and the set of codes whose first generator digit is a one. We then choose the first generator digit to have the value that led to the better set of codes as evaluated by our evaluation procedure.

In the second step, we choose the value of the second generator digit. (The value of the first digit remains unchanged.) We evaluate the set of codes whose second generator digit is equal to zero, and the set of codes whose second generator digit is equal to 1. We choose the second generator digit to have the value that led to the better set of codes as thus evaluated.

We continue in this fashion, determining one generator digit in each step. Eventually, after a given number of generator digits are chosen, the procedure stops, and the remaining generator digits, if there are any, are filled in at random.

IV. EVALUATION METHOD

The heart of the suboptimization procedure is the method of simply and accurately evaluating the effect of fixing the first few generator digits. The evaluation method that is described below is basically a method of accurately bounding the average probability of error of the set of convolution codes with a given first few generator digits. This method is based on a simpler, but less rigorous, evaluation method developed by Wozencraft (11), which uses only the messages with the smallest or the next smallest number of ones.

To simplify the explanation, we first describe the evaluation method for a convolution code of rate 1/2, in which the first m generator digits are given, and in which information and check digits alternate. Later, the evaluation method will be generalized to convolution codes with periodic patterns of information and check digits.

Since convolution codes are group codes, the probability of error is the same for all messages. Thus we can find the probability of error (hereafter denoted by P_e) for the code by evaluating P_e when a standard message – for example, the zero message – is sent. In a convolution code with rate 1/2, P_e for each digit, when previous digits are known, is identical. Hence, we can evaluate a convolution code by considering P_e of digit 1 when the zero message is sent. If the zero message is sent, and the first 2m received digits are decoded optimally, digit 1 will be decoded incorrectly only in the following case. One of the 2^{m-1} messages whose first digit is a one differs from the received message in as few digits or fewer digits than the zero message. Hence, P_e of digit 1 is bounded by the sum over all the 2^{m-1} messages of the probability of each message being as close, or closer, to the received message as the zero message is. That is,

$$P_e \text{ for digit 1 when the zero message is sent} \leq \sum_{i=1}^{2^{m-1}} \begin{pmatrix} \text{probability that message } i \\ \text{is as close, or closer, to} \\ \text{the received message than} \\ \text{the zero message is} \end{pmatrix} \tag{2}$$

If a given message has d ones and the channel is a binary symmetric channel, the probability that the given message will be as probable as the zero message when the zero message is sent, is the probability that d/2, or more, of the positions in which the given message has ones are received incorrectly. This probability is

$$\sum_{j=\frac{d+1}{2}}^{d} \binom{d}{j} p^j q^{d-j}, \qquad \text{with d odd}$$

or

$$\sum_{j=\frac{d}{2}}^{d} \binom{d}{j} p^j q^{d-j}, \qquad \text{with d even}$$

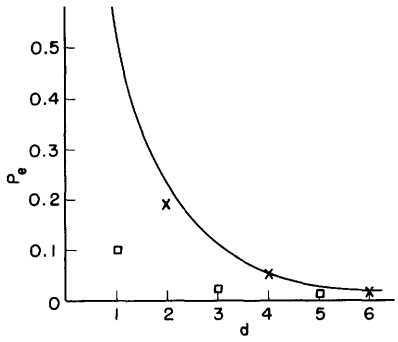

Fig. 3. Error bounds: $\sum_{j=(d+1)/2}^{d} \binom{d}{j} p^j q^{d-j}$ represented by □; $\sum_{j=d/2}^{d} \binom{d}{j} p^j q^{d-j}$ represented by ×; $(2\pi/d)^{1/2} (4pq)^{d/2} [1-(p/q)]^{-1}$ represented by the curve.

As Fig. 3 shows, P_e as calculated by these formulas does not lie on a smooth curve. Hence, in the following discussion we shall bound P_e for a message with d ones by the smooth formula

$$\frac{(2/\pi d)^{1/2} (4pq)^{d/2}}{(1-p/q)}$$

which bounds both of the formulas that have just been stated and is asymptotic to the formula for d even. The derivation requires Stirling's formula and some algebraic manipulation.

If a given message has d ones and the channel is a binary erasure channel, the probability that the given message appear as probable as the zero message, when the zero message is sent, is p^d. This is true, since the given message will have a nonzero a posteriori probability, if and only if all the d positions in which the given message has ones are erased.

Using Eq. 2 and the formulas just described, we can bound P_e for digit 1 when the zero message is sent by

$$P_e \text{ for digit 1 in a binary symmetric channel} = \sum_{d=0}^{2m} \frac{w_d (2/\pi d)^{1/2} (4pq)^{d/2}}{(1-p/q)} \qquad (3a)$$

$$P_e \text{ for digit 1 in a binary erasure channel} = \sum_{d=0}^{2m} w_d p^d \qquad (3b)$$

where w_d is the number of messages of length 2m that have d ones and whose first digit is a one. The w_d can be found by listing all 2^{m-1} messages that have a given number of ones.

Since a convolution code with a rate 1/2 can be characterized by P_e for digit 1 when the zero message is transmitted, Eqs. 3 permit the evaluation of convolution codes.

This is the evaluation used in our procedure.

The evaluation is similar for convolution codes that have a different periodic pattern of information and check digits (see Fig. 4c). In such codes, P_e for the zero message is, as before, identical with that for any other message. Also, when the previous digits are known, P_e is the same for all digits in corresponding positions in the periodic pattern. Thus these codes can be evaluated by means of bounds on P_e of a representative digit in each of the positions of the periodic pattern. The bound on each digit position can be found by means of Eq. 3, when w_d is defined as follows: w_d equals the number of messages that have zeros preceding the given digit, a one in the given digit, and d ones in the digits up to, and including, the last check digit determined by the given generator in conjunction with the given digit.

V. EXAMPLE OF THE SUBOPTIMIZATION PROCEDURE

We illustrate the suboptimization procedure with the following example. It is assumed that the channel is a binary erasure channel, that the rate of transmission is equal to 1/2, and that $p = 0.1$; $q = 0.9$.

In the first step, there is just one message whose first digit is a one.

Assuming that first generator digit is a zero, we have	Assuming that first generator digit is a one, we have
$w_o = w_2 = 0$; $w_1 = 1$	$w_o = w_1 = 0$; $w_2 = 1$
$P_e \leq .1$	$P_e \leq .01$

We choose the first generator digit to be a one.

Assuming the second generator digit is a zero, we have	Assuming the second generator digit is a one, we have
$w_2 = w_4 = 1$; $w_3 = 0$	$w_2 = w_4 = 0$; $w_3 = 2$
$P_e \leq .0101$	$P_e \leq .002$

We choose the second generator digit to be a one.

Assuming the third generator digit is a zero, we have	Assuming the third generator digit is a one, we have
$w_3 = w_5 = 1$; $w_4 = 2$	$w_3 = w_5 = 1$; $w_4 = 2$
$P_e \leq .00121$	$P_e \leq .00121$

Since the bounds are the same for both codes, it does not matter which one is chosen. We shall assume that the third generator digit is a zero.

Assuming the fourth generator digit is a zero, we have	Assuming the fourth generator digit is a one, we have
$w_3 = w_4 = 1$; $w_5 = w_6 = 3$	$w_3 = 0$; $w_4 = w_5 = 3$; $w_6 = w_7 = 1$
$P_e \leq .001133$	$P_e \leq .0003311$

We choose the fourth generator digit to be a one. The probability bound for a random code is 0.13, as compared with 0.00033 for the chosen code. This completes the example.

VI. EXPONENTIAL DECREASE OF ERROR FOR SUBOPTIMIZATION CODES

We shall now prove that for rates less than the critical rate (see Appendix A), codes generated by the suboptimization procedure have a P_e that decreases with the same exponent as Elias' bounds for a random code. In Elias' bounds, for rates less than the critical rate and with $R = 1/2$, the bound for the binary symmetric channel is multiplied by the factor $1/2\ [1+(4pq)^{1/2}]^2$, and that for the binary erasure channel by the factor $1/2\ (1+p)^2$ for every increase of two digits in length.

First, we show that if a generator digit is added at random to any code with $R = 1/2$, then the bounds of Eqs. 3a and 3b are multiplied by a factor as small as, or smaller than, these factors. Next, we show that in any step of the suboptimization procedure, the bounds of Eqs. 3 decrease by a factor that is as small as, or smaller than, the factor for a random code. These proofs can be extended in a natural fashion to rates that are not 1/2. Thus codes generated by the suboptimization procedure (or codes with some digits generated by our procedure, and some chosen at random) have a P_e that decreases with the same exponent as Elias' bounds, or more rapidly. Moreover, the initial values of Eqs. 3 are smaller than the initial values of Elias' bounds for all cases in the binary erasure channel, and for many cases in the binary symmetric channel. For such cases, the bounds of Eqs. 3 for a random code are smaller than Elias' bounds.

By investigating what happens to a particular message of length 2m, whose first digit is a one, when the $m+1^{th}$ generator digit is chosen at random, we can prove that the bounds for a code with $R = 1/2$ decrease by an appropriate factor with each additional generator digit chosen at random. In step m+1, a particular message gives rise to four messages of length 2m+2 because there are two choices for the $m+1^{th}$ generator digit, each with probability 1/2, and two choices for the $m+1^{th}$ information digit. (In any given code the generator digit has a given value and only two of these messages appear.) Since the first information digit of the message is a one, the value of the $m+1^{th}$ check digit can be expressed in terms of the coding matrix b_{ij} as follows:

$$d_{p(m+1)} = b_{m+1,\,1} + \sum_{j=2}^{p(m+1)} b_{m+1,\,j}\, d_j \qquad (4)$$

where $b_{m+1,\,1}$ is the $m+1^{th}$ generator digit, and p(m+1) is the position of the $m+1^{th}$ check digit. Thus, if the $m+1^{th}$ check digit has the value zero for a given choice of the $m+1^{th}$ generator digit and for the $m+1^{th}$ information digit equal to zero (one), the $m+1^{th}$ check digit will have the value one for the other choice of the $m+1^{th}$ generator digit and for the $m+1^{th}$ information digit equal to zero (one). Therefore, a given message of length 2m and weight d gives rise to four messages of length 2m+2: (a) a message of weight d that is composed of the old message and two new digits that are both zero; (b) a message of weight d+1 that is composed of the old message, a new information digit equal to zero, and a new check digit equal to one; (c) a message of

I_1	C_1	I_2	C_2	I_3	C_3	I_4	C_4
d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
1	1						
1		1	1				
0		1		1	1		
?		0		1		1	1

(a)

d_1	d_2	d_3	d_4	d_5	d_6
1	1	0	1	0	0

(b)

d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
1	1	0	1	0	0	0	0
1	1	0	1	0	0	1	1

(c)

d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8
1	1	0	1	0	0	0	1
1	1	0	1	0	0	1	0

(d)

Fig. 4. (a) Convolution matrix, rate = 1/2, generator of length 3 = 011. (b) An original message corresponding to the matrix of (a), length = 6. (c) New messages if $b_{4,1} = 0$. (d) New messages if $b_{4,1} = 1$.

weight d+1 that is composed of the old message, a new information digit equal to one, and a new check digit equal to zero; and (d) a new message of weight d+2 that is composed of the old message and two new digits that are both equal to 1 (see Fig. 4).

Each of the new messages occurs with 1/2 the probability of the new message because generator digits are selected at random. Thus we find that the contribution to the bound of Eq. 3a for the four new messages, if the channel is a binary symmetric channel, is

$$1/2\left(\begin{matrix}\text{probability}\\ \text{of original}\\ \text{message}\end{matrix}\right)\left(\frac{2}{\pi}\right)^{1/2}(1-p/q)^{-1}\left(\frac{(4pq)^{d/2}}{\sqrt{d}}+\frac{2(4pq)^{(d+1)/2}}{(d+1)^{1/2}}+\frac{(4pq)^{(d+2)/2}}{(d+2)^{1/2}}\right)$$

$$<\left(\begin{matrix}\text{bound on the}\\ \text{old message}\end{matrix}\right)1/2\left[1+(4pq)^{1/2}\right]^2 \qquad (5)$$

If the channel is a binary erasure channel, the contribution to the bound of Eq. 3b for the four new messages is

$$1/2\left(\begin{matrix}\text{probability of the}\\ \text{old message}\end{matrix}\right)(p^d+2p^{d+1}+p^{d+2}) = \left(\begin{matrix}\text{bound on the}\\ \text{old message}\end{matrix}\right)1/2\,(1+p)^2 \qquad (6)$$

If a new generator digit is selected at random for a code (set of codes) of given length, each message of the code (set of codes) will generate four messages for which the bound is less than, or equal to, the appropriate factor multiplied by the bound on the original message. Thus the bound for the new code, which is the sum of the bounds for each

message, is less than, or equal to, the appropriate factor multiplied by the bound on the original code.

This proof can be extended in a natural fashion to codes whose rate is not 1/2. The correct factor per digit is $2^{-1+R}[1+(4pq)^{1/2}]$ for the binary symmetric channel, and $2^{-1+R}(1+p)$ for the binary erasure channel. The factors $[1+(4pq)^{1/2}]$ and $1+p$ result from the fact that, with each added digit, each message of weight d gives rise to two messages, one of the same weight, and one of weight d+1. The factor 2^{-1+R} results from the fact that every time a check digit occurs, it is equal to zero, or 1, with probability 1/2, and from the fact that check digits occur a fraction of the time, which is equal to (1-R).

At every step the suboptimization procedure decreases the bounds of Eqs. 3 as fast as, or faster than, choosing a generator digit at random. If a generator digit is chosen at random, the bound is the average of the bounds for the two possible generator choices. On the other hand, the suboptimization procedure chooses the lower bound for the generator choices. Therefore the suboptimization procedure does as well as, or better than, the average procedure. This completes the proof that P_e for codes generated by the suboptimization procedure decreases at the appropriate rate.

VII. APPLICATION OF THE SUBOPTIMIZATION PROCEDURE

The most useful application of the suboptimization procedure is in the generation of convolution codes that can be decoded with little effort. Wozencraft (12) and the author (13, 14) have examined decoding methods that require little effort. The average amount of computation for these decoding methods is a very strong function of the choice of the first few generator digits. Hence, this suboptimization procedure is useful in choosing codes for these decoding methods because it carefully chooses the first few generator digits of a convolution code. The usefulness of the suboptimization procedure is illustrated by the following example. Using the usual bounds on P_e, the author (14) found that 200 was a bound on the average number of binary additions for decoding a convolution code of $R = 1/2$ that was transmitted through a binary erasure channel with $p = 0.1$. In contrast, it was found (15), by using the better bound on P_e for a random code developed in our proof, that the bound for the same situation is 7.78; and that the bound for a code whose generator digits are chosen at random, except for the first four generator digits which are those of our example, is 0.36.

The suboptimization procedure can also be used to generate convolution codes whose P_e per digit is known to be small. Moreover, in evaluating P_e for this procedure we have also proved new bounds on P_e per digit for convolution codes. The bounds for a code of length n are $(2/\pi)^{1/2}(1-p/q)[(1+(4pq)^{1/2})2^{-1+R}]^n$ and $[(1+p)2^{-1+R}]^n$ for the binary symmetric channel and the binary erasure channel, respectively.

There is a rather sharp practical limit to the number of generator digits that can be chosen by this procedure. The number of messages listed, and hence the effort per generator digit, increases exponentially with code length. A list of 1000 messages is the approximate limit for manual computation, and a list of 10^6 messages is the approximate limit for present computer computation. Thus, for a rate of 1/2, with the procedure described above, the limit for manual computation is approximately length 20, and for present digital computers, approximately length 40. However, the limit for digital computers can be extended to 50 digits, or more, by eliminating the messages with the largest number of ones when the computational capacity is overtaxed, and considering only the descendants of the messages with the least number of ones (and hence, the largest P_e). This modified suboptimization procedure will have to stop as soon as P_e for the eliminated messages is larger than P_e of the messages that are still being considered.

VIII. SUMMARY

A procedure for choosing a convolution code has been described. This procedure chooses generator digits sequentially and requires little effort, while the probability of error decreases as fast as, or faster than, the bound on a random code. There is a practical limit to the length of a code that can be described by this method, since the effort per digit increases exponentially.

A number of new problems is raised by this procedure. In the example that has been described the probability of error decreases much faster than the usual bounds. This raises the question as to whether this method, in the limit, has an exponential decrease of error that is better than the random bound. This question is especially relevant to probabilities of error close to zero, since it is known (10) that optimum codes behave in this manner. Other problem areas are the description of better sub-optimization procedures and the generalization of these procedures to other codes and channels. A different evaluation procedure which applies to any rate below capacity, including rates above R_{crit}, is described in Appendix B.

APPENDIX A

BOUNDS ON THE PROBABILITY OF ERROR

Elias (6, 10, 12-14) has found bounds on the probability of error per digit for a convolution code chosen at random from the set of all convolution codes of length n. Simplified forms of these bounds, with minor modifications, are given in Eqs. A-2 and A-3. In each formula, K is a constant for fixed rate R and channel capacity C. These bounds, in certain cases, may be improved at the cost of increased complexity by returning to Elias' original bounds.

R_{crit} for the binary symmetric channel is defined as the capacity of a binary symmetric channel whose probability of incorrect reception p_{crit} is $\sqrt{p}/(\sqrt{p}+\sqrt{q})$. We also define an auxiliary quantity p_1 in terms of the rate R:

$$1 + p_1 \log_2 (p_1) + q_1 \log_2 (q_1) = R, \qquad p \leq 1/2 \tag{A-1}$$

$$\left.\begin{aligned} &\text{The probability of error per digit for a binary symmetric channel} \leq K n^{-y} e^{-\ln(X)} = K n^{-y} X^{-n} \\ &X = \left(\frac{p_1}{p}\right)^{p_1} \left(\frac{q_1}{q}\right)^{q_1}; \; y = \frac{1}{2} \quad \text{for } R_{crit} < R \\ &X = \frac{2^{1-R}}{1 + \sqrt{4pq}}; \; y = 0 \quad \text{for } R_{crit} \geq R \end{aligned}\right\} \tag{A-2}$$

$$\left.\begin{aligned} &\text{The probability of error per digit for a binary erasure channel} = K n^{-y} e^{-\ln(X)} = K n^{-y} X^{-n} \\ &X = \left(\frac{R}{C}\right)^{R} \left(\frac{1-R}{1-C}\right)^{1-R}; \; y = \frac{1}{2} \quad \text{for } q/(1+p) = R_{crit} < R \\ &X = \frac{2^{1-R}}{1+p}; \; y = 0; \; K = \frac{1+p}{\sqrt{\pi}} + \frac{2}{(1+p)\sqrt{\pi}} + \frac{1+p}{2(\pi pq)^{1/2}} \quad \text{for } R_{crit} \geq R \end{aligned}\right\} \tag{A-3}$$

APPENDIX B

ANOTHER EVALUATION METHOD

The evaluation method that will now be described has the correct exponential decrease in probability of decoding error for rates above and below the critical rate. (This has been proved when the convolution code has rate 1/2 and the channel is a binary erasure channel. It is not known whether the result holds for codes with different rates, or for the binary symmetric channel.) This is contrary to the method described in Section IV, which has the correct decrease only for rates below the critical rate. The new procedure is more complicated, and more restricted, and greater effort is required for choosing each generator digit.

Before we can understand this new evaluation method, we need to describe the sequential decoding procedure for the binary erasure channel (13, 14). The sequential decoding procedure for convolution codes decodes digit 1 in a number of steps. In the m^{th} step, an attempt is made to decode digit 1 by means of the first m parity-check equations. The procedure ends when digit 1 is decoded, or when there are no more equations that check digit 1. At the end of step m, there is a certain number, r, of erased digits in the first m equations, and a certain number, r-k, of independent equations in the erased digits.

We describe the situation at the end of step m by means of states, the state assigned to a given situation being a function of whether or not digit 1 is decoded, and of k (that is, the number of erased digits minus the number of independent equations). State S_0 characterizes all the situations in which digit 1 is decoded. States S_1, S_2, S_3, ... characterize the situations in which digit 1 is not decoded. The subscript for these last states is equal to the number of erased digits minus the number of independent equations. Note that each possible situation corresponds to one, and only one, state. To summarize,

$$\left.\begin{aligned} S_0 &= \text{the state if digit 1 is decoded} \\ S_k &= \text{the state if digit 1 is not decoded and the number of erased digits minus the number of independent equations equals k, for } k = 1, 2, 3, \ldots \end{aligned}\right\} \qquad \text{(B-1)}$$

At the end of step m for a given code and a given erasure pattern, the decoding procedure is in a given state. (See Fig. 5.)

The evaluation method for the first m digits of a generator sequence is as follows: Given an erasure pattern for the first 2m digits, we can find the equations for the decoding procedure, and we can find the state at the end of step m. Hence, by determining the state at the end of step m for each of the 2^{2m} possible erasure patterns, and by using the probability of occurrence of the erasure patterns, we can find the probabilities $p_m(k)$, defined below.

$$p_m(k) = \begin{array}{l}\text{probability of being}\\ \text{in state } k \text{ at the end}\\ \text{of step } m\end{array} = \sum_{\substack{\text{erasure patterns}\\ \text{that give state } k}} \begin{array}{l}\text{probability of the}\\ \text{erasure pattern}\end{array} \qquad \text{(B-2)}$$

Furthermore, we define $A_{m,n}(k)$ to be the number that bounds the probability that we are not in state S_o at the end of step n, given that we are in state k at the end of step m and that generator digits m+1, m+2, . . . , n are chosen at random. Using

I C I C		
1 1	Erasure patterns that give state S_1: XX00; XX0X; XXX0.	$p_2(0) = 0.99$
0 1 1	Erasure patterns that give state S_2: XXXX.	$p_2(1) = 0.0099$
	All other patterns of length 4 give state S_o.	$p_2(2) = 0.0001$
(a)	(b)	(c)

Fig. 5. Evaluation of $p_m(k)$. (a) Convolution coding matrix. (b) Erasure patterns: erasures are indicated by X, transmission by 0. (c) Values of $p_2(k)$ for $p = 0.1$, $q = 0.9$.

the distribution of states at the end of step m and the $A_{m,n}(k)$, we can bound the probability of ambiguity for a code whose first m generator digits are the given set, and whose remaining generator digits are chosen at random. This leads to

$$\begin{array}{l}\text{The probability of ambiguity for a convolution code of length n whose first m generator digits are the given set, and whose remaining generator digits are chosen at random}\end{array} \leq \sum_k p_m(k)\, A_{m,n}(k) \qquad \text{(B-3)}$$

The $A_{m,n}(k)$ for codes of rate 1/2 are determined by the following equations.

$$A_{m,n}(0) = 0 \qquad m = 1, 2, 3, \ldots, n \qquad \text{(B-4a)}$$

$$A_{n,n}(k) = 1 \qquad k = 1, 2, 3, \ldots \qquad \text{(B-4b)}$$

$$A_{m,n}(1) = (2pq + (1/2)q^2)\, A_{m+1,n}(1) + p^2\, A_{m+1,n}(2) \qquad \text{(B-5a)}$$

$$A_{m,n}(k) = q^2\, A_{m+1,n}(k-1) + 2pq\, A_{m+1,n}(k) + p^2\, A_{m+1,n}(k+1) \qquad \text{for } k = 2, 3, \ldots \qquad \text{(B-5b)}$$

Equations B-4 determine the A's for m equal to n. Equations B-5 are then used to calculate the A's for m equal to n-1, n-2, and so on. A table of $A_{m,n}(k)$ for $n = 5$ is shown in Fig. 6.

The bound of Eq. B-3, if the values of $A_{m,n}(k)$ determined above are used, is the

m \ k	0	1	2	3	4	5
1	0	.131				
2	0	.216	.421			
3	0	.358	.672	1		
4	0	.595	1	1	1	
5	0	1	1	1	1	1

Fig. 6. Table of $A_{m,5}(k)$; $p = 0.1$, $q = 0.9$.

new evaluation. It has been proved (15) that the use of this evaluation in the suboptimization procedure will result in the correct exponential decrease for a code of rate 1/2. The exponent is correct, whether or not the channel capacity is such that rate 1/2 is below the critical rate. This evaluation method can be extended to codes whose rate is not 1/2. However, it has not been proved that this method, when it is extended to rates other than 1/2, has the correct exponent.

Acknowledgment

The research reported here was carried out under Professor P. Elias and Professor R. M. Fano, of the Department of Electrical Engineering, M. I. T. The author wishes to acknowledge the assistance of members of the Research Laboratory of Electronics and Lincoln Laboratory, M. I. T. In particular, he is indebted to Professor J. M. Wozencraft for his many helpful comments on the suboptimization procedure.

References

1. C. E. Shannon, The Mathematical Theory of Communication (University of Illinois Press, Urbana, Illinois, 1949).
2. A. Feinstein, A new basic theorem in information theory, Trans. IRE, PGIT-4, pp. 2-22 (September 1954).
3. C. E. Shannon, Certain results in coding theory for noisy channels, Information and Control 1, 6-25 (September 1957).
4. M. J. E. Golay, Notes on digital coding, Proc. IRE 37, 657 (1949).
5. D. Slepian, A class of binary signalling alphabets, Bell System Tech. J. 35, 203-234 (1956).
6. A. B. Fontaine and W. W. Peterson, On coding for the binary symmetric channel, Communication and Electronics 39, 638-647 (1958).
7. R. W. Hamming, Error detecting and error correcting codes, Bell System Tech. J. 29, 147-160 (1950).
8. I. S. Reed, A class of multiple error-correcting codes and the decoding scheme, Trans. IRE, PGIT-4, pp. 38-49 (September 1954).
9. P. Elias, Error-free coding, Trans. IRE, PGIT-4, pp. 29-37 (September 1954).
10. P. Elias, Coding for two noisy channels, Information Theory, edited by C. Cherry (Butterworths Scientific Publications, London, 1956).
11. J. M. Wozencraft, private communication, Massachusetts Institute of Technology, January 1958.
12. J. M. Wozencraft, Sequential decoding for reliable communication, Technical Report 325, Research Laboratory of Electronics, M. I. T., Aug. 9, 1957.
13. M. A. Epstein, Coding for the binary erasure channel, Sc. D. Thesis, Department of Electrical Engineering, M. I. T., September 1958.
14. M. A. Epstein, Algebraic decoding for a binary erasure channel, IRE National Convention Record, Part 4, pp. 56-69 (1958).
15. M. A. Epstein, Sc. D. Thesis, op. cit., see Appendix IV.