9

# CODING FOR TWO NOISY CHANNELS*

PETER ELIAS

*Department of Electrical Engineering and Research Laboratory of Electronics,*
*Massachusetts Institute of Technology, Cambridge, Massachusetts*

## INTRODUCTION

SHANNON's original demonstration[1,2] that information could be transmitted over a noisy channel at a positive rate with an arbitrarily small probability of error at the receiver suggested that there was a relationship between permissible delay or block size, transmission rate, channel capacity and error probability. RICE[3], investigating a special case, got an indication of exponential decrease of error probability with increasing coding delay. FEINSTEIN[4] showed an exponential decrease in an upper bound to error probability in a more general case, for transmission rates near to channel capacity. The present paper derives much more detailed results for two particular channels with binary input. It is shown that random coding, as initially discussed by SHANNON[1], is substantially as good as anything else for a considerable range of signalling rates near to channel capacity, but that for very small rates appreciably better codes exist.

For the binary symmetric noisy channel, HAMMING[5], GILBERT[6], PLOTKIN[7] and GOLAY[8] have constructed a variety of error-correcting and detecting codes and found some of the basic properties of the channel. LAEMMEL[9], MULLER[10] and REED[11] have also constructed specific codes and classes of codes. The first constructive coding procedure for error-free transmission at a non-zero rate was discovered recently by the author[12]. All of this work has been concerned primarily with systematic, or check-symbol, codes. The question arises whether this convenient restriction reduces the permissible signalling rate or increases the error probability appreciably over what is obtainable with non-systematic coding. It is shown in reference 16 that for the two channels considered here it does not do so. Check-symbol codes are as good as any other kind in terms of both maximum transmission rate and error probability.

Some of the results presented here for the binary symmetric channel were discussed by the author in an earlier paper[13]. The overlap is not complete for this channel, however, and all of the results for the erasure channel are new†.

* This work was supported in part by the Signal Corps, the Office of Scientific Research, Air Research and Development Command, and the Office of Naval Research, of the United States.

† This paper is an abbreviated version of reference 16, which is the full paper as distributed at the Symposium. All proofs are omitted here but will be found in reference 16.

## THE CHANNELS

The coding problems that we shall discuss are illustrated in *Figure 1.* The first problem is to match the output of an ideal binary message source to a binary symmetric noisy channel (BSC). The second problem is to match the output of the same source to a channel which occasionally erases one of the transmitted symbols: this channel will be called a binary erasure channel (BEC).

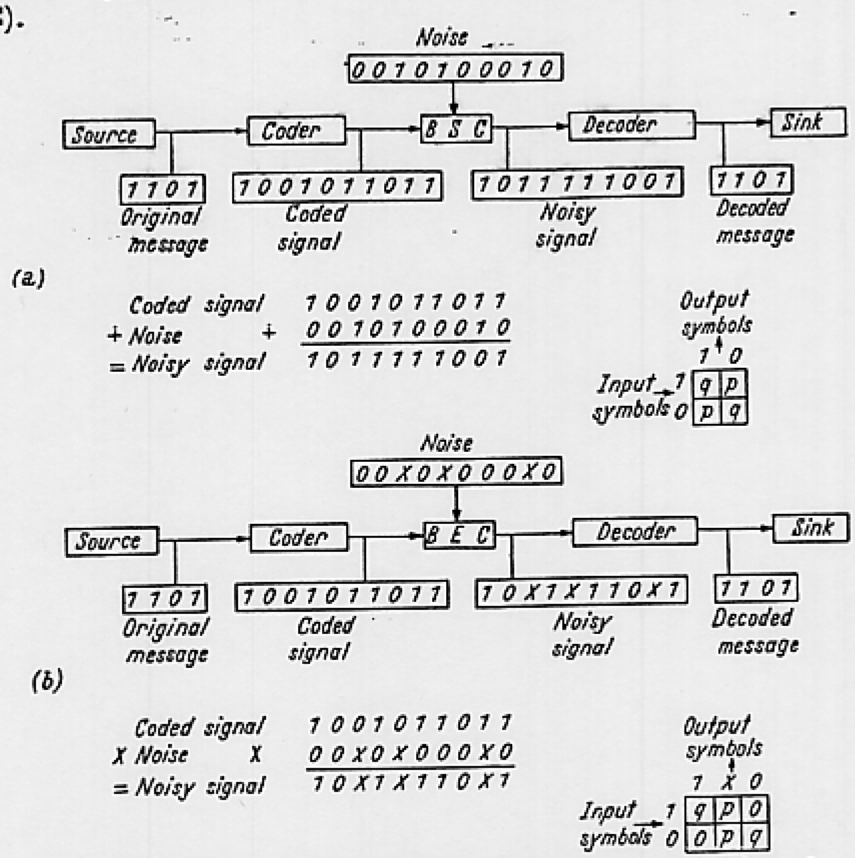


*Figure 1.* (*a*) Binary symmetric channel; (*b*) binary erasure channel.

The message source generates a sequence of zeros and ones. These two symbols are selected with equal probability, and successive selections are statistically independent.

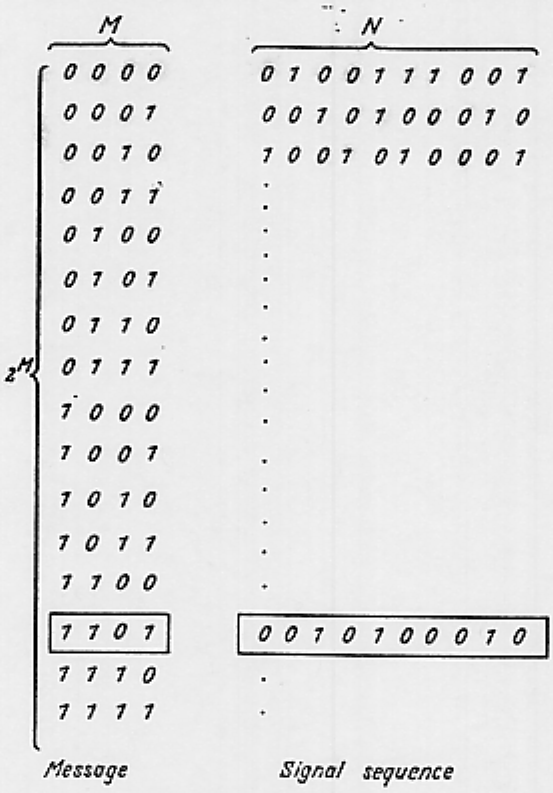
The BSC accepts binary symbols as an input and produces binary symbols as an output. Each input symbol has a probability $p < \frac{1}{2}$ of being received in error, and a probability $q = 1 - p$ of being received as transmitted. The transmission error probability $p$ is a constant, independent of the value of the symbol being transmitted: the channel is as likely to turn a one into a zero as to turn a zero into a one. The channel, in effect, adds a noise sequence to the input sequence to produce the output sequence; the noise is a random sequence of zeros and ones, synchronous with the signal sequence, in which the ones have probability $p$ and the addition is addition modulo two of each signal digit to the corresponding noise digit ($1+1=0+0=0$; $0+1=1+0=1$).

The BEC accepts binary symbols as an input and produces ternary

symbols as
erased, an
erasure pr
An erasur
receiver re

symbol c
about the
If the r
fraction c
procedur
The outp
length $M$
equal pro
output bl
The in
sequences
between
'code' is
output bl
the $2^M$ p
order. F
correspo
The sy
that is co

symbols as an output. Each input symbol has probability $p < 1$ of being erased, and a probability $q = 1 - p$ of being received as transmitted. The erasure probability $p$ is independent of the value of the transmitted symbol. An erasure is indicated to the receiver by the received symbol $X$. If the receiver receives a zero or a one, it knows that it has received the transmitted symbol correctly, but reception of an $X$ gives the receiver no information about the transmitted symbol.



| Message | Signal sequence |
|---|---|
| 0 0 0 0 | 0 1 0 0 1 1 1 0 0 1 |
| 0 0 0 1 | 0 0 1 0 1 0 0 0 1 0 |
| 0 0 1 0 | 1 0 0 1 0 1 0 0 0 1 |
| 0 0 1 1 | . |
| 0 1 0 0 | . |
| 0 1 0 1 | . |
| 0 1 1 0 | . |
| 0 1 1 1 | . |
| 1 0 0 0 | . |
| 1 0 0 1 | . |
| 1 0 1 0 | . |
| 1 0 1 1 | . |
| 1 1 0 0 | . |
| $\boxed{1\ 1\ 0\ 1}$ | $\boxed{0\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 0}$ |
| 1 1 1 0 | . |
| 1 1 1 1 | . |

*Figure 2.* A codebook.

If the message source were connected directly to either of the channels, a fraction of the received symbols would be in error or erased; a coding procedure for reducing the effect of the errors is illustrated in *Figures 1* and *2*. The output of the message source is segmented into consecutive blocks of length $M$. There are $2^M$ such blocks, and they are selected by the source with equal probability. To each input block of $M$ binary symbols is assigned an output block of $N$ binary symbols, $N > M$.

The input sequences of length $M$ are the messages to be sent; the output sequences of length $N$ are the transmitted signals; and the correspondence between input and output blocks is the code used. The use of the word 'code' is justified by *Figure 2*, where the correspondence between input and output blocks is given in the form of a codebook. On the left is a column of the $2^M$ possible messages, listed as $M$-digit binary numbers in numerical order. Following each message is the $N$-digit binary sequence which is the corresponding signal, so that the codebook has $2^M$ entries in all.

The system in operation is shown in *Figure 1*. The source selects a message that is coded into a transmitted signal and sent over the noisy channel. The received block of $N$ (the received, or noisy, signal) differs from the transmitted signal in about $pN$ of its $N$ symbol values. The decoder receives this noisy signal and reproduces one of the $2^M$ possible messages, with an average probability $P_e$ of making an incorrect choice. In order to minimize $P_e$ the decoder must operate so that the message selected, when a given noisy signal is received, is the one corresponding to the signal most likely to have been transmitted.

For the BSC, the signal most likely to have been transmitted is the one that differs from the received signal in the fewest symbol positions. This follows from the fact that a particular group of $j$ errors has probability $p^j q^{N-j}$ of being introduced by the channel. This probability decreases as $j$ increases, for any $p < \frac{1}{2}$. If the received sequence differs from two or more signals in the same (minimal) number of places, the decoding decision is ambiguous and an error may be made. If the noise has altered half or more of the positions in which the transmitted signal differs from some other permissible signal, then the decoding decision will be incorrect or ambiguous, and an error may be made.

For the BEC, the signal most likely to have been transmitted is the one that agrees with all of the received symbols that have not been erased. If there is more than one such signal, the decoding process is ambiguous and may lead to an error. In later computation it is more convenient to discuss the probability $Q$ of an ambiguity or error in the decoding operation than to discuss the error probability $P_e$ itself. The probability of ambiguity is greater than the probability of error, since some guesses in ambiguous situations will be correct. However for ambiguity to exist there must be at least two equiprobable alternatives. Thus we have the inequalities

$$Q \geqslant P_e \geqslant \tfrac{1}{2}Q$$

For given $M$ and $N$ the probability of ambiguity $Q$ depends critically on the set of signal sequences that are used in the code. This may be discussed most easily in a geometric language introduced by HAMMING[5]. Each signal sequence is taken as a point or vector in an $N$-dimensional space, with co-ordinates equal to the values (zero or one) of its $N$ binary symbols. The distance between two points is defined as the number of co-ordinates in which they differ. Then the probability $Q$ will be large if all of the signal sequences used are clustered together in a small region of the space: it will be small if they are far from one another.

Shannon's second coding theorem, as specially referred to these two channels, states an asymptotic relation between $M$, $N$, and $P_e$ or $Q$ for a suitable selection of signal sequences. A number of definitions of channel and code parameters are necessary before stating some stronger versions of this theorem for the binary symmetric and binary erasure channels.

### CODE AND CHANNEL PARAMETERS

Given a BSC, with transmission error probability $p < \frac{1}{2}$, and $q = 1 - p$, its capacity $C = C(p)$ can be defined in terms of the entropy $H(p)$ of the $p$, $q$ distribution.

$$H(p) = -p \log p - q \log q, \text{ and} \qquad \ldots.(1)$$

$$C(p) = 1 - H(p) \qquad \ldots.(2')$$

For a BEC with erasure probability $p < 1$, the capacity $C(p)$ is given by

$$C(p) = 1 - p = q \qquad \ldots.(2'')$$

(Equations that refer to both the BSC and the BEC have unprimed numbers. Equations that refer to the BSC alone have a single prime on the equation number; those that refer to the BEC alone have a double prime.)

Either of the parameters, $C$ or $p$, completely defines a channel of either type. To define completely a code of the type shown in *Figure 2* (a block code) we need a specification of all its signal points. The most important single parameter of the code, however, is the number of signal points, which is determined by the rate $R$ of transmission, in bits per symbol. In terms of *Figure 2*, we have

$$R = \frac{M}{N} \qquad \ldots.(3)$$

and the total number of signal points in the code is $2^M = 2^{NR}$.

In order to transmit over a noisy channel with arbitrarily small error probability, $R$ must be less than $C$. Just as $C$ may be defined in terms of $p$, it is convenient to introduce an auxiliary probability $p_1$ which may be used to define $R$. This probability is selected to make $Np_1$ an integer:

$$k_1 = Np_1 \qquad \ldots.(4)$$

For the BSC, it is required that

$$p < p_1 < \tfrac{1}{2} \qquad \ldots.(5')$$

while for the BEC

$$p < p_1 < 1 \qquad \ldots.(5'')$$

$R$ is defined in terms of $p_1$ (or $k_1$) as the maximum rate at which it is possible, by the rules of information theory, to transmit information over a noisy channel and correct all sets of $k_1$ or fewer errors in each block of $N$ transmitted symbols. For the BSC, if all sets of $k_1$ or fewer errors are to be corrected by the decoding procedure, then the code could transmit error-free information over a binary channel in which all such sets of errors occurred in each block of $N$ with equal probability. Such a channel would have an equivocation per block $N$ equal to the logarithm of the number, $V_N(k_1)$, of possible error patterns of $k_1$ or less out of $N$, given by

$$V_N(k_1) = \sum_0^{k_1} \binom{N}{j} \qquad \ldots.(6)$$

where the terms in the sum are the binomial coefficients

$$\binom{N}{j} = \frac{N!}{j!(N-j)!} \qquad \ldots.(7)$$

The maximum permissible rate $R$, in bits per symbol, is given by

$$R = 1 - (1/N) \log V_N(k_1) \qquad \ldots.(8')$$

a result obtained by Hamming by a slightly different argument[5].

For the BEC the argument is simpler. If $k_1$ erasures are made, only $N - k_1$ unerased symbols remain, and only $2^{N-k_1}$ messages can possibly be distinguished without ambiguity. Thus the rate $R$ is just

$$R = \frac{N - k_1}{N} = 1 - p_1 = q_1 \qquad \ldots .(8'')$$

Shannon's original version of the second coding theorem states that, for fixed $R < C$, block codes exist where the ambiguity probability $Q_b$ may be made arbitrarily small by choosing $N$ sufficiently large. Feinstein strengthened this result by showing that $Q_b$ may be bounded above by a decreasing exponential in $N$. We shall show that for the BSC and the BEC, $Q_b$ is bounded above and below by decreasing exponentials in $N$, and that for a considerable range of channel and code parameters the two exponents agree.

To define this range, two more parameters $p_{\text{crit}}$ and $q_{\text{crit}}$ are needed. These are defined for the BSC by

$$p_{\text{crit}} = \frac{p^{\frac{1}{2}}}{p^{\frac{1}{2}} + q^{\frac{1}{2}}}, \qquad q_{\text{crit}} = \frac{q^{\frac{1}{2}}}{p^{\frac{1}{2}} + q^{\frac{1}{2}}}$$

and for the BEC by

$$p_{\text{crit}} = \frac{2p}{1 + p}, \qquad q_{\text{crit}} = \frac{q}{1 + p}$$

## LOWER BOUND

First we need an ambiguity probability $Q_{\text{opt}}$ which is smaller than the smallest attainable ambiguity probability for block coding, $Q_b$. For the BSC, with transmission rate $R$ given by equation 8′, the ambiguity probability will be minimized if every possible received sequence differs from one (and only one) of the $2^{NR}$ signal sequences in $k_1$ or fewer positions. This follows from the fact that the probability $p^j q^{N-j}$ of a particular set of $j$ errors is a monotonic decreasing function of $j$ for $p < \frac{1}{2}$. This minimum ambiguity probability $Q_{\text{opt}}$ is just the tail of the binomial distribution—the probability of more than $k_1$ errors in transmission. Thus

$$Q_{\text{opt}} = \sum_{j=k_1+1}^{N} \binom{N}{j} \qquad \ldots .(9)$$

For the BEC, with transmission rate $R$ given by equation 8″, the same equation (equation 9) holds. For, if more than $k_1$ errors occur, it is not possible to distinguish all of the $2^{NR}$ messages, and ambiguity must arise.

## UPPER BOUND

An upper bound to the ambiguity probability $Q_b$ is computed by Shannon's original procedure of random coding, here carried to a quantitative conclusion. 'Random coding' means that the $2^{NR}$ signal sequences are selected from the $2^N$ possibilities independently at random, with equal probabilities assigned to each possible sequence. $Q_{\text{av}}$, the average of the ambiguity

probabilit
ability of
The re:
nature of
for asymp
factorials

of the ass
*Theorem*
probabilit
8 is boun

In term
asymptoti

$Q_b \leqslant$

$\geqslant$

and the e

$A_b =$

The co

The as

The ex

$A_b =$

probabilities of all such codes, is certainly larger than the ambiguity probability of the best of them[1]. This quantity is evaluated in reference 16.

The results may be summarized in a theorem. Since the exponential nature of the dependence of $Q_b$ on $N$ is of interest, results will also be given for asymptotic bounds obtained by using Stirling's approximation for all the factorials in the binomial coefficients, and for the exponents

$$A = \lim_{N\to\infty} \frac{\log Q}{N} \qquad \ldots.(10)$$

of the assorted kinds of $Q$'s.

*Theorem 1.* (a) For fixed $p$ and $p_1$, with $p < p_1 < p_{\text{crit}}$, the ambiguity probability $Q_b$ for the best block code of transmission rate $R$ given by equation 8 is bounded as a function of $N$ for the BSC by

$$\begin{aligned} Q_b \leqslant Q_{\text{av}} &\leqslant p^{Np_1} q^{Nq_1} \binom{N}{Np_1} \left\{ \frac{pq_1}{p_1 - p} + \frac{1}{1 - (q/p)(p_1/q_1)^2} \right\} \\ &\geqslant Q_{\text{opt}} \geqslant p^{Np_1} q^{Nq_1} \binom{N}{Np_1} \frac{q_1}{p_1 + (1/N)} \end{aligned} \qquad \ldots.(11')$$

In terms of rate $R$ (equation 8′), and capacity $C$ (equation 2′), we have the asymptotic bounds

$$\begin{aligned} Q_b \leqslant Q_{\text{av}} &\leqslant \left\{ \frac{pq_1}{p_1 - p} + \frac{1}{1 - (q/p)(p_1/q_1)^2} \right\} . 2^{-N[-C+R+(p_1-p)\log(q/p)]} \\ &\geqslant Q_{\text{opt}} \approx \left\{ \frac{pq_1}{p_1 - p} \right\} 2^{-N[-C+R+(p_1-p)\log(q/p)]} \end{aligned} \qquad \ldots.(12')$$

and the exponent

$$\begin{aligned} A_b = \lim_{N\to\infty} \frac{\log Q_b}{N} &= -[-C + \lim_{N\to\infty} R + (p_1 - p) \log (q/p)] \\ &= -[-H(p_1) + H(p) + (p_1 - p) \log q/p] \end{aligned} \qquad \ldots.(13')$$

The corresponding results for the BEC give the same lower bound:

$$\begin{aligned} Q_b \leqslant Q_{\text{av}} &\leqslant p^{Np_1} q^{Nq_1} \binom{N}{Np_1} \left\{ \frac{pq_1}{p_1 - p} + \frac{1}{1 - (q/p)(p_1/2q_1)} \right\} \\ &\geqslant Q_{\text{opt}} \geqslant p^{Np_1} q^{Nq_1} \binom{N}{Np_1} \frac{q_1}{p_1 + (1/N)} \end{aligned} \qquad \ldots.(11'')$$

The asymptotic bounds are

$$\begin{aligned} Q_b \leqslant Q_{\text{av}} &\leq (2\pi N p_1 q_1)^{-\frac{1}{2}} \left\{ \frac{pq_1}{p_1 - p} + \frac{1}{1 - (q/p)(p_1/2q_1)} \right\} \\ &\quad \times 2^{-N[H(p)-H(p_1)+(p_1-p)\log(q/p)]} \\ &\geqslant Q_{\text{opt}} \approx (2\pi N p_1 q_1)^{-\frac{1}{2}} \left( \frac{pq_1}{p_1 - p} \right) 2^{-N[H(p)-H(p_1)+(p_1-p)\log(q/p)]} \end{aligned} \qquad \ldots.(12'')$$

The exponent is unchanged:

$$A_b = \lim_{N\to\infty} \frac{\log Q_b}{N} = -[H(p) - H(p_1) + (p_1 - p) \log (q/p)] \qquad \ldots.(13'')$$

(*b*) For $p_1 > p_{\text{crit}}$, the lower bound is unchanged. For the BSC the upper bound has the exponent

$$A_{\text{av}} = \lim_{N\to\infty} \frac{\log Q_{\text{av}}}{N} = -(C_{\text{crit}} - \lim_{N\to\infty} R) + A_{\text{crit}}$$

$$= -(H(p_1) - H(p_{\text{crit}})) + A_{\text{crit}} \qquad \ldots.(14')$$

and for the BEC

$$A_{\text{av}} = -(C_{\text{crit}} - R) + A_{\text{crit}} = -(p_1 - p_{\text{crit}}) + A_{\text{crit}} \qquad \ldots.(14'')$$

where $A_{\text{crit}}$ is $A_{\text{av}}(p_1)$ evaluated at $p_1 = p_{\text{crit}}$, and $C_{\text{crit}}$ is $C(p)$ at $p = p_{\text{crit}}$

This theorem is proved in reference 16.

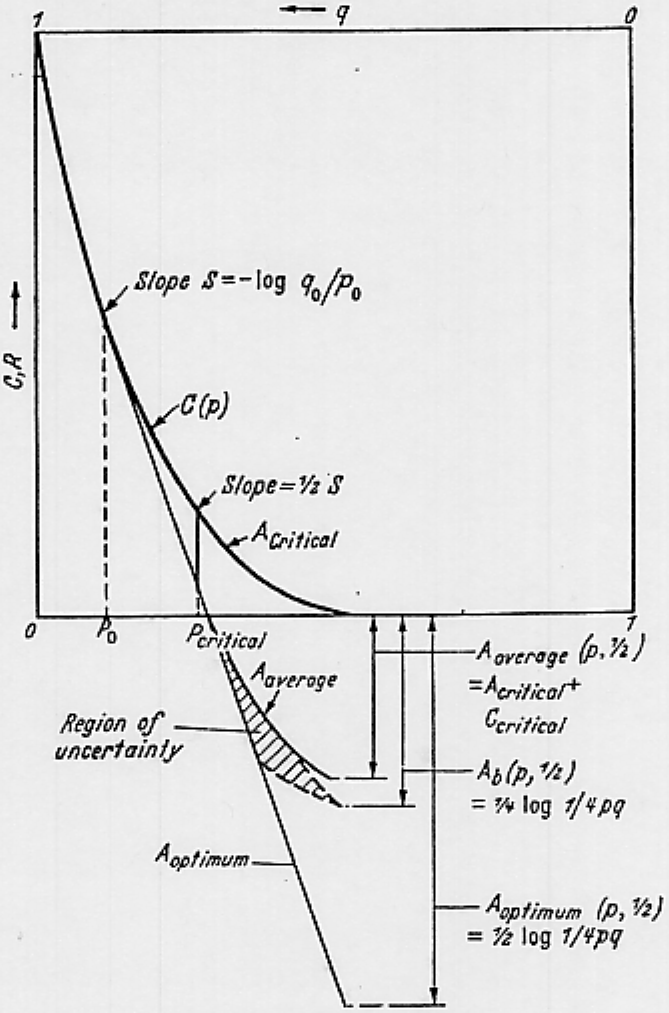


*Figure 3*(*a*). BSC exponent.

## GEOMETRICAL INTERPRETATION

The most dramatic implications of the theorem concern the nature of the exponential decrease in probability of ambiguity. $Q_{\text{av}}$ and $Q_{\text{opt}}$, the upper and lower bounds to $Q_b$, have the same exponent for $p_1 < p_{\text{crit}}$, and their ratio approaches a constant for large $N$. Furthermore, as $p_1 \to p$ the asymptotic ratio of $Q_{\text{av}}$ to $Q_{\text{opt}}$ approaches unity. This means that random coding is as good as anything else for transmission rates near to channel capacity. In fact, for any $p_1 < p_{\text{crit}}$, random coding is very nearly as good as anything else. For any $\epsilon > 0$ at sufficiently large values of $N$, $Q_{\text{opt}}$ for coding in blocks of length $N$ will be greater than $Q_{\text{av}}$ for coding in blocks of length $(1 + \epsilon)N$.

The behaviour of the exponents of $Q_{\text{opt}}$ and $Q_{\text{av}}$ have a simple geometric interpretation. The expression $[H(p) - H(p_1) + (p_1 - p) \log (q/p)]$ is the difference between the change in $H$, between points $p$ and $p_1$, and the change in a tangent to $H$ at $p$:

$$dH(p)/dp = \log (q/p) \qquad \ldots.(15)$$

This geometry is illustrated in *Figure 3*. *Figure 3a* shows the BSC exponent. The capacity curve $C(p) = 1 - H(p)$ is plotted against $p$. For given $p$, a tangent is drawn to this curve at $(p, C(p))$. For any $p_1 < p_{\text{crit}}$, the length of a perpendicular dropped from the capacity curve to the tangent line is the exponent of the ambiguity probability, for either optimum or random coding. For $p_1 > p_{\text{crit}}$, the optimum coding exponent is still the length of a perpendicular from the capacity curve to the tangent line, but the average coding exponent is smaller, and its perpendicular terminates on a curve lying above the tangent line. For $p_1 \to \frac{1}{2}$, the values of $A_{\text{av}}$ and $A_{\text{opt}}$ approach the limits shown in the illustration. It also can be deduced from results given by Plotkin[7] and Gilbert[6] that $A_b$ approaches a limit different from either of these. For $p_1$ near $\frac{1}{2}$, and thus for signalling rates very near to channel capacity, the best block code is definitely better than average coding but not so good as optimum coding. (The two dotted lines diverging from this point and bounding the region of uncertainty in the illustration are derived in Appendix 2 of reference 16 from Plotkin and Gilbert's work.)

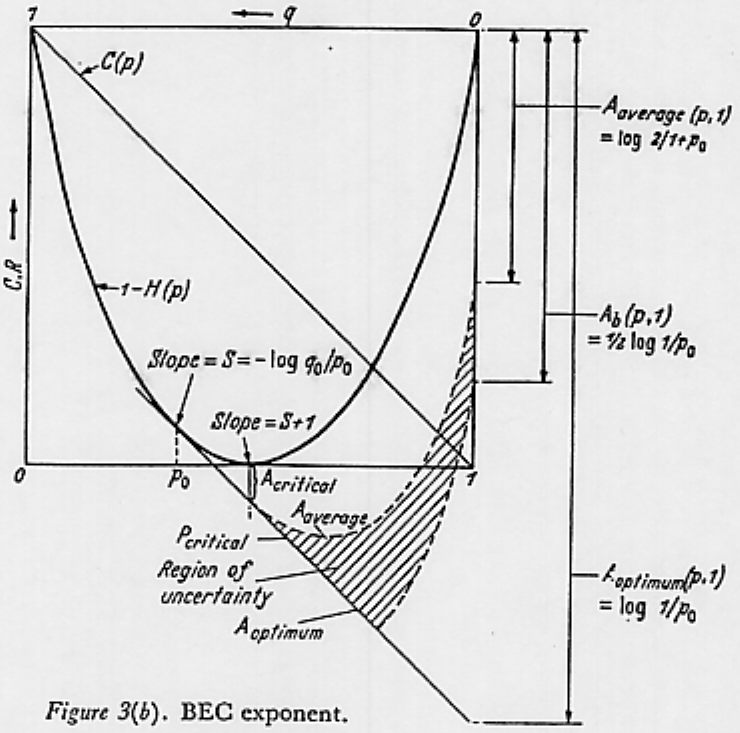

1
← q
0
C(p)
C,R →
1−H(p)
Slope = S = − log $q_0/p_0$
Slope = S+1
0
$P_0$
$A_{critical}$
1
$A_{average}$
$P_{critical}$
Region of uncertainty
$A_{optimum}$
$A_{average}$ (p, 1) = log 2/1+$p_0$
$A_b$ (p,1) = ½ log 1/$p_0$
$A_{optimum}$(p.1) = log 1/$p_0$

*Figure 3(b)*. BEC exponent.

*Figure 3b* shows the same sort of plot for the BEC. In this case the capacity curve is a straight line. The exponent is still the length of a perpendicular dropped from the curve $1 - H(p)$ to the tangent line for optimum coding, to the curve which diverges from it at $p_{\text{crit}}$ for random coding, and to somewhere in the region of uncertainty for best possible block coding. The limiting values of the exponents as $p_1 \to 1$ are shown on the right.

One conclusion which may be drawn from these figures is that it only pays to be clever in designing a code when $p_1 > p_{\text{crit}}$; that is, when transmission is at rates appreciably below channel capacity. Another conclusion follows from the fact that, for both channels, $A_b$ approaches $\frac{1}{2}A_{\text{opt}}$ as $C$ approaches 0. This means that for transmission at very low rates, in order to obtain a given ambiguity probability with the best possible block code, it is necessary to use a block length $N^2$, where $N$ is the length which would suffice if a noiseless feedback channel were available and optimum coding could be used.

### CONSTRUCTION

Theorem 1 shows that for signalling in blocks of length $N$ at rates $R$ near to the channel capacity, random coding is essentially as good as optimum coding. Optimum coding requires a noiseless feedback channel. Plotkin's results show this for the BSC for $p \geqslant \frac{1}{4}$, and it seems likely to hold for any $p > 0$ for sufficiently large $N$. For the BEC, we shall show that no block code can be optimum in its behaviour for any fixed $p > 0$ and sufficiently large $N$. If the feedback channel is not available, then random coding is the only quasi-constructive procedure suggested by the theorem for taking advantage of the exponential decrease of $Q_b$ with $N$.

Random coding has been criticized on the basis of lack of uniformity. Since only average error probabilities are computed, there is no guarantee that any one code, or any one signal sequence in a code, will be near to the average in its behaviour; but, Shannon has pointed out[1,14] if the average of a set of positive quantities is $\epsilon$, then at most $(1/n)$ of them can be as large as $n\epsilon$. Thus most codes are good codes and, in a good code, by throwing away the worst half of the signal sequences and thus reducing $R$ by only $(1/N)$ bits per symbol, the signal sequences may be made uniformly good.

There is one far more practical objection, however. There are $2^{NR}$ entries in a codebook and $N$ binary digits in each entry. This codebook must be stored at transmitter and receiver, which is impractical for values of $N$ and $R$ large enough to be useful in greatly reducing the ambiguous detection probability $Q_b$. Furthermore, the receiver must compare each incoming sequence with every entry in the codebook, which takes a great deal of computing time. It is, of course, possible to devise coding schemes that are systematic, and have simple schemes giving the signal sequence in terms of the message sequence, and the message sequence in terms of the corrupted sequence[11,12]. However, the only scheme of this kind that has been shown to transmit information at a positive rate does not attain channel capacity[12], nor does its error probability decrease as rapidly as it should. This is reasonable enough, for the codes describable by a simple set of rules are a very small fraction of all codes, when $N$ is large, and the fact that the average behaviour of all codes is good is no guarantee of the behaviour of this very small subset.

Thert
that ha
set of al
ability c
practica
actual c
This
scheme
procedu
have bee

In the
sequenc

check sy
Hammin
symbols.
$(N - k_1$
ones anc
the mat
and the
digits ar

There is an alternative procedure. This is to find a small subset of codes that have a simple encoding and decoding procedure and are typical of the set of all possible codes, in the sense that they have the same average probability of error. The random selection of one of these codes will provide a practical solution to the problem, although it may not be as satisfying as the actual construction of a single coding scheme known to be well behaved.

This programme is carried out below for the BEC, and a random coding scheme with algebraic (rather than codebook) encoding and decoding procedures is shown to behave as well as the unrestricted random coding we have been discussing. Similar results for the BSC have been given elsewhere[13].

## RANDOM PARITY-CHECK CODING FOR THE BEC

In the codes to be constructed, the first $(N - k_1)$ symbols in the signal sequence are simply the message symbols. The remaining $k_1$ symbols are check symbols, each of which completes a parity-check, like those used by HAMMING[5], on a random selection of about half of the first $(N - k_1)$ symbols. The procedure is illustrated in *Figure 4*. The binary digits in the $(N - k_1) \times k_1$ coefficient matrix are selected independently at random, with ones and zeros equiprobable. The $N - k_1$ message digits are written above the matrix. Denoting the matrix elements by $a_{ij}$, the message digits by $m_j$, and the check digits by $c_i$, where $1 \leqslant i \leqslant k_1$, $1 \leqslant j \leqslant N - k_1$, the check digits are determined by

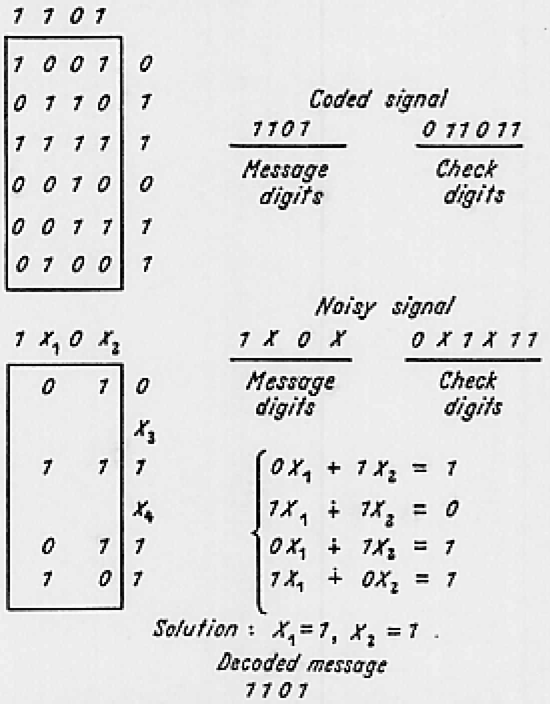


*Figure 4.* Parity-check coding and decoding.

$$c_i = \sum_{j=1}^{N-k_1}{}' a_{ij} m_j \qquad \ldots .(16'')$$

where the summation is modulo two. Thus $c_i$ is unity if there is an odd number of columns in which the message and the $i$th matrix row both have ones present, and is zero otherwise. The block of $N$ digits is transmitted by sending the message digits reading from left to right across the top of the matrix, and then the check digits reading down the right side.

In transmission over the channel, suppose that $j$ erasures occur, $j_1$ of them among the message digits, and $j_2$ among the check digits. The receiver has available the coefficient matrix $\| a_{ij} \|$ used by the transmitter. Writing the received sequence above and to the right of the matrix, in the order used by the transmitter, the receiver selects those columns of the matrix which correspond to erased message digits and those rows which correspond to unerased check digits. These are written as the coefficient matrix of a set of equations, with the erased message digit values as the unknowns. The right-hand terms $d_i$ of this set of equations are formed by adding to each unerased check digit $c_i$ the sum of $a_{ij}m_j$,

$$d_i = c_i + \sum_{j=1}^{N-k_1}{}' a_{ij}m_j, \qquad \ldots . (17'')$$

where the addition and summation are again modulo two, the prime indicates summation only over unerased $m_j$, and the equation holds for unerased $c_i$.

This gives a set of $k_1 - j_2$ equations, modulo two, in the $j_1$ unknown message digit values. If these equations are soluble, the erased symbols are determined. If $j_1 > k_1 - j_2$, so that $j = j_1 + j_2 > k_1$, then there are not enough equations to determine the missing digit values, and there will be ambiguity in the decoding. If $k_1 - j_2 > j_1$, then there are more equations than unknowns. No question of over-determination arises, since one solution of the set of equations certainly exists: the digits present in the original message, which have been erased by the channel. There will be no ambiguity if $j \leqslant k_1$, then, unless fewer than $j_1$ of the equations are linearly independent. Thus the probability $Q_{\text{av}}(j)$ of ambiguous decoding, which is unity for $j > k_1$, is for $j < k_1$ just the probability of an indeterminate set of equations. This is evaluated in reference 16. It gives the same bound on $Q_{\text{av}}$.

This finishes the demonstration that random parity-check coding has essentially the same ambiguity probability as does the random coding. Thus the remainder of the random coding derivation applies unaltered, and the resulting statements about $Q_{\text{av}}$ still hold. In fact, we have

*Theorem 2.* The results of Theorem 1 are unchanged by the restriction of permissible codes to codes of the parity-check type.

For the BEC, this follows from the identical behaviour of random sequence coding and random parity-check coding. A restriction on the class of permissible codes can only increase the minimum attainable error probability, so that the error probability of the optimum code is a lower bound to that of the parity-check code *a fortiori*. For the BSC it is also possible to use a random parity-check code, and to show that it leads to the same bound on $Q_{\text{av}}$ (see reference 12, Appendix) and this completes the proof of Theorem 2. It is even possible to make a stronger statement: the bounds used in Appendix 2 of reference 16 for obtaining the behaviour of the error probability at low transmission rates, as illustrated in *Figure 3* by the dotted lines, all apply to parity-check codes as well, so that the remainder of *Figure 3* and the formulas of Appendix 2 of reference 16 are still valid.

## OTHER RANDOM PARITY-CHECK CODES

The random selection involved in constructing a random parity-check code is the choice of the $(N - k_1) \times k_1$ matrix of *Figure 4*, which requires $N^2 p_1 q_1 \leqslant (\frac{1}{4})N^2$ random binary digits, rather than the $N \cdot 2^{NR}$ required for random selection of signal sequences. It is possible to construct a random parity-check code requiring only $(N - 1)$ random binary selections by modifying the coefficient matrix, as shown in *Figure 5*. Here, the first $N - k_1$ random digits are used for the first row in the matrix, the second to the $(N - k_1 + 1)^{\text{st}}$ for the second row . . . , the $k_1$'th to the $(N - 1)^{\text{st}}$ for the last row. It is again possible to show that for both the BEC and the BSC the same bounds still hold for $Q_{\text{av}}$, and Theorem 2 still applies. (The demonstration is outlined for the BSC in reference 13 and will not be given here.) If in a code of this type, which may be called a sliding parity-check code, the check digits are interspersed among the information digits, then the block length $N$ can be made indefinitely large. If the receiver waits long enough, for any $p_1 > p$ he is assured, with probability unity, that he will ultimately obtain an independent set of equations large enough to determine all of the symbols which have been erased up to that point. This is, of course, not a highly practical procedure, since it has an indefinitely large memory requirement, but it serves to give the $N$ in the expressions for $Q_{\text{av}}$ the interpretation, not of a block length, but of a delay time, which is much more intuitively appealing. With a sliding code, $N$ symbols after the receipt of a given message symbol the receiver can produce a decoded version of that symbol with an error probability bounded by the expressions given in Theorem 1. If he is willing to wait longer he may get the lower probability

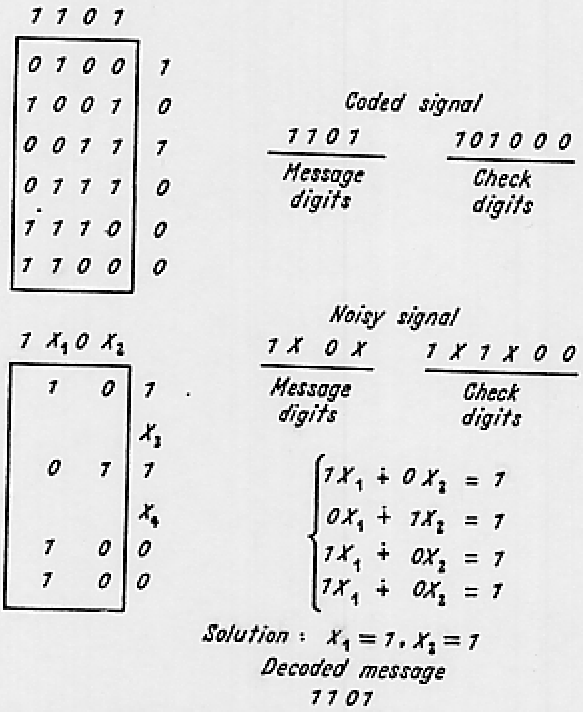


*Figure 5.* Sliding parity-check coding and decoding.

associated with a larger value of $N$, without requiring the transmitter to modify the coding procedure. Such a coding procedure is error-free, in the sense defined in reference 12, and answers the question raised there as to whether or not error-free coding was possible at no sacrifice in transmission rate of error probability.

*After the analysis reported in references 16 and 13 was completed, but before it had been organized for presentation, I discovered that C. E. Shannon was also working on the problem of error probability. In discussing our results he mentioned the geometric interpretation giving the error exponent for small $(p_1 - p)$ as the difference between the capacity curve and a tangent line, used in Figure 3. The geometric picture is essential: any other presentation of the results gets lost in families of curves, since the exponent is a function of both $p$ and $p_1$.*

*This work has benefited from the interest of my colleagues, Professors Fano, Huffman, and Yngve of the Research Laboratory of Electronics at M.I.T.*

## REFERENCES

[1] Shannon, C. E. 'A Mathematical Theory of Communication', *Bell System Tech. J.*, 27 (1948) 379, 623

[2] — 'Communication in the Presence of Noise', *Proc. I.R.E.*, 37 (1949) 10

[3] Rice, S. O. 'Communication in the Presence of Noise—Probability of Error of Two Encoding Schemes', *Bell System Tech. J.*, 29 (1950) 60

[4] Feinstein, A. 'A new Basic Theorem of Information Theory', *Trans. I.R.E.* (PGIT), 4 (1954) 2

[5] Hamming, R. W. 'Error Detecting and Error Correcting Codes', *Bell System Tech. J.*, 29 (1950) 147

[6] Gilbert, E. N. 'A Comparison of Signalling Alphabets', *Bell System Tech. J.*, 31 (1952) 504

[7] Plotkin, M. 'Binary Codes with Specified Minimum Distance', *Univ. of Penna., Moore School Research Division Report* 51–20 (1951)

[8] Golay, M. J. E. 'Binary Coding', *Trans. I.R.E.* (PGIT), 4 (1954) 23

[9] Laemmel, A. E. 'Efficiency of Noise-reducing Codes', pp. 111–118, in *Communication Theory*, reference 4 above

[10] Muller, D. E. 'Metric Properties of Boolean Algebra and their Application to Switching Circuits', *University of Illinois, Digital Computer Laboratory Report* No. 46

[11] Reed, I. S. 'A Class of Multiple Error-correcting Codes and the Decoding Scheme', *Trans. I.R.E.* (PGIT), 4 (1954) 38

[12] Elias, P. 'Error-free Coding', *Trans. I.R.E.* (PGIT), 4 (1954) 30

[13] — 'Coding for Noisy Channels', pp. 37–46 *Record of the 1955 I.R.E. National Convention*, part 4, 1955

[14] Shannon, C. E. Comment at the 1954 Symposium on Information Theory, M.I.T., Cambridge, Mass., U.S.A.

[15] Feller, W. *An Introduction to Probability Theory and Its Applications*, New York; Wiley and Sons, Inc., 1950

[16] Elias, P. *Coding for Two Noisy Channels*, to be published

## DISCUSSION

D. Slepian: Dr. Elias has shown that it is possible to signal at rates arbitrarily close to the capacity of the binary symmetric channel with arbitrarily small probability of error, using codes of certain restricted classes. One such class of codes is the parity-check codes first introduced by Hamming. These codes have some very special properties to make them attractive from the practical point of view: one such obvious and well-known property is the ease with which they can be generated. I have recently investigated this class of code in more detail and have found several other simple, practical properties which they possess. In the first place, maximum likelihood detection is much easier to accomplish with these codes than one might have expected beforehand. Secondly, for certain values of the signal parameters, there exist no better codes; this is true in the practical case of a channel with little noise. Thirdly, all transmitted messages have the same probability of being decoded correctly. I have compiled a list of best parity-check codes of ten binary digits length or less. This list, together with the theory of these codes, is available in my paper which has been circulated to you but has not been read at this Symposium*.

I should like now to comment on a remark made by Dr. Golay. There has been considerable confusion in the literature between codes and decoding schemes. Shannon's encoding theorem is proved on the basis of using the best possible detection scheme, mainly a maximum likelihood detector which identifies each possible received signal with the nearest possible sent signal. Now most computations that I have seen on the probability of error associated with parity-check codes have not used maximum likelihood detection. Rather these computations are based on the notion of drawing disjoint spheres about the possible sent messages. Received messages inside a sphere are associated with the centre of that sphere. Received signals which do not lie in any of these spheres are ignored; that is, they are counted as errors in computing the probability that a sent message be decoded correctly. Let us call this method of detection 'bounded distance detection'. It is not hard to show that one cannot achieve the ideal signalling of Shannon's theorem with bounded distance detection if $p$, the probability of error per binary digit on the channel, is between $\frac{1}{4}$ and $\frac{1}{2}$. In fact, in the appendix of Elias's full paper you will find a proof of this fact with the figure $\frac{1}{4}$ being replaced by a somewhat smaller number. These remarks hold, of course, for parity-check codes. It is known that they are not good if one uses bounded distance detection and if $p$ is in the range just referred to. It is my guess that with bounded distance detection they cannot be used to approach the codes of Shannon's theorem for any value of $p$. What Elias showed at the New York meeting of the I.R.E. this Spring, and has mentioned, is that if one uses maximum likelihood detection, then parity-check codes are as good as the average code and hence can be used to approach the results of Shannon's theorem.

D. A. Bell: The British Standards Institute *Glossary of Terms used in Telecommunication* defines *Channel* as 'a means of one-way communication' and, in this sense, the author is correct in speaking of a 'binary symmetric channel'. (Any system in which the occurrence of an error was made known to the transmitter would require at least *two* channels on this definition.) Yet in spite of the Glossary I am still tempted to think of a channel as that which exists between terminal equipments *e.g.* an allocation of bandwidth for radio transmission, an open-wire line or a cable in which any repeaters are linear amplifiers. Now the binary-symmetric or binary erasure characteristic of a channel is a function of its terminal equipment, and I suggest therefore that the investigation of the merits of a coding system for a channel (in the author's sense of channel) is not complete until we have considered also the merits of the terminal equipment. If over the greater part of the distance between terminals there is a channel having continuous properties which can be defined in terms of bandwidth and noise power, the communication rate achieved by a code should ultimately be compared with the maximum rate predicted by Shannon for a continuous channel.

B. Mandelbrot: Some of Dr. Elias' results can also be deduced by continuing the argument of Feinstein (*cf.* B. Mandelbrot, *Ann. Telecomm.*, June 1955). I should like to ask Dr. Elias if he can say more about the relationship between Feinstein's work and random coding.

* A summary of Dr. Slepian's paper appears on page 399.

P. Elias in reply: Dr. Slepian's statements are quite correct, and the confusion which has existed between bounded distance detection and maximum likelihood detection was admirably illustrated in a note of Dr. Zaremba's circulated to the participants of this symposium.

Dr. Bell is of course correct in saying that if a channel which is not binary or symmetric is available, then its capacity should not be computed as if it were binary and symmetric; but his definition of 'channel' sounds exceedingly narrow. It would, for example, rule out scatter channels in which noise is not additive but in part multiplicative, and would also rule out a human operator repeating his best guess at a noisy received signal.

Dr. Mandelbrot's question is difficult to answer briefly, but in general Feinstein's work may be considered as random coding operating under constraints. These constraints do not reduce channel capacity, nor do they alter the exponent in the exponentially decreasing error probability, so far as the leading term for rates very near channel capacity is concerned. However, they do increase the error probability for somewhat lower transmission rates compared with what unconstrained random coding can do.

TI

*Depart*

*I*

Linear se
filters. Tl
over an i
character
operator,
networks
adders ar
steady-sta
Several m
one of wh
attention
since, if a
are clearl

A binary
outputs,
symbols d

with the a
the preser
voltage,
further as
tion with

* This w
(Air Resear
States.