#5

LIST DECODING FOR NOISY CHANNELS

PETER ELIAS

Technical Report 335

September 20, 1957

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

Reprinted from the
1957 IRE WESCON CONVENTION RECORD
Part 2

PRINTED IN THE U.S.A.

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-64637, Department of the Army Task 3-99-06-108 and Project 3-99-00-100.

LIST DECODING FOR NOISY CHANNELS

PETER ELIAS

Technical Report 335

September 20, 1957

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

Reprinted from the
1957 IRE WESCON CONVENTION RECORD
Part 2

PRINTED IN THE U.S.A.

LIST DECODING FOR NOISY CHANNELS

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

Summary. Shannon's fundamental coding theorem for noisy channels states that such a channel has a capacity C, and that for any transmission rate R less than C it is possible for the receiver to use a received sequence of n symbols to select one of the 2^{nR} possible transmitted sequences, with an error probability P_e which can be made arbitrarily small by increasing n, keeping R and C fixed. Recently upper and lower bounds have been found for the best obtainable P_e as a function of C,R and n. This paper investigates this relationship for a modified decoding procedure, in which the receiver lists L messages, rather than one, after reception. In this case for given C and R, it is possible to choose L large enough so that the ratio of upper and lower bounds to the error probability is arbitrarily near to 1 for all large n. This implies that for large L, the average of all codes is almost as good as the best code, and in fact that almost all codes are almost as good as the best code.

Introduction

The binary symmetric channel (abbreviated BSC) is illustrated in Fig. 1. It accepts two input symbols, 0 and 1, and produces the same two symbols as outputs. There is probability p that an input symbol will be altered in transmission, from 0 to 1 or from 1 to 0, and probability q = 1-p that it will be received correctly. Successive transmitted digits are treated by the channel with statistical independence.

The capacity C of this channel is given by

$$C = 1 - H(p) \quad \text{bits/symbol} \qquad (1)$$

where

$$H(p) = -p \log p - q \log q \qquad (2)$$

is the entropy of the p,q probability distribution. C is plotted in Fig. 1 as a function of p. Note that if $p > q$, the output symbols can be reinterpreted, an output 1 being read as a 0 and vice versa, so that only values of $p < q$ will be considered. (The logarithms in Eq. 2 and throughout the paper are taken to the base 2.)

C can be given two interpretations. First, if 0's and 1's are transmitted with equal probability and statistical independence, then one bit per symbol is supplied at the input, and an average of H(p) bits per symbol of equivocation remains in the output. Second, by the fundamental coding theorem for noisy channels, if the input information rate is reduced from 1 bit per symbol to any rate $R < C$ bits per symbol, it is possible for the receiver to reconstruct the transmitted sequence of input symbols with as low an error probability as is desired, at the cost of a delay of n symbols between transmission and decoding, where n must be increased to reduce the error probability.

In block coding, the input information rate is reduced by using only a restricted number,

This work was supported in part by the Army (Signal Corps), the Air Force (Office of Scientific Research, Air Research and Development Command), and the Navy (Office of Naval Research).

$M = 2^{nR}$, of the 2^n possible sequences of length n. This gives an average rate of $(1/n)\log M = R$ bits per symbol, the sequences being used with equal probability. The receiver then lists the single transmitter sequence which is most probable, on the evidence of the received sequence of n noisy symbols. It is sometimes wrong, with a probability which depends on the particular set of sequences which have been selected for transmission. We define P_{opt}, a function of n, R and C, as the minimum average probability of error for a block code: i.e. as the average error probability for the code whose transmitter sequences have been chosen most judiciously, the average being taken over the different transmitter sequences.

Unfortunately we do not know the optimum codes for large values of n: for small n, a number are known (see Slepian (3,4)). However it is possible to define a lower bound to P_{opt}, which we denote by P_t, which is a fairly simple function of the code and channel parameters. It is also possible to define P_{av}, the average error probability of all possible codes, which is an upper bound to P_{opt}. The behavior of P_{opt} is discussed in some detail in references (1,2). As the transmission rate approaches the channel capacity, P_{av} approaches P_t, so that the error probability P_{opt} is well defined. For somewhat lower rates the two bounds differ, but for fixed R and C their ratio is bounded above and below by constants as n increases. Since each is decreasing exponentially with n, at least the exponent of P_{opt} is well-defined. For still lower rates, however, P_{av} and P_t decrease with different exponents as n increases with C and R fixed. In this region therefore even the exponent of P_{opt} is not known, although it is bounded. Finally, for very low rates, it can be shown that the exponent of P_{opt} is definitely different from that for either P_{av} or P_t.

As can be seen from this brief description, the behavior of P_{opt} is complicated, and not well-defined, even for the channel which has been studied most intensively. The present paper discusses a somewhat more complicated transmission procedure for use with the BSC, which has the advantage that the description of its error probability behavior is much simpler. Our procedure - block coding with list decoding - leaves a little equivocation in the received message. Instead of a unique selection of one message out of M, the receiver provides a list of L messages, $L < M$. As soon as a little residual equivocation is permitted, it turns out that P_{av} has the same exponent as P_t, and thus as P_{opt}, for <u>all</u> transmission rates and not just those which are near capacity. This is essentially Theorem 1. It also turns out that as L increases, P_{av} actually approaches P_t, so that P_{opt} itself, and not just its exponent, becomes well-defined. This is the content of Theorem 2. As a result of this simplification, the cause of detection errors in a system using list detection can be interpreted simply: mistakes are made when, in a particular block of n symbols, the channel gets too noisy. In the case of ordinary block coding, on the other hand, errors also occur because of unavoidable weak spots in the code, which causes the complication in the description of P_{opt} for such codes.

Operation

The operation of a block code with list decoding is illustrated in Fig. 2. There are M equiprobable messages, represented by the integers from 1 to M, and the source selects one, say the integer m = 2. This is coded into a sequence $u_m = u_2$ of n binary digits, 011 in the example, for which n = 3. The sequence u_m is transmitted over the noisy BSC. The channel alters some of the digits: its effect is to add a noise sequence w_r, in this case the sequence 010, to the transmitted sequence u_m to produce the received noisy sequence, v_2 = 001. (The addition is modulo 2, digit by digit: 1+1 = 0+0 = 0, 1+0 = 0+1 = 1.) The received sequence is decoded into a list of L messages - L = 3 in the Figure. If the list includes the message which the source selected, the system has operated correctly and no error has occurred. If not there has been an error. Since 2 is on the list 1,2,3, the transmission illustrated was successful.

The behavior of the system is determined by the choice of $u_1,\ldots,u_m,\ldots,u_M$, the transmitter sequences. An encoding codebook, illustrated for the example in Fig. 3, assigns M = 4 such sequences each of length n = 3, to the four messages m = 1,2,3,4. The $2^n = 8$ possible noise sequences of length n are ranked, by an index r, so that the sequence with no 1's, w_1 = 000, comes first, a sequence with k 1's precedes one with k+1 1's, and sequences with the same number of 1's are ranked by their size as binary numbers. In the example the noise sequence w_3 = 010 is added to the transmitted sequence to produce the received sequence v_2: the 2^n possible received sequences $v_1,\ldots,v_{2n}$ being ranked in the same manner as the noise sequences. The received sequence is decoded by means of a decoding codebook, as illustrated in Fig. 3, which provides a list of L = 3 messages for each received sequence.

The decoding codebook is constructed by listing, after each received sequence, the L messages whose transmitter sequences are converted into the given received sequence by noise sequences of the lowest possible rank. Because of the equal probability of the transmitter sequences and the statistical independence of transmitter sequence and noise sequence in the BSC, the L messages which are most probable on the evidence of the received sequence are those which are converted into the received sequence by the L most probable noise sequences. These are just the ones which will be listed in the codebook, except that the ranking of the noise sequences will provide an unambiguous decision when several noise sequences have the same probability. This follows because the probability $\text{Prob}[w_r]$ of a noise sequence w_r with k 1's is just

$$\text{Prob}[w_r] = p^k q^{n-k} = q^n (p/q)^k, \qquad (3)$$

which is monotone decreasing in k for $p < q$, while rank is monotone increasing with k.

The behavior of the system under all possible circumstances is illustrated in the table of possible events in Fig. 3. The message sequence u_2 and the noise sequence w_3 determine a row and a column in the table. The received sequence v_2=001 which is their digit-by-digit sum modulo two is entered at the intersection. The decoding codebook shows that v_2 will be decoded as the list 1,2,3 which includes the message m = 2 which was

actually transmitted. This entry in the table is therefore circled, as an event which the system will handle without error, and so are all such events. The probability of error for the system is then just the sum of the probabilities of the uncircled events, which are all of the entries in the last two columns.

Rate of Transmission

The average amount of information received over this kind of system when no error has been made depends on the probabilities of the different messages on the decoded list. To make it constant, we assume that the messages on the list are shuffled in order before being given to the sink, so that they are equiprobable to him. Then the rate is

$$R = \frac{\log M - \log L}{n} \quad \text{bits/symbol}, \quad (4)$$

and to transmit at rate R requires that

$$M = L \cdot 2^{nR} \quad (5)$$

messages be transmitted. In the example of Figs.2 and 3, we have M = 4, L = 3, n = 3, and R = (1/3) log (4/3) $\doteq$ 0.14 bits/symbol.

Note that the largest M which can be used effectively is $M = 2^n$, since there are only that many distinct sequences. Thus $L = 2^{n(1-R)}$ is as large a list as would be of any interest. This would correspond to transmitting uncoded information-i.e., all possible sequences.

Error Probability: Lower Bound

In any row of the table in Fig. 3, all 2^n possible received sequences appear, since each of the 2^n noise sequences produces a distinct received sequence when added to the given transmitter sequence. Therefore the number of circled entries in row m is equal to the total number of times that the integer m appears in all of the 2^n lists of L messages in the decoding codebook. Thus a total of $L \cdot 2^n$ entries are circled, out of the total of $M \cdot 2^n$ entries in the table, or an average of $2^{n(1-R)}$ circles per message.

Clearly the minimum probability of error would be obtained if all of the circles were as far to the left as possible in the table. This is true for the code in Fig. 3, but will not be possible for other values of M, L and n. However it provides a lower bound to error probability in any case. We define P_t as the error probability for a code (perhaps unrealizeable) with such a table.

Let r_1 be the last column in which any circled entries appear:

$$\begin{aligned} r_1 M &\geq L \cdot 2^n > (r_1 - 1)M, \text{ or} \\ r_1 &\geq 2^{n(1-R)} > r_1 - 1 \quad . \end{aligned} \quad (6)$$

Then P_t is bounded:

$$\sum_{r=r_1}^{2^n} \text{Prob}[w_r] > P_t \geq \sum_{r=r_1+1}^{2^n} \text{Prob}[w_r]. \quad (7)$$

The quantity P_t is characteristic of the channel, and of its use at rate R in blocks of N symbols, but is independent of the code used. The channel parameters p,q enter in evaluating $\text{Prob}[w_r]$

$$\text{Prob}[w_r] = p^{k(r)} q^{n-k(r)} \quad , \quad (8)$$

where k(r) is determined by means of the binomial coefficients $\binom{n}{j} = n!/j!(n-j)!$,

$$\sum_{j=0}^{k(r)} \binom{n}{j} \geq r > \sum_{j=0}^{k(r)-1} \binom{n}{j} \qquad (9)$$

Using Eq. 8, and introducing k_1 as an abbreviation for $k(r_1)$ defined through Eq. (9), we have the following bounds for P_t:

$$\sum_{r=r_1}^{2^n} p^{k(r)} q^{n-k(r)} > P_t \geq \sum_{r=r_1+1}^{2^n} p^{k(r)} q^{n-k(r)} \qquad (10)$$

$$\sum_{k=k_1}^{n} \binom{n}{k} p^k q^{n-k} > P_t \geq \sum_{k=k_1+1}^{n} \binom{n}{k} p^k q^{n-k} \qquad (11)$$

The first of these is needed for Theorem 2. The second, which is simpler to evaluate, suffices for Theorem 1.

Error Probability: Upper Bound

An upper bound to the best attainable average error probability is obtained by Shannon's procedure of random coding - i.e. by computing the average error probability for all possible codes. This turns out to be much easier than computing the error probability for any one particular code. Since a code is uniquely determined by its set $u_1, \ldots, u_M$ of transmitter sequences, there are as many codes possible, for rate R, list size L and length n, as there are selections of $M = L \cdot 2^{nR}$ objects out of 2^n possibilities, repetitions permitted, or

$$2^{nM} = 2^{nL \cdot 2^{nR}}$$

in all. We define P_{av} as the average error probability, averaged over all transmitter sequences in any one code and over all codes, giving equal weights to each.

P_{av} can be expressed as a sum over r of the probability of occurrence of the rth noise sequence w_r, times the ensemble average conditional probability $Q_L(r)$ of making a decoding error when the noise sequence w_r occurs:

$$P_{av} = \sum_{r=1}^{2^n} p^{k(r)} q^{n-k(r)} Q_L(r) \qquad (12)$$

Now we can bound P_{av} above and below. Since it is an average of quantities all of which are greater than P_t, P_t is a lower bound. Since $Q_L(r)$ is a conditional probability and thus less than one, it can be bounded above by one for a portion of the range of summation. Thus, using Eq. 10,

$$\begin{aligned} P_t \leq P_{av} &\leq \sum_{r=1}^{r_1} p^{k(r)} q^{n-k(r)} Q_L(r) \\ &\quad + \sum_{r_1+1}^{2^n} p^{k(r)} q^{n-k(r)} \qquad (13) \\ &\leq \sum_{r=1}^{r_1} p^{k(r)} q^{n-k(r)} Q_L(r) + P_t \end{aligned}$$

More crudely, we define $Q_L^*(k)$:

$$Q_L^*(k(r)) = Q_L\left(\sum_{j=0}^{k(r)} \binom{n}{j} \right) \geq Q_L(r) , \qquad (14)$$

by Eq. 9 and the fact that $Q_L(r)$ is monotone increasing in r (See Eq. 16, below). Then

$$P_t \leq P_{av} \leq \sum_{k=0}^{k_1} \binom{n}{k} p^k q^{n-k} Q_L^*(k) + P_t . \qquad (15)$$

Finally, an explicit expression can be given for $Q_L(r)$. Given that the rth noise sequence has occurred in a transmission, the probability of a

decoding error is just the probability that L or more of the other M-1 transmitter sequences differ from the received sequence by error sequences of rank $\leq r$. This is the probability of L successes in M-1 statistically independent tries, when the average probability of success is $r/2^n$, for in the ensemble of codes the different transmitter sequences are statistically independent. Thus $Q_L(r)$ is a binomial sum:

$$Q_L(r) = \sum_{j=L}^{M-1} \binom{M-1}{j} (r/2^n)^j (1-r/2^n)^{M-1-j} . \quad (16)$$

The Parameter p_1 . Theorem 1 .

It would be desirable to express the error probability directly as a function of rate, capacity, block size n and list size L. However it turns out to be more convenient for Theorem 1 to use, instead of rate, the parameter $p_1 = k_1/n$, where $k_1 = k(r_1)$ is defined by Eq.9. Fixing p_1 and n determines the rate R through Eq. 6. For large n the relation is simple: if we define the rate R_1 by the equation

$$R_1 = 1 - H(p_1) \quad (17)$$

as illustrated in Fig. 1b, then for fixed p_1, R approaches R_1 from above as n increases. This is shown in references (1,2).

We can now state

Theorem 1

Given a BSC with error probability p, and given any p_1, with $0 < p < p_1 < \frac{1}{2}$, and any $L > L_o(p,p_1)$, the ratio of P_{av} to P_t for list decoding with list size L is bounded, independent of n:

$$1 \leq P_{av}/P_t \leq A(p,p_1,L) . \quad (18)$$

The value of L_o needed for the theorem is

$$L_o = \frac{\log(q/p)}{\log(q_1/p_1)} - 1 . \quad (19)$$

Corollary

Under the same conditions, the exponent of P_{opt} as n increases is

$$\begin{aligned} &\lim_{n\to\infty} \{(1/n)\log P_{opt}\} \\ &= \lim_{n\to\infty} \{(1/n)\log P_{av}\} \\ &= \lim_{n\to\infty} \{(1/n)\log P_t\} \\ &= H(p_1) - H(p) - (p_1-p)\log(q/p) . \end{aligned} \quad (20)$$

The proof of the corollary given the theorem is direct. Taking logarithms and limits in Eq. 18 shows that P_{av} and P_t have a common exponent. Since they bound P_{opt} above and below, it shares the same exponent. The value given in Eq. 20 is the exponent of P_t, as found in references (1,2). It is illustrated geometrically in Fig. 1b.

Proof of Theorem 1.

Bounding P_{av} in terms of P_t requires bounding the sum in Eq. 15. For the last term in the sum we use the bound $Q_L^*(k_1) = 1$. For the earlier terms we use the bound of Eq. A6, derived in Appendix A. Eq. 15 then becomes

$$\begin{aligned} P_{av} \leq e^L \binom{n}{k_1-1} \sum_{k=0}^{k_1-1} p^k q^{n-k} \left\{ \binom{n}{k} \Big/ \binom{n}{k_1-1} \right\}^{L+1} \\ + \binom{n}{k_1} p^{k_1} q^{n-k_1} + P_t . \end{aligned} \quad (21)$$

Now the sum in Eq. 21 can be bounded by a geometric series, which sums:

$$\sum_{k=o}^{k_1-1} p^k q^{n-k} \left\{ \binom{n}{k} \Big/ \binom{n}{k_1-1} \right\}^{L+1}$$

$$\leq \frac{p^{k_1-1} q^{n-k_1+1}}{1 - (q/p)(p_1/q_1)^{L+1}} \quad , \tag{22}$$

where we have used $k_1 = np_1$ and $n-k_1 = nq_1$. The convergence of the series is guaranteed by the requirement that $L > L_o$ as given by Eq. 19, so that the denominator on the right in Eq. 22 is positive.

Substituting Eq. 22 in Eq. 21 and regrouping,

$$P_{av} \leq \binom{n}{k_1} p^{k_1} q^{n-k_1} \left\{ 1 + \frac{qp_1}{pq_1} \frac{e^L}{1 - (q/p)(p_1/q_1)^{L+1}} \right\} + P_t \quad . \tag{23}$$

Now, from Eq. 11,

$$P_t \geq \binom{n}{k_1+1} p^{k_1+1} q^{n-k_1-1}$$

$$\geq (p/q) \binom{n}{k_1} p^{k_1} q^{n-k_1} \quad . \tag{24}$$

Substituting Eq. 24, read in reverse, in Eq. 23 gives

$$P_{av}/P_t \leq (q/p) \left\{ 1 + \frac{qp_1}{pq_1} \frac{e^L}{1 - (q/p)(p_1/q_1)^{L+1}} \right\} + 1 \quad , \tag{25}$$

QED.

Theorem 2.

The proof of Theorem 1 suggests that $A(p,p_1,L)$ increases with L, but this is due to the weakness of the bound of Eq. A6 on $Q_L^*(k)$ for k near k_1. Actually $A(p,p_1,L)$ decreases as L increases. In fact we have

Theorem 2

Given a BSC with capacity C, and a rate R with $0 < R < C < 1$, and given any $\epsilon > 0$, it is possible to find an $L(\epsilon)$ so large that for all $n > n_o(L)$,

$$1 \leq P_{av}/P_t \leq 1 + \epsilon \quad . \tag{26}$$

Corollary

For sufficiently large L and n, almost all codes are almost as good as the best code.

Given the theorem, the corollary follows by an argument of Shannon's in reference (5). The difference between P_{opt} and the error probability of any code is positive or zero, and the average value of the difference, by Theorem 2, is $< \epsilon$. Thus only a fraction 1/T of the codes may have such a difference $> T\epsilon$, and at least a fraction (1-1/T) have an error probability $< (1+T\epsilon)P_{opt}$, QED. The result here is stronger than in Shannon's application because of the nearby lower bound.

Proof of Theorem 2.

For a constant rate R, $p_1 = k_1/n$ is not a constant, but depends on n. However for sufficiently large n it is arbitrarily close to the limiting value given by setting $R = R_1$ in Eq. 17. Let p_1^* be this limit. Then

$$\lim_{n \to \infty} \left\{ \binom{n}{k_1-j} \Big/ \binom{n}{k_1-1} \right\} = (p_1^*/q_1^*)^j \quad . \tag{27}$$

Choose j so that

$$(p_1/q_1)^j = e^{-(1+a)} \quad , a > 0 . \tag{28}$$

Then breaking the sum in Eq. 13 at

$$r_o = \sum_{k=o}^{k_1-j} \binom{n}{k} \tag{29}$$

and using the bound of Eq. A6 with the substitution of Eq. 27 below this point gives a term which is bounded by

$$\frac{e^{-(L+1)a} \binom{n}{k_1} p^{k_1} q^{n-k_1}}{1 - (q/p)(p_1^*/q_1^*)^{L+1}} \tag{30}$$

Next, the portion of the sum in Eq. 13 which is between r_o, given in Eq. 29, and r_1, is bounded. Since $p^k q^{n-k}$ is monotone decreasing in k, we have

$$\sum_{r=r_o}^{r_1} p^{k(r)} q^{n-k(r)} Q_L(r) < (q/p)^j p^{k_1} q^{n-k_1} \sum_{r=r_o}^{r_1} Q_L(r). \tag{31}$$

We now bound $Q_L(r)$ by Eq. B4, derived in appendix B. This gives

$$\begin{aligned} \sum_{r=r_o}^{r_1} Q_L(r) &< \sum_{r=r_o}^{r_1} e^{-(L/2)(r_1-r)^2/r_1^2} \\ &< 1 + \int_0^\infty e^{-Lr^2/2r_1^2} \, dr \\ &< 1 + (r_1/2)\sqrt{2\pi/L} . \end{aligned} \tag{32}$$

From Eq. 9, using a geometric series to bound the sum of binomial coefficients as in references (1,2,6) gives

$$r_1 < \binom{n}{k_1} \frac{1}{1 - (p_1/q_1)} . \tag{33}$$

Substituting this in Eq. 32, we have

$$\sum_{r=r_o}^{r_1} Q_L(r) < p^{k_1} q^{n-k_1} \binom{n}{k_1} \left\{ \frac{C_1}{\sqrt{L}} + \frac{C_2}{\binom{n}{k_1}} \right\} . \tag{34}$$

Now the denominator under C_1 increases arbitrarily with L, and the denominator under C_2 increases arbitrarily with n. Thus the summation in Eq. 34 becomes negligible for large n and L compared to the term

$$p^{k_1} q^{n-k_1} \binom{n}{k_1} . \tag{35}$$

So does the expression in Eq. 30 which bounds the remainder of the sum in Eq. 13, because of its exponential factor. But from Eq. 24, this is not true for P_t, which is in fact of the order of the expression in Eq. 35. Thus in the limit the right side of Eq. 13 approaches P_t, QED.

Conclusion.

Two points should be noted in conclusion. First, if L grows exponentially with n - that is, if there is a residual <u>percentage</u> of equivocation, rather than a residual <u>amount</u> - then the ratio P_{av}/P_t approaches 1 exponentially, since $\sqrt{L}$ is then a negative exponential in n. Second, the discussion of Theorem 2 shows that P_{opt} is very near P_t, but does not say just how big either of them are. P_t is of course given by Eq. 9, with r_1

defined by Eq. 6, and the exponent is given by Eq. 20. But no really tight bounds for P_t or P_{opt} in terms of C, R, and finite n and L have been given. These take a good deal more of detailed algebra, which will be presented elsewhere.

Acknowledgements.

Theorem 1 was discovered independently by Professor J. M. Wozencraft in the course of his doctoral research. It does not appear in his thesis (10) since it was not directly relevent to the final topic, but he will publish a note on the subject shortly. Shannon introduced the author to the work of Chernoff and Cramer (7,8), which makes the bound of Appendix B possible. Without this kind of bound the argument for Theorem 2 is even more tedious.

APPENDIX A: A BOUND ON $Q_L(r)$.

The probability of L or more successes in M-1 tries, with probability $r/2^n$ of success on each try, can be bounded by the product of $(r/2^n)^L$, the probability of success in a particular L of the M-1 tries, times the number of ways of selecting L from M-1 :

$$Q_L(r) \leq \binom{M-1}{L} (r/2^n)^L < (M^L/L!)(r/2^n)^L \tag{A1}$$

Stirling's lower bound to L! gives

$$\begin{aligned} Q_L(r) &\leq (1/\sqrt{2\pi L})(e/L)^L M^L (r/2^n)^L \\ &\leq (e^L/\sqrt{2\pi L})(Mr/L2^n)^L \\ &\leq (e^L/\sqrt{2\pi L})(r/2^{n(1-R)})^L , \end{aligned} \tag{A2}$$

making use of Eq. 5. From Eq. 6 we have

$$Q_L(r) = (e^L/\sqrt{2\pi L})(r/(r_1-1))^L \tag{A3}$$

Now from Eq.'s 6 and 9,

$$2^{n(1-R)} = \sum_{j=0}^{k_1-1} \binom{n}{j} \tag{A4}$$

And, by term-by-term comparison, for $k \leq k_1-1$,

$$\sum_{j=0}^{k} \binom{n}{j} \Big/ \sum_{j=0}^{k_1-1} \binom{n}{j} \leq \binom{n}{k} \Big/ \binom{n}{k_1-1} , \tag{A5}$$

so from Eq. A2 and the definition in Eq. 14 we have for $k \leq k_1-1$,

$$\begin{aligned} Q_L^*(k) &= (e^L/\sqrt{2\pi L}) \left\{ \binom{n}{k} \Big/ \binom{n}{k_1-1} \right\}^L , \\ &< e^L \left\{ \binom{n}{k} \Big/ \binom{n}{k_1-1} \right\}^L \end{aligned} \tag{A6}$$

APPENDIX B: ANOTHER BOUND ON $Q_L(r)$.

A result due to Chernoff (7), which may also be derived from Cramer (8), provides bounds on the tails of distributions of sums of random variables in terms of the moment-generating function of the parent distribution. Applying this to the binomial distribution, Shannon (9) shows that for $k_b = Np_b > Np_a$,

$$\sum_{k=k_b}^{N} \binom{N}{k} p_a^k q_b^{N-k} \leq \left\{ \left(\frac{p_a}{p_b}\right)^{p_b} \left(\frac{q_a}{q_b}\right)^{q_b} \right\}^N \tag{B1}$$

Taking $1/N$ times the natural logarithm of the right side of Eq. B1, and defining $\delta = p_b - p_a > 0$, gives

$$p_b \log_e(1-\delta/p_b) + q_b \log_e(1+\delta/q_b) = -\sum_{j=2}^{\infty} (\delta^j/j)\left\{\frac{1}{p_b^{j-1}} (-)^j \frac{1}{q_b^{j-1}}\right\}. \quad \text{(B2)}$$

We take $q_b > p_b$. Then $1/q_b < 1/p_b$, so the coefficient of δ^j in Eq. B2 is > 0. Then the sum in Eq. B2 is more negative than its leading term, and

$$\sum_{k=k_b}^{N} \binom{N}{k} p_a^k q_b^{N-k} \leq \exp\left(- N(\delta^2/2)[(1/p_b) + (1/q_b))\right) \leq \exp\left(- N\delta^2/2p_b q_b\right) < \exp\left(- N\delta^2/2p_b\right) . \quad \text{(B3)}$$

In applying this to $Q_L(r)$, we have $p_a = r/2^n$, $p_b = r_1/2^n$, $\delta = (r_1 - r)/2^n$, and we may use $N = M$, since increasing M-1 to M in Eq. 16 increases the sum. Thus, using Eqs. 5 and 6,

$$Q_L(r) \leq \exp\left(- (M/2)(r_1-r)^2/r_1 . 2^n \right) \leq \exp\left(- (L/2)(r_1-r)^2/2^{n(1-R)} r_1 \right) \quad \text{(B4)}$$

$$\leq \exp\left(- (L/2)(r_1-r)^2/r_1^2 \right) , \quad \text{QED.}$$

References

1. Elias, P., "Coding for Noisy Channels", I.R.E. Convention record part 4, p.37-46 (1955).

2. Elias, P., "Coding for Two Noisy Channels", in Cherry, (editor) Information Theory, p.61-74, Butterworth (1956).

3. Slepian, D., "A Class of Binary Signalling Alphabets", Bell Syst. Tech. J. 35, p.203-234 (1956).

4. Slepian, D., "A Note on Two Binary Signalling Alphabets", Trans. I.R.E. PGIT IT-2, 61-74 (1956).

5. Shannon, C. E., "The Mathematical Theory of Communication", University of Illinois Press (1949) see p.40 .

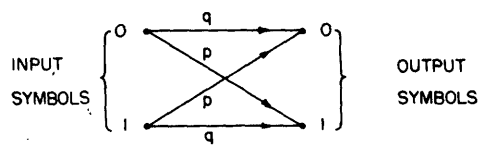
6. Feller, W., Probability Theory and its Applications, Wiley (1950) see p.126, Eq. (8.1) .

7. Chernoff, H. "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on a Sum of Observations", Ann. Math. Stat. 23, (1952).

8. Cramer, H. "Sur un Nouveau Theoreme-Limite de la Theorie des Probabilites" in Actualites Sci. et Indust., no. 736, Hermann et cie (1938).

9. Shannon, C. E., Information Seminar Notes , M.I.T. (1956) (unpublished).

10. Wozencraft, J. M., Sequential Decoding for Reliable Communications , Dissertation submitted to Elect. Eng. Dept., M.I.T. (1957).



THE BINARY SYMMETRIC CHANNEL

Fig. 1a.

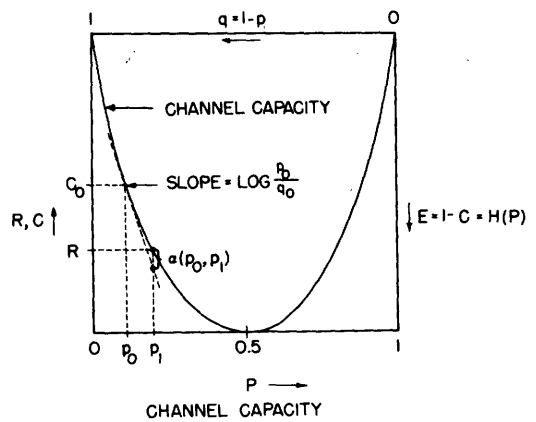
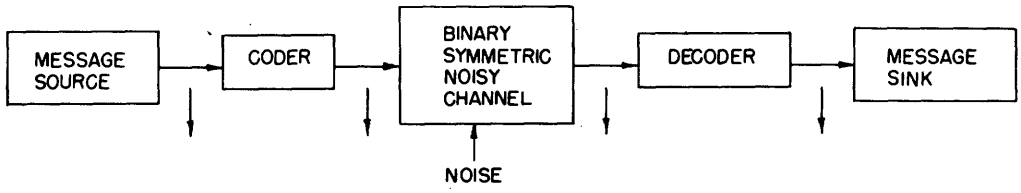


Fig. 1b.



MESSAGE 2 → TRANSMITTED SEQUENCE 011 (n SYMBOLS) → RECEIVED SEQUENCE 001 (n SYMBOLS) → DECODED MESSAGE LIST 1,2,3 (L MESSAGES)

NOISY CHANNEL OPERATION:

$\oplus$ 011 TRANSMITTED SEQUENCE
010 NOISE SEQUENCE
$=$ 001 RECEIVED SEQUENCE

BLOCK CODING AND LIST DECODING

Fig. 2.

ENCODING CODE BOOK		NOISE SEQUENCES		DECODING CODE BOOK	
MESSAGE m	TRANSMITTER SEQUENCE u_m	RANK r	SEQUENCE w_r	RECEIVED SEQUENCE v_s	DECODED MESSAGE LIST
1	$u_1 = 000$	1	$w_1 = 000$	$v_1 = 000$	1,2,3
2	$u_2 = 011$	2	$w_2 = 001$	$v_2 = 001$	1,2,3
3	$u_3 = 101$	3	$w_3 = 010$	$v_3 = 010$	1,2,4
4	$u_4 = 110$	4	$w_4 = 100$	$v_4 = 100$	1,3,4
		5	$w_5 = 011$	$v_5 = 011$	1,2,4
		6	$w_6 = 101$	$v_6 = 101$	1,3,4
		7	$w_7 = 110$	$v_7 = 110$	2,3,4
		8	$w_8 = 111$	$v_8 = 111$	2,3,4

EXAMPLE OF LIST DECODING SYSTEM

Fig. 3a.

NOISE SEQUENCES	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8
PROB $[w_r]$	q^3	pq^2	pq^2	pq^2	p^2q	p^2q	p^2q	p^3
	(v_1)	(v_2)	(v_3)	(v_4)	(v_5)	(v_6)	v_7	v_8
u_1	(v_5)	(v_3)	(v_2)	(v_8)	(v_1)	(v_7)	v_6	v_4
u_2	(v_6)	(v_4)	(v_8)	(v_2)	(v_7)	(v_1)	v_5	v_3
u_3	(v_7)	(v_8)	(v_4)	(v_3)	(v_6)	(v_5)	v_1	v_2
u_4								

TRANSMITTER SEQUENCES

CIRCLED ENTRIES ARE CORRECTLY DECODED

EXAMPLE OF LIST DECODING SYSTEM

Fig. 3b.