THE MAXIMUM ACHIEVABLE LENGTH OF AN ERROR CORRECTING CODE[1]

BY
J. WOLFOWITZ

1. Introduction

The present paper continues the investigation initiated in [1]. It is written so that it can be read without a knowledge of [1] at least until the details of the proof begin.

We shall assume that the alphabets involved contain exactly two symbols, 0 and 1. This involves no loss of generality, as the extension to alphabets with any finite number of symbols is immediate and straightforward. Suppose that a person has a vocabulary of S words (or messages), any or all of which he may want to transmit, in any frequency and in any order, over a "noisy channel" of memory m (m an integer $\geqq 0$).

We shall now explain what is meant by a noisy channel of memory m. A sequence of $(m + 1)$ elements, each zero or one, will be called an α-sequence. Let the α-sequences be numbered in some fixed but arbitrary manner from 1 to 2^{m+1}. The channel probability function $p(i \mid j)$ is a nonnegative function defined for $i = 0, 1$ and $j = 1, \cdots, 2^{m+1}$, such that for every j, $p(0 \mid j) + p(1 \mid j) = 1$. Any sequence of n elements, each of which is zero or one, is called an x-sequence. To describe the operation of the channel it will be sufficient to describe how it transmits any given x-sequence, say x_1. Let α_1 be the α-sequence of the first $(m + 1)$ elements of x_1. The channel "performs" a chance experiment with possible outcomes 0 and 1 and respective probabilities $p(0 \mid \alpha_1)$ and $p(1 \mid \alpha_1)$, and transmits the outcome of this chance experiment. It then performs another chance experiment, independently of the first, with possible outcomes 0 and 1 and respective probabilities $p(0 \mid \alpha_2)$ and $p(1 \mid \alpha_2)$, where α_2 is the α-sequence of the 2^{nd}, 3^{rd}, $\cdots$ $(m + 2)^{\text{nd}}$ elements of the sequence x_1. This is repeated until $(n - m)$ independent experiments have been performed. The probabilities of the outcomes 0 and 1 in the i^{th} experiment are, respectively, $p(0 \mid \alpha_i)$ and $p(1 \mid \alpha_i)$, where α_i is the α-sequence of the i^{th}, $(i + 1)^{\text{st}}$, $\cdots$, $(i + m)^{\text{th}}$ elements of x_1. The x-sequence x_1 is called the transmitted sequence. The chance sequence $Y(x_1)$ of outcomes of the experiments in consecutive order is called the received sequence. Any sequence of $(n - m)$ elements, each zero or one, is called a y-sequence.

Let $P\{\ \}$ denote the probability of the relation in braces, and let $0 < \lambda < 1$. A "code" of length t is a set $\{(x_i, A_i), i = 1, \cdots, t\}$, where

Received January 13, 1958.

[1] This research was supported by the U. S. Air Force through the Office of Scientific Research.

each x_i is an x-sequence, each A_i is a set of y-sequences, for each i we have

$$P\{Y(x_i) \in A_i\} \geqq 1 - \lambda,$$

and the sets $A_1, \cdots, A_t$ are disjoint. The coding problem which is a central concern of the theory of transmission of messages may be described as follows: For given S, to find an n and then a code of length S. The practical applications of this will be as follows: When one wishes to transmit the i^{th} word one transmits the x-sequence x_i. Whenever the receiver receives a y-sequence which is in A_j, he always concludes that the j^{th} word has been sent. Whenever the receiver receives a y-sequence not in $A_1 \cup A_2 \cup \cdots \cup A_S$, he may draw any conclusion he wishes about the word that has been sent. The probability that any word transmitted will be correctly received is $\geqq 1 - \lambda$.

The purpose of this paper is to prove the following:

THEOREM 5.[2] *There exists a functional J of the channel probability function p with the following property: Let $\varepsilon > 0$ be arbitrary, and let n be sufficiently large. There exists a code of length $2^{n(J-\varepsilon)}$. No code can have a length greater than $2^{n(J+\varepsilon)}$.*

The functional J will be defined below by specified algebraic and analytic operations on p. It seems reasonable to call J "the capacity" of the channel.

2. Relation to previous results

When $m = 0$, a functional C_0 of p is defined in [1] (Section 6) and there called the capacity (when $m = 0$). It is proved in [1] that, when $m = 0$, there is a functional $K > 0$ of p such that, for any n, there exists a code of length $2^{nC_0 - Kn^{1/2}}$ (Theorem 1), and no code can have a length greater than $2^{nC_0 + Kn^{1/2}}$ (Theorem 2; see also Theorem 4 of [2]). Hence, when $m = 0$, $C_0 = J$, and Theorem 5 is weaker than Theorems 1 and 2 of [1].

When $m \geqq 0$, functionals C_1 and C_2 of the channel probability function are defined respectively in footnotes 5 and 9 of [1]. (Unfortunately the letters C_1 and C_2 are also used in another sense in [1], but only in Section 4 and for a particular argument. We shall not make use of these latter quantities again). From their definition we always have $C_0 \leqq C_1 \leqq C_2$. It was proved in [3] and [1] that, for any $\varepsilon > 0$ and n sufficiently large, there exists a code of length greater than $2^{n(C_2 - \varepsilon)}$. Hence $C_0 \leqq C_1 \leqq C_2 \leqq J$. We have seen that when $m = 0$ these four quantities coincide. What their relation is for general m is at present unknown. The functional C_1 is easier to compute than J, and the functional C_2 is at least as difficult to compute as J. If further investigation should prove that $C_1 < J$, then it may not be of practical value to determine whether $C_2 < J$. If, however, it should turn out that $C_1 = J$, then this would be very useful for computational purposes. Moreover, then the positive part of Theorem 5 could be strengthened by using Theorem 1 (of [1]) which asserts that there is a functional $K' > 0$ of the channel prob-

[2] The first four theorems of the present investigation appeared in [1] and [2].

ability function such that there exists a code of length $2^{nC_1-K'n^{1/2}}$. The author conjectures that, for $m > 1$, $C_1 < J$.

If one neglects the ε's which occur in the statement of Theorem 5, then the latter roughly states that 2^{nJ} is the maximum achievable length of code. Any improvement on this result must consist in a more precise analysis of the εn term in the exponent. For example, in some of the theorems cited above, the εn term was replaced by a $Kn^{1/2}$ term.

One could also consider a channel such that the probability that a particular symbol be received in a given place is a function not only of the symbol transmitted in that place and the m preceding symbols transmitted, as in the present case, but also of the m' preceding symbols received. Such a channel can be treated by the same method as that used below, with a result analogous to Theorem 5.

3. Proof of the theorem

Let l be any integer $> m$. Only for the purposes of this proof let us call any sequence of l elements, each zero or one, an l-sequence, and any sequence of $(l - m)$ elements, each zero or one, an l'-sequence. Let the l-sequences (l'-sequences) be numbered in some fixed but arbitrary manner from 1 to 2^l (from 1 to 2^{l-m}). Let $\phi(i \mid j)$, $i = 1, \cdots, 2^{l-m}$; $j = 1, \cdots, 2^l$, be the probability that the i^{th} l'-sequence is received when the j^{th} l-sequence is sent.

Suppose we restrict ourselves to n which are multiples of l. Then any x-sequence can be regarded as a sequence of n/l disjoint l-sequences. Suppose further that in any y-sequence (which of course has $(n - m)$ elements) we "pay no attention" to the elements in the places whose serial numbers are congruent to $(l - m + 1)$ or $(l - m + 2)$ or $\cdots$ or l, modulo l; this means that we do not distinguish between y-sequences which differ *only* in elements in such places. Then what we have is a transmission system in which the input alphabet (alphabet of symbols transmitted) has the elements $1, \cdots, 2^l$, the output alphabet (alphabet of symbols received) has the elements $1, \cdots, 2^{l-m}$, the memory is zero, and the channel probability function is ϕ. The place of n in the original system is taken by n/l in the new system. Since y-sequences different in the original system may be identified in the new system, but never vice versa, it follows that, for any fixed n, any length of code achievable in the new system is also achievable in the original system.

We now proceed with the new system exactly as in Section 6 of [1], except for the unimportant difference that the alphabets may now contain more than two symbols. Let $X_1, X_2, \cdots$ be independent, identically distributed, chance variables such that

$$P\{X_1 = i\} = \rho_i \qquad i = 1, \cdots, 2^l$$

and $\sum_i \rho_i = 1$. Let $Y_1, Y_2, \cdots$ be chance variables which take only the values $1, \cdots, 2^{l-m}$, and such that the conditional probability that $Y_j = i$, given $X_1, X_2, \cdots, Y_1, \cdots, Y_{j-1}$, is $\phi(i \mid X_j)$. Then the entropy of the

Y process is ([1], equation (6.8))

$$H(Y_0) = -\sum_j \left(\sum_i \rho_i \phi(j \mid i)\right) \log_2 \left(\sum_i \rho_i \phi(j \mid i)\right),$$

where the summation with respect to i is from 1 to 2^l, and the summation with respect to j is from 1 to 2^{l-m}. Also the conditional entropy of the Y-process relative to the X-process is ([1], equation (6.9))

$$H_X(Y_0) = -\sum_j \sum_i \rho_i \phi(j \mid i) \log_2 \phi(j \mid i)$$

with the same limits of summation as before. Let $C_0(l)$ be the maximum of

$$H(Y_0) - H_X(Y_0)$$

with respect to $\rho_1, \cdots, \rho_{2^l}$, which are subject to the conditions

$$\text{all } \rho_i \geqq 0, \qquad \sum_i \rho_i = 1.$$

Then $C_0(l)$ is the capacity ([1], Section 6) of the new system described in the preceding paragraph. It may be mentioned in passing that $C(l_1) + C(l_2) \leqq C(l_1 + l_2)$. Although we shall not use this fact below, it may make easier the computation of J.

We now define

$$J = \sup_{l>m} C_0(l)/l.$$

Let $\varepsilon > 0$ be given. Let l be an integer $> m$ such that $C_0(l)/l > J - \varepsilon/2$. Then

$$(n/l)\,(C_0(l) - \varepsilon/2) = n(C_0(l)/l - \varepsilon/2l) > n(J - \varepsilon).$$

By Theorem 1 of [1], for n sufficiently large, there exists in the new system a code of length greater than

$$2^{(n/l)(C_0(l)-\varepsilon/2)} > 2^{n(J-\varepsilon)}.$$

Hence for the same n there exists in the original system a code of length greater than $2^{n(J-\varepsilon)}$. This establishes the first part of the theorem.

We now prove the second part of the theorem. Suppose that, from every y-sequence, we delete the elements in all places whose serial numbers are congruent to $(l - m + 1), (l - m + 2), \cdots, l$, modulo l. Let us call the resulting sequences y'-sequences. Every x-sequence may be considered as a sequence of n/l disjoint l-sequences. If, for transmission, we used x-sequences and y'-sequences, then there would be no memory between (or among) the l-sequences which compose the x-sequence, and Theorem 2 of [1] would apply. (While Theorem 2 was proved for the case when both sending and receiving alphabets contain two symbols, it obviously applies when there are any finite number of symbols in each; it is especially easy to see how the proof of Theorem 4 of [2] carries over, and the latter implies Theorem 2.) We obtain the following result: For $\varepsilon > 0$ and n sufficiently large, there cannot exist a code of length greater than $2^{(n/l)(C_0(l)+\varepsilon/2)}$. However, this is not yet the desired result, because in the system under consider-

ation by us it is y-sequences and not y'-sequences which are received sequences.

If one considers the proof of Theorem 2 of [1], or, better still, of the more general Theorem 4 of [2], one observes that the proof rests on bounding the number of y-sequences generated by an x-sequence. Every y'-sequence is a subsequence of $2^{(n/l-1)m}$ y-sequences. Suppose we change the definition of a y-sequence generated by an x-sequence as follows: Consider a y-sequence to be the composition of the y'-sequence which is a subsequence of it, hereafter to be called the first sequence, and the subsequence of all elements in the y-sequence but not in the y'-sequence, hereafter to be called the second sequence. Say that the y-sequence is generated by the x-sequence if the first and second sequences each fulfill the corresponding requirements of being generated in the original sense (by the sequences of α-sequences which produced them). Then all the arguments of [1] and [2] are valid with this new definition. It is easy to see that the number of y-sequences generated (in the new sense) by all x-sequences is less than $2^{(n/l-1)m}$ times the number of y'-sequences generated (in the old sense) by all x-sequences. Consequently the proof of Theorem 4 of [2], which applies to the system of x-sequences and y'-sequences, applies also to our original system of x-sequences and y-sequences, if we multiply by $2^{(n/l-1)m}$ the number of y'-sequences generated by all x-sequences. This will multiply the upper bound on the length of the code by $2^{(n/l-1)m}$. We conclude that in the original system of x-sequences and y-sequences no code can have a length greater than

$$2^{(n/l)m} \cdot 2^{(n/l)(C_0(l)+\varepsilon/2)}.$$

We have not yet specified l. Let l be so large that $m/l < \varepsilon/2$. Then

$$(n/l) \cdot m + (n/l)(C_0(l) + \varepsilon/2) < n(J + \varepsilon).$$

This completes the proof of the second part of the theorem and of the theorem itself.

From the argument of the second part of the theorem and the conclusion of the first part, one easily obtains that, whenever l is so large that $m/l < \varepsilon_1$, $J - \varepsilon_1 < C_0(l)/l$. This fact simplifies the task of computing J.

References

1. J. Wolfowitz, *The coding of messages subject to chance errors*, Illinois J. Math., vol. 1 (1957), pp. 591–606.
2. ———, *An upper bound on the rate of transmission of messages*, Illinois J. Math., vol. 2 (1958), pp. 137–141.
3. A. Khintchine, *On the fundamental theorems of the theory of information*, Uspehi Matem. Nauk (N.S.), vol. 11 no. 1 (67), (1956), pp. 17–75.

Cornell University
Ithaca, New York
Israel Institute of Technology
Haifa, Israel