A Comparison of Signalling Alphabets

By E. N. GILBERT

(Manuscript received March 24, 1952)

Two channels are considered; a discrete channel which can transmit sequences of binary digits, and a continuous channel which can transmit band limited signals. The performance of a large number of simple signalling alphabets is computed and it is concluded that one cannot signal at rates near the channel capacity without using very complicated alphabets.

INTRODUCTION

C. E. Shannon's encoding theorems[1] associate with the channel of a communications system a capacity C. These theorems show that the output of a message source can be encoded for transmission over the channel in such a way that the rate at which errors are made at the receiving end of the system is arbitrarily small provided only that the message source produces information at a rate less than C bits per second. C is the largest rate with this property.

Although these theorems cover a wide class of channels there are two channels which can serve as models for most of the channels one meets in practice. These are:

1. The binary channel

This channel can transmit only sequences of binary digits 0 and 1 (which might represent hole and no hole in a punched tape; open-line and closed line; pulse and no pulse; etc.) at some definite rate, say one digit per second. There is a probability p (because of noise, or occasional equipment failure) that a transmitted 0 is received as 1 or that a transmitted 1 is received as 0. The noise is supposed to affect different digits independently. The cpacity of this channel is

$$C = 1 + p \log p + (1 - p) \log (1 - p) \tag{1}$$

bits per digit. The log appearing in Equation (1) is log to the base 2; this convention will be used throughout the rest of this paper.

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Tech. J.*, **27,** p. 379–423 and pp. 623–656, 1948, theorems 9, 11, and 16 in particular.

2. The low-pass filter

The second channel is an ideal low-pass filter which attenuates completely all frequencies above a cutoff frequency W cycles per second and which passes frequencies below W without attenuation. The channel is supposed capable of handling only signals with average power P or less. Before the signal emerges from the channel, the channel adds to it a noise signal with average power N. The noise is supposed to be white Gaussian noise limited to the frequency band $|\nu| < W$. The capacity of this channel is

$$C = W \log\left(1 + \frac{P}{N}\right) \tag{2}$$

bits per second.

Shannon's theorems prove that encoding schemes exist for signalling at rates near C with arbitrarily small rates of errors without actually giving a constructive method for performing the encoding. It is of some interest to compare encoding systems which can easily be devised with these ideal systems. In Part I of this paper some schemes for signalling over the binary channel will be compared with ideal systems. In Part II the same will be done for the low-pass filter channel.

Part I

THE BINARY CHANNEL

1. Error-Correcting Alphabets

Imagine the message source to produce messages which are sequences of letters drawn from an alphabet containing K letters. We suppose that the letters are equally likely and that the letters which the source produces at different times are independent of one another. (If the source given is a finite state source which does not fit this simple description, it can be converted into one which approximately does by a preliminary encoding of the type described in Shannon's Theorem 9.) To transmit the message over the binary channel we construct a new alphabet of K letters in which the letters are different sequences of binary digits of some fixed length, say D digits. Then the new alphabet is used as an encoding of the old one suitable for transmission over the channel. For example, if the source produced sequences of letters from an alphabet of 3 letters, a typical encoding with $D = 5$ might convert the message into a binary sequence composed of repetitions of the three letters.

00000
11100
and 00111

If $K = 2^D$, the alphabet consists of all binary sequences of length D and hence if any of the digits of a letter is altered by noise the letter will be misinterpreted at the receiving end of the channel. If K is somewhat smaller than 2^D it is possible to choose the letters so that certain kinds of errors introduced by the noise do not cause a misinterpretation at the receiver. For example, in the three letter alphabet given above, if only one of the five digits is incorrect there will be just one letter (the correct one) which agrees with the received sequence in all but one place. More generally if the letters of the alphabet are selected so that each letter differs from every other in at least $2k + 1$ out of the D places, then when k or fewer errors are made the correct interpretation of the received sequence will be the (unique) letter of the alphabet which differs from the received sequence in no more than k places. An alphabet with this property will be called a *k error correcting alphabet*[2].

Error correcting alphabets have the advantage over the random alphabets which Shannon used to prove his encoding theorems that they are uniformly reliable whereas Shannon's alphabets are reliable only in an average sense. That is, Shannon proved that the probability that a letter *chosen at random* shall be received incorrectly can be made arbitrarily small. However, a certain small fraction of the letters of Shannon's alphabets are allowed a much higher probability of error than the average. This kind of alphabet would be undesirable in applications such as the signalling of telephone numbers; one would not want to give a few subscribers telephone numbers which are received incorrectly more often than most of the others. It is only conjectured that the rate C can be approached using error correcting alphabets. The alphabets which are to be considered here are all error correcting alphabets.

A geometric picture of an alphabet is obtained by regarding the D digits of a sequence as coordinates of a point in Euclidean D dimensional space. The possible received sequences are represented by vertices of the unit cube. A k error correcting alphabet is represented by a set of vertices, such that each pair of vertices is separated by a distance at least $\sqrt{2k + 1}$

Let $K_0(D, k)$ be the largest number of letters which a D dimensional

[2] R. W. Hamming, "Error Detecting and Error Correcting Codes," *Bell System Tech. J.*, **29**, pp. 147–160, 1950.

k error correcting alphabet can contain. Except when $k = 1$, there is no general method for constructing an alphabet with $K_0(D, k)$ letters, nor is $K_0(D, k)$ known as a function of D and k. Crude upper and lower bounds for $K_0(D, k)$ are given by the following theorem.

Theorem 1. The largest number of letters $K_0(D, k)$ satisfies

$$\frac{2^D}{N(D, 2k)} \leq K_0(D, k) \leq \frac{2^D}{N(D, k)} \tag{3}$$

where

$$N(D, k) = \sum_{r=0}^{k} C_{D,r}$$

is the number of sequences of D digits which differ from a given sequence in 0, 1, $\cdots$, or k places.

Proof

The upper bound is due to R. W. Hamming and is proved by noting that for each letter S of a k error correcting alphabet there are $N(D, k)$ possible received sequences which will be interpreted as meaning S. Hence $N(D, k)\, K_0(D, k) \leq 2^D$, the total number of sequences.

The lower bound is proved by a random construction method. Pick any sequence S_1 for the first letter. There remain $2^D - N(D, 2k)$ sequences which differ from S_1 in $2k + 1$ or more places. Pick any one of these S_2 for the second letter. There remain at least $2^D - 2N(D, 2k)$ sequences which differ from both S_1 and S_2 in $2k + 1$ or more places. As the process is continued, there remain at least $2^D - rN(D, 2k)$ sequences, which differ in $2k + 1$ or more places from $S_1, \cdots, S_r$, from which S_{r+1} is chosen. If there are no choices available after choosing S_K, then $2^D - KN(D, 2k) \leq 0$ so the alphabet $(S_1, \cdots, S_K)$ has at least as many letters as the lower bound (3).

For all the simple cases (D and k not very large) investigated so far the upper bound is a better estimate of $K_0(D, k)$ than the lower bound. The upper and lower bounds differ greatly, as may be seen from a quick inspection of Table I. For example, in the case of a ten dimensional two error correcting alphabet, the bounds are 2.7 and 18.3.

2. Efficiency Graph

The first step in constructing an efficiency graph for comparing alphabets is to decide on what constitutes reliable transmission. The criterion used here is that on the average no more than one letter in 10^4 shall be misinterpreted.

TABLE I

TABLE OF $2^D/N(D, k)$

k =	1	2	3	4	5	6	7
$D =$ 3	2						
4	3.2						
5	5.3	2					
6	9.1	2.9					
7	16	4.4	2.9				
8	28.4	6.9	2.8				
9	51.2	11.1	3.9	2			
10	93.1	18.3	5.8	2.7			
11	170.7	30.6	8.8	3.6	2		
12	315.8	51.8	13.7	5.2	2.6		
13	585.2	89.0	21.6	7.5	3.4	2	
14	1092.3	154.4	34.9	11.1	4.7	2.5	
15	2048	270.8	56.8	16.8	6.6	3.3	2

Missing entries are numbers between 1 and 2.

This sort of criterion might be appropriate for a channel transmitting English text. For other messages it is not always appropriate. For example, if the messages are telephone numbers, one would naturally require that the probability of mistaking a telephone number be small, say less than 10^{-4}. If the telephone numbers are L decimal digits long, and if the alphabet has K different letters in it (so that it takes about $L \log 10/\log K$ letters to make up a telephone number) the probability of making a mistake in a single letter should be required to be less than about

$$\frac{10^{-4} \log K}{L \log 10}$$

which gives alphabets with large K an advantage over alphabets with small K.

Since the probability that exactly r binary digits out of D shall be received incorrectly is $C_{D,r}p^r (1 - p)^{D-r}$, we achieve the required reliability with a D-dimensional k-error correcting alphabet provided p satisfies

$$\sum_{r=k+1}^{D} C_{D,r}p^r(1 - p)^{D-r} \leq 10^{-4}. \tag{4}$$

The value of p which makes the inequality hold with the equals sign determines the noisiest channel over which the alphabet can be used safely.

Let K be the number of different letters in the alphabet. Then the rate in bits per digit at which information is being recieved is

$$R = \frac{\log K}{D}. \tag{5}$$

In Equation (5) we have neglected a term which takes account of the information lost due to channel noise. This is legitimate because all but 10^{-4} of the letters are received correctly.

The worst tolerable probability p of (4) and the rate R of Equation (5) determine the noise combating ability of an alphabet. To compare different alphabets one may represent them as points on an efficiency graph of R versus p. Fig. 1 is an efficiency graph on which the values (p, R) for a number of simple error correcting alphabets have been plotted. Each point on the graph is labelled with the two numbers k, D in that order. The alphabets represented were not found by any systematic process and are not all proved to be best possible (i.e., to have the largest K) for the stated values of k and D. Fortunately, R depends on K only logarithmically so that it is not likely the points representing the best possible alphabets lie far away from the plotted points.

The solid line represents the curve

$$R = C = 1 + p \log p + (1 - p) \log (1 - p).$$

According to Shannon's theorems, all alphabets are represented by points lying below this line.

The efficiency graph only partially orders the alphabets according to

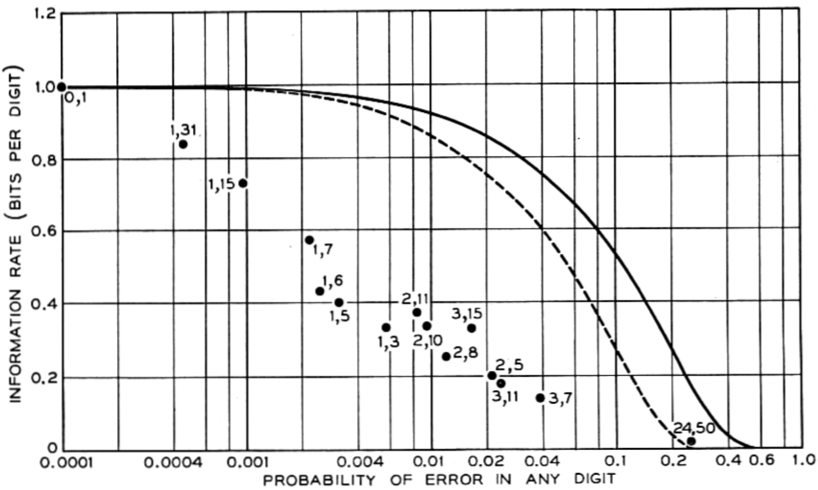

Fig. 1—Probability of error in a letter is 10^{-4}.

their invulnerability to noise. For example, it is clear that the alphabet 3, 15 is better than 2, 8. However, without further information about the channel, such as knowledge of p, there is no reasonable way of choosing between 3, 15 and 3, 7.

3. *Large Alphabets*

We have been unable to prove that there are error correcting alphabets which signal at rates arbitrarily close to C while maintaining an arbitrarily small probability of error for any letter. A result in this direction is the following theorem.

Theorem 2. Let any positive ϵ and δ be given. Given a channel with $p < \frac{1}{4}$ there exists an error correcting alphabet which can signal over the channel at a rate exceeding $R_0 - \epsilon$ where

$$R_0 = 1 + 2p \log 2p + (1 - 2p) \log (1 - 2p)$$

bits per digit and for which the probability of error in any letter is less than δ.

Proof

The probability of error in any letter is the sum on the left of (4). This is a sum of terms from a binomial distribution which, as is well known, tends to a Gaussian distribution with mean Dp and variance $Dp(1 - p)$ for large D. Hence there is a constant $A(\delta)$ such that all k error correcting alphabets with sufficiently large D have a letter error probability less than δ provided

$$k \geq Dp + A(\delta)\,(Dp(1 - p))^{1/2} \tag{6}$$

Let $k(D)$ be the smallest integer which satisfies (6) and consider an alphabet which corrects $k(D)$ errors and contains $K_0(D, k(D))$ letters. By Equation (5) and the lower bound of Theorem 1, this alphabet signals at a rate $R(D)$ satisfying

$$1 - \frac{1}{D} \log N(D, 2k(D)) \leq R(D).$$

Since $p < \frac{1}{4}$, $2k(D) < D/2$ for large D and hence

$$N(D, 2k(D)) < (2k(D) + 1)C_{D,2k(D)}.$$

Then an application of Stirling's approximation for factorials shows that as $D \to \infty$

$$1 - \frac{1}{D} \log N(D, 2k(D)) \to R_0 .$$

Hence by taking D large enough one obtains an alphabet with rate exceeding $R_0 - \epsilon$ and letter error probability less than δ.

The rate R_0 appears on the efficiency graph as a dotted line.

It has not been shown that no error-correcting alphabet has a rate exceeding R_0. In fact, one alphabet which exceeds R_0 in rate is easy to construct. If the noise probability p is greater than $\frac{1}{4}$, then $R_0 = 0$. The alphabet with just two letters

$$0\,0\,0\,0 \ldots 0$$

and

$$1\,1\,1\,1 \ldots 1$$

will certainly transmit information at a (small) positive rate, and with a 10^{-4} probability of errors if D is large enough, as long as $p < \frac{1}{2}$.

Using a more refined lower bound for $K_0(D, k)$ it might be shown that there are error-correcting alphabets which signal with rates near C. If one repeats the calculation that led to R_0 using the upper bound (3) (which seems to be a better estimate of the true $K_0(D, k)$) instead of the lower bound (3), one is led to the rate C instead of R_0.

The condition (4) is more conservative than necessary. The structure of the alphabet may be such that a particular sequence of more than k errors may occur without causing any error in the final letter. This is illustrated by the following simple example due to Shannon: the alphabet with just two letters

$$\begin{array}{c} 0\,0\,0\,0\,0\,0 \\ 1\,1\,1\,0\,0\,0 \end{array}$$

corrects any single error but also corrects certain more serious errors such as receiving 0 0 1 1 1 1 for 0 0 0 0 0 0. An alphabet designed for practical use would make efficient enough use of the available sequences so that any sequence of much more than k errors causes an error in the final letter; the random alphabets constructed above probably do not. If this kind of error were properly accounted for, the rate R_0 could be improved, perhaps to C.

4. Other Discrete Channels

If instead of transmitting just 0's and 1's the channel can carry more digits

$$0, 1, 2, \cdots, n$$

a similar theory can be worked out. The simplest kind of noise in this channel changes a digit into any one of the n other possible numbers with probability p/n. Then the capacity of the channel is

$$C = \log(n+1) + p \log \frac{p}{n} + (1-p) \log (1-p).$$

Error-correcting alphabets for this channel can also be constructed and the criterion (4) for good transmission remains unchanged. The proof of theorem 1 can be repeated with little change using

$$N(D, k) = \sum_{r=0}^{k} C_{D,r} n^r$$

as the number of sequences which can be reached after k or fewer errors [the terms 2^D in (1) and (3) are replaced by $(n+1)^D$]. Once more, using the lower bound, one finds an expression for R_0 which is the same as the one for C but with p replaced by $2p$.

Part II

THE LOW PASS FILTER

1. Encoding and Detection

If $f(t)$ is a signal emerging from a low pass filter (so that its spectrum is confined to the frequency band $|\nu| < W$ cycles per second) then $f(t)$ has a special analytic form given by the sampling theorem[3]

$$f(t) = \sum_{m=-\infty}^{\infty} f\left(\frac{m}{2W}\right) \frac{\sin \pi (2Wt - m)}{\pi(2Wt - m)} \tag{7}$$

Thus the signal is completely determined by the sequence of sample values $f(m/2W)$. The average power of the signal $f(t)$ is measured by

$$P = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f^2(t)\, dt$$

which can be expressed in terms of the sample values as follows

$$P = \lim_{M\to\infty} \frac{1}{2M} \sum_{m=-M}^{M} f^2\left(\frac{m}{2W}\right). \tag{8}$$

As in Part I, consider a message source producing a sequence of letters from an alphabet of K equally likely letters. To transmit this information over the low pass filter we must encode the sequence into a function

[3] C. E. Shannon, "*Communication in the Presence of Noise,*" *Proc. I. R. E.*, **37**, pp. 10–21, Jan. 1949.

$f(t)$ of the form (7), or in other words into a sequence of sample values $f(m/2W)$. To do this, we construct a new alphabet containing K letters which are different sequences of real numbers of some fixed length, say D places. When we let the letters of the new alphabet correspond to letters of the old one the message is translated into a sequence of real numbers which we use for the sequence $f(m/2W)$.

If the K letters of the sequence alphabet are

$$\begin{array}{l} S_1:\ a_{11}, \cdots, a_{1D} \\ S_2:\ a_{21}, \cdots, a_{2D} \\ \quad \vdots \\ S_K:\ a_{K1}, \cdots, a_{KD}, \end{array}$$

the expression (8) for the average power of the function $f(t)$ becomes

$$P = \frac{1}{DK}(d_1^2 + d_2^2 + \cdots + d_K^2) \tag{9}$$

where

$$d_i^2 = \sum_{j=1}^{D} a_{ij}^2.$$

If the D numbers in the sequence S_i are regarded as coordinates of a point in Euclidean D dimensional space, d_i^2 represents the square of the distance from the point representing S_i to the origin.

When $f(t)$ is transmitted, the received signal will be $f(t) + n(t)$ where $n(t)$ is some (unknown) white Gaussian noise signal. The noise signals $n(t)$ are characterized by the fact that their sample values $n(m/2W)$ are independently distributed according to Gaussian laws. That is,

$$\text{Prob}\left(n\left(\frac{m}{2W}\right) \leq X\right) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{X} e^{-y^2/2\sigma^2}\, dy. \tag{10}$$

The variance σ^2 of the distribution of noise samples is, by an application of (8), the power of this ensemble of noise signals.

At the receiving end of the channel, there is a detector which observes each block of D sample values $f(m/2W) + n(m/2W)$ and tries to decide which one of the K letters $S_1, \cdots, S_K$ was sent. In terms of the geometric picture, the detector divides all of D dimensional space into K non-overlapping regions $U_1, \cdots, U_K$ with the property that, if the D received sample values are represented by a point in U_i, the detector decides that S_i was sent. By Equation (10), the probability that the detector picks the wrong letter when S_i is sent is

$$p_i = \frac{1}{(2\pi)^{D/2}\sigma^D} \int \int_{\bar{U}_i} \cdots \int e^{-r_i^2/2\sigma^2} \, dy_1 \cdots dy_D \tag{11}$$

where $\bar{U}_i$ is the set of all points not in U_i and r_i is the distance from $(y_1, \cdots, y_D)$ to the point representing S_i.

For any given alphabet the best possible detector (in the sense that it minimizes the average probability of making an error in guesssing a letter) is called a *maximum likelihood detector*. The region U_i for a maximum likelihood detector consists of all points $(y_1, \cdots, y_D)$ which are closer to the point S_i than to any other letter point $S_j (r_i < r_j$ for all $j \neq i)$. To prove that this choice of U_i is best possible consider any other detector such that U_i contains a set V of points in which $r_i > r_j$. A direct calculation shows that the detector obtained by removing V from U_i and making V part of U_j has a smaller probability of error per letter. The set of points equidistant from two given points is a hyperplane. The region U_i of a maximum likelihood detector is a convex region bounded by segments of the hyperplanes

$$r_i = r_1, \qquad r_i = r_2, \cdots .$$

To compare signalling alphabets under the most favorable possible circumstances, we always compute letter error probabilities assuming that the detector is a maximum likelihood detector.

2. *Computation of error probabilities*

Exact evaluation of the letter error probability integral (11) is impossible except in a few special cases. Fortunately we are only interested in (11) when σ is small enough in comparison to the size of U_i to make the integral small. Then fairly accurate approximate formulas can be derived.

Theorem 3. Let R_{ij} be the distance between letter points S_i and S_j. Then

$$1 - \prod_{j \neq i} (1 - Q_{ij}) \leq p_i \leq \sum_{j \neq i} Q_{ij} \tag{12}$$

where

$$Q_{ij} = \frac{1}{\sqrt{2\pi}} \int_{R_{ij}/2\sigma}^{\infty} e^{-x^2/2} \, dx.$$

The proof of Theorem 3 follows from the fact that Q_{ij} is the probability that, when S_i is transmitted, the received sequence will be closer to S_j than to S_i.

In the cases to be computed Q_{ij} is a rapidly decreasing function of R_{ij} and the only terms worth keeping in (12) are the ones for which R_{ij} is the smallest of the numbers $R_{i1}, \cdots, R_{iK}$. Moreover since the Q_{ij} are all small enough so that the upper and lower bounds differ only by a few per cent, the upper bound is a good approximation to p_i. Then a simple approximate formula for the average letter error probability $p = (p_1 + \cdots + p_K)/K$ is

$$p = \frac{N}{\sqrt{2\pi}} \int_{r_0/\sigma}^{\infty} e^{-x^2/2} \, dx \tag{13}$$

where $2r_0$ is the smallest of the $K(K-1)/2$ distances R_{ij} and N is the average over all letters in the alphabet of the number of letter points which are a distance $2r_0$ away.

3. Efficiency graph

The efficiency graph to be described was constructed originally to compare alphabets for signalling telephone numbers of length equal to ten decimal digits. It was desired that on the average only one telephone number in 10^4 should be received incorrectly. As described in Part I section 2, if the telephone numbers are encoded into sequences of letters from an alphabet of K letters, we must require that the average probability of error in any letter be

$$p = 10^{-5} \log_{10} K \tag{14}$$

or smaller.

Given an alphabet, one can compute with the help of (13) and (14) and a table of the error integral the largest value of the noise power σ^2 which can be tolerated. The average power of the transmitted signal is P given by Equation (9). Hence we can compute the smallest signal to noise ratio

$$Y = P/\sigma^2 \tag{15}$$

which will be satisfactory.

A letter containing $\log K$ bits of information is transmitted during an interval of $D/2W$ seconds. Hence the rate at which information is received is

$$R = \frac{2W \log K}{D} \tag{16}$$

bits per second. Again Equation (16) ignores a term representing information lost due to channel noise which is negligible because the error probability is low.

The efficiency graph, Fig. 2, is a chart on which the signal to noise ratio Y in db [computed from Equation (15)] is plotted against the signalling rate per unit bandwidth $R/W = (2 \log K)/D$ for different alpha-

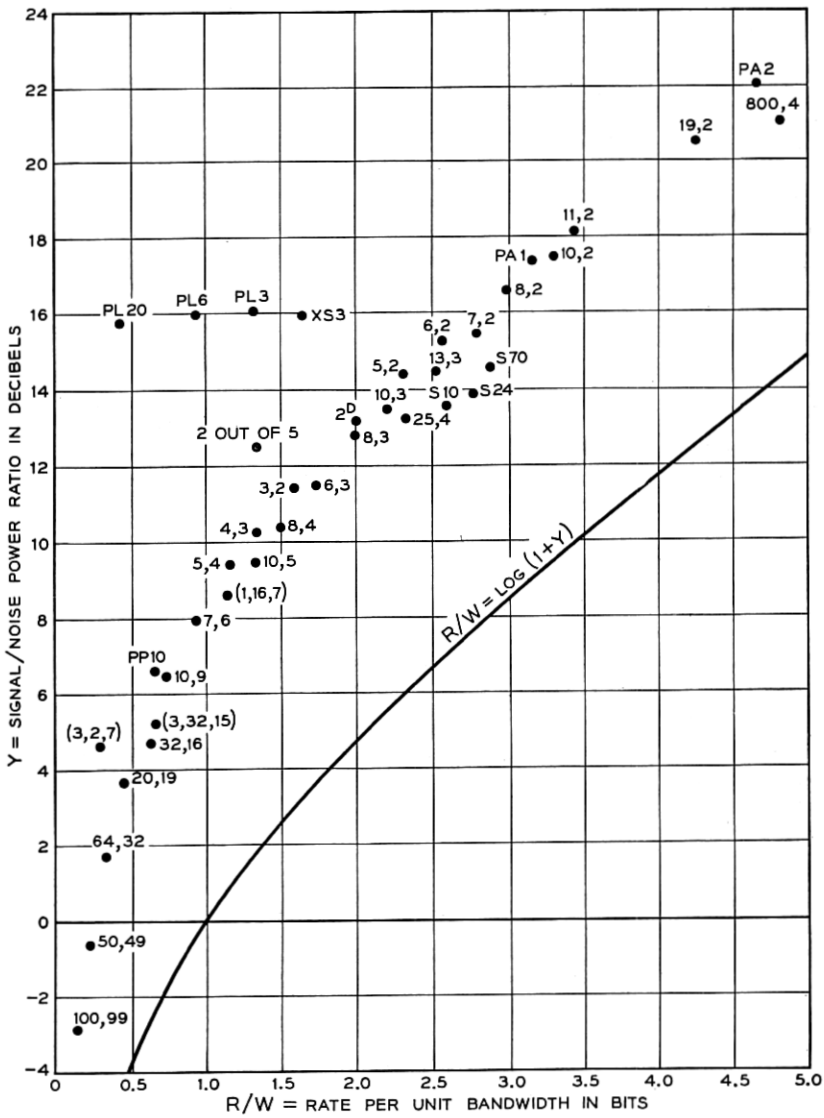

Fig. 2—Probability is 10^{-4} that an error is made in a 10 digit decimal number.

bets. An alphabet is considered poor if its point on the efficiency graph lies far above the ideal curve $R/W = C/W = \log(1 + Y)$.

4. The alphabets

The alphabets which appear on the efficiency graph are the following:

excess three (*XS3*): the ten sequences of 4 binary digits which represent 3, 4, $\cdots$, and 12 in binary notation;

two out of five: the ten sequences of five binary digits which contain exactly two ones;

pulse position (*PP10*): the ten sequences of ten binary digits which contain exactly one one;

2^D *binary*: all of 2^D sequences of D binary digits.

pulse amplitude (*PAn*): the $2n + 1$ sequences of length 1 consisting of $-n, -n + 1, \cdots, n$. This alphabet gives rise to a sort of quantized amplitude modulation.

pulse length (*PLn*): the $n + 1$ sequences of n binary digits of the form $11 \cdots 10 \cdots 0$, i.e., a run of ones followed by a run of zeros.

Minimizing alphabets (K, D): The above alphabets are taken from actual practice. They are convenient because, aside from PAn, they require a signal generator with only two amplitude levels. If we ignore ease of generating the signals as a factor, a great many geometric arrangements of points suggest themselves as possible good alphabets. The principle by which one arrives at good alphabets may be described as follows. When a D and K have been determined which give the desired information rate R [by Equation (16)] try to arrange the K letter points in D dimensional space in such a way that the distances between pairs of points are all greater than some fixed distance and that the average of the K squared distances to the origin is minimized. By Equations (9) and (13) it is seen that, apart from the small influence of the factor N, this process must minimize the signal to noise ratio Y required.

Ordinarily it is difficult to prove that a configuration is a minimizing one. Even to recognize a configuration which leads to a relative minimum (*i.e.* a minimum over all nearby configurations) is not always easy. The eight vertices of a cube, for example, do not give a relative minimum. Consequently, most of the alphabets to be described are only conjectured to be "best possible." Each of them satisfies one necessary requirement of minimizing alphabets that the centroid of the point configuration (assuming a unit mass at each letter point) lies at the origin. That this condition is necessary follows from the easily derived identity

$$r_2^2 = r_1^2 - R_0^2$$

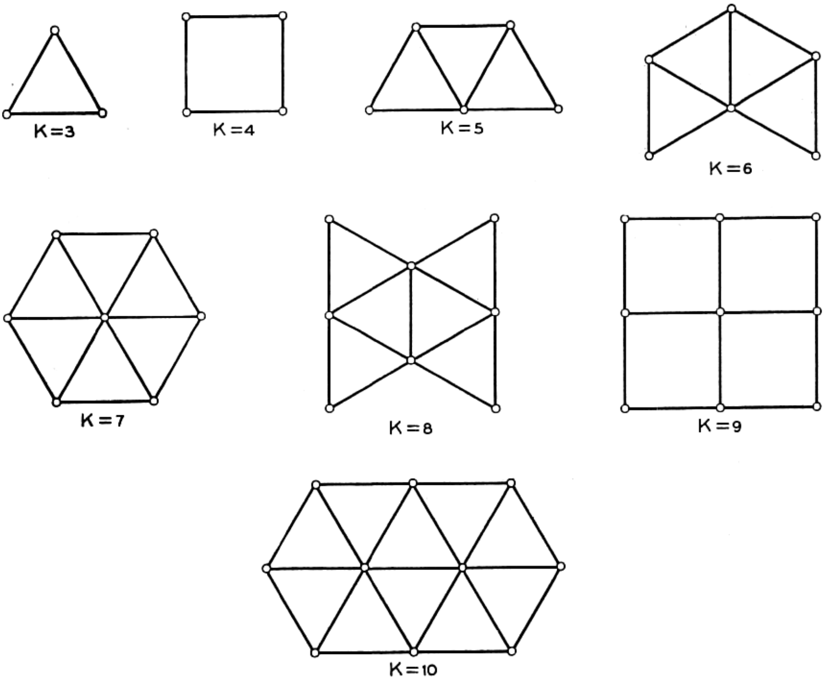

Fig. 3—Two dimensional alphabets.

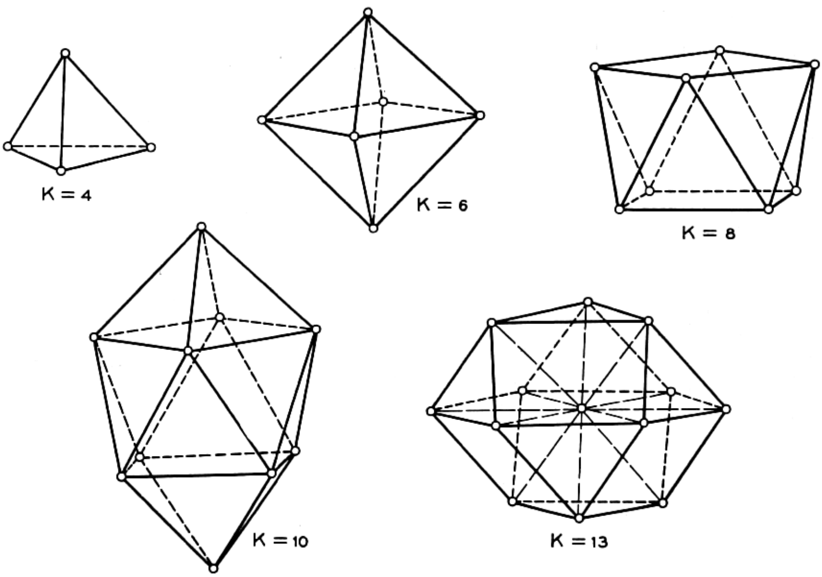

Fig. 4—Three dimensional alphabets.

where r_1 is the rms distance from the origin to the points of a configuration A, R_0 is the distance from the origin to the centroid of A, and r_2 is the rms distance from the points of A to the centroid of A.

In plotting points on the efficiency graph the notation K, D is used for the best K-letter D-dimensional alphabet which has been found. The arrangement of points for various K, 2 and K, 3 alphabets is given in Figs. 3 and 4. In these figures two points are joined by a straight line if the distance between them is 1 (which is the value we have adopted for the minimum allowed separation $2r_0$). Although not shown, the origin is always at the centroid of the figure. To aid interpretation of these diagrams we have included Fig. 5 which demonstrates how all the signals of a typical alphabet can be generated. The functions of time shown in

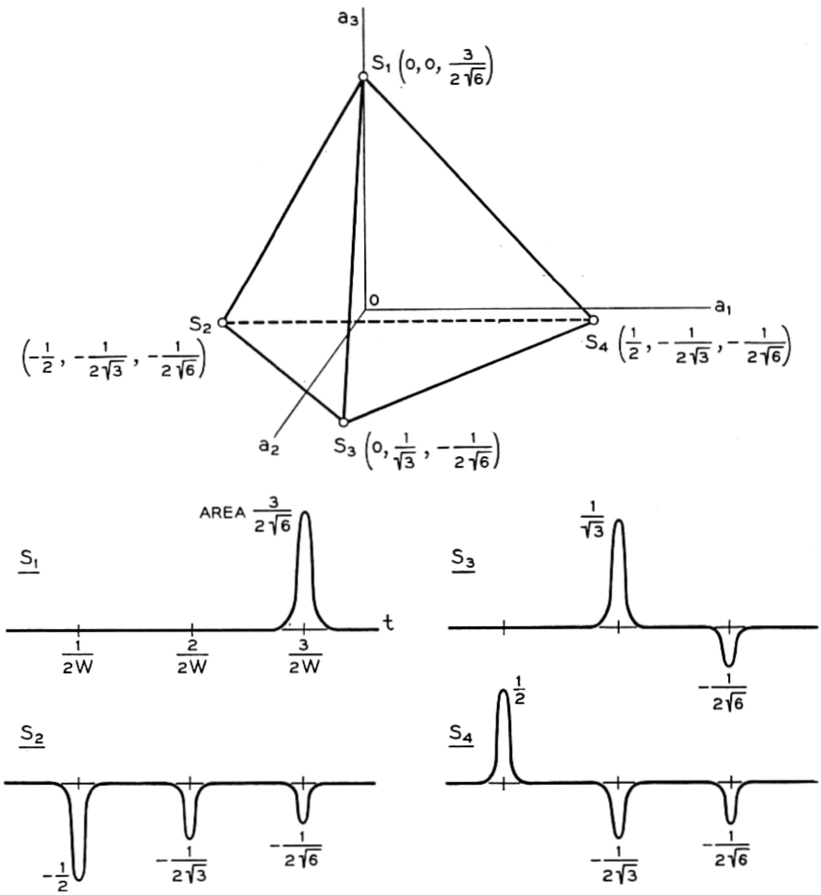

Fig. 5—Generation of the 4,3 code signals.

Fig. 5 are not the code signals themselves but impulse functions which are to be passed through a low pass filter with cutoff at W c.p.s. to form the code signals.

The best possible higher dimensional alphabets can be described more easily verbally than pictorially. In four dimensions we have found four alphabets.

The *25,4* alphabet consists of the origin and all 24 points in 4 dimensional space having two coordinates equal to zero and the remaining two equal to $1/\sqrt{2}$ or $-1/\sqrt{2}$. Each of the 24 points lies a unit distance away from the origin and its 10 other nearest neighbors; they are, in fact, the vertices of a regular solid. This alphabet has an advantage beyond its high efficiency. The code signals are composed entirely of positive and negative pulses of fixed energy and so should be easier to generate than most of the other codes which appear in this paper.

The *800, 4* alphabet is constructed in the following way: Consider a lattice of points throughout the entire 4-dimensional space formed by taking all the linear combinations with integer coefficients of a basic set of four vectors. That is, the lattice points are of the form $C_1v_1 + C_2v_2 + C_3v_3 + C_4v_4$ where $C_1, \cdots, C_4$ are integers and the v_i are the four given vectors. In connection with our problem it is of interest to know what lattice, (i.e. what choice of v_1, v_2, v_3, v_4) has all lattice points separated at least unit distance from one another and at the same time packs as many points as possible into the space per unit volume. When a solution to this "packing problem" is known, it is clear that a good alphabet can be obtained just by using all the lattice points which are contained inside a hypersphere about the origin as the letter points. Many of the two dimensional alphabets illustrated in the sketches are related in this way to the corresponding two dimensional packing problem (which is solved by letting v_1 and v_2 be a pair of unit vectors 60° apart). A solution to the four dimensional packing problem is afforded by

$$v_1 = \frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}},\ 0,\ 0$$

$$v_2 = \frac{1}{\sqrt{2}},\ 0,\ \frac{1}{\sqrt{2}},\ 0$$

$$v_3 = \frac{1}{\sqrt{2}},\ 0,\ 0,\ \frac{1}{\sqrt{2}}$$

$$v_4 = 0,\ \frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}},\ 0.$$

This lattice contains two points per unit volume (twice as dense as the cubic lattice in which $v_1, \cdots, v_4$ are orthogonal to one another) and each point has 18 nearest neighbors. A hypersphere of radius 3 about the origin has a volume $(\pi^2/2)3^4$, about 400. Thus it contains about 800 lattice points. Take these as the code points of the 800, 4 code. Their average squared distances from the origin can be estimated as

$$\frac{\int_0^3 r^5\,dr}{\int_0^3 r^3\,dr} = \frac{2}{3}(3)^2 = 6.$$

N in Equation (13) may be estimated at 18; this is conservative because some lattice points outside the sphere are being counted.

The two remaining four dimensional alphabets belong to two families of D-dimensional alphabets.

The 4, 3; 5, 4; $\cdots$; $D + 1$, D $\cdots$ alphabets are the vertices of the simplest regular solid in D-dimensional space. For example, 4, 3 is a tetrahedron. Such a solid can be constructed from $D + 1$ vertices whose coordinates are the first $D + 1$ rows of the scheme

$$\begin{array}{cccccc}
0 & 0 & 0 & 0 & 0 & \cdots \\
1 & 0 & 0 & 0 & 0 & \cdots \\
\frac{1}{2} & \frac{3}{2\sqrt{3}} & 0 & 0 & 0 & \cdots \\
\frac{1}{2} & \frac{1}{2\sqrt{3}} & \frac{4}{2\sqrt{6}} & 0 & 0 & \cdots \\
\frac{1}{2} & \frac{1}{2\sqrt{3}} & \frac{1}{2\sqrt{6}} & \frac{5}{2\sqrt{10}} & 0 & \cdots \\
\frac{1}{2} & \frac{1}{2\sqrt{3}} & \frac{1}{2\sqrt{6}} & \frac{1}{2\sqrt{10}} & \frac{6}{2\sqrt{15}} & \cdots \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdots
\end{array}$$

The vertices all lie a distance $\sqrt{D/2(D+1)}$ from the centroid of the figure.

6, 3; 8, 4; $\cdots$; $2D$, D, $\cdots$ are obtained by placing a point wherever any positive or negative coordinate axis intersects the sphere of radius

$1/\sqrt{2}$ about the origin. Thus it follows that 6, 3 consists of the vertices of an octohedron.

Error correcting alphabets $((k, K, D))$: The error correcting alphabets discussed in Part I can be converted into good alphabets for this channel by replacing all digits which equalled 0 by -1. Three error correcting alphabets appear on the chart; each is labelled by three numbers signifying (k, K, D).

Slepian alphabets (SD): Using group theoretic methods, D. Slepian has attempted to construct families of alphabets which signal at rates approaching C. Although this goal has not yet been reached, families of alphabets depending on the parameter D have been found which approach the ideal curve to within 6.2 db and then get worse as $D \rightarrow \infty$. In the simplest of these families of alphabets, $D = 2m$ is even and the letters consist of all the $2^m C_{2m,\, m}$ sequences containing m zeros, the remaining places being filled by ± 1. The best alphabet in this family is the one with $D = 24$. It lies 6.23 db away from the ideal curve and contains 1.1×10^{10} letters. The alphabets of this family for $D = 10, 24,$ and 70 appear on the efficiency graph labelled $S10$, $S24$, and $S70$.

The conclusion to which one is forced as a result of this investigation is that one cannot signal over a channel with signal to noise level much less than 7 db above the ideal level of Equation (2) without using an unbelievably complicated alphabet. No ten digit alphabet tolerates less than 7.7 db more than the ideal signal to noise ratio.

It would be interesting to know more about good higher dimensional alphabets. They are very much more difficult to obtain. The regular solids, which provided some good alphabets in 3 and 4 dimensions, provide nothing new in 5 or more dimensions; there are only three of them and they correspond to our $D + 1, D$; $2D, D$, and 2^D binary alphabets. Worse still, the packing problem also becomes unmanageable after dimension 5.

ACKNOWLEDGMENT

The author wishes to thank R. W. Hamming, L. A. MacColl, B. McMillan, C. E. Shannon, and D. Slepian for many helpful suggestions during the investigation summarized by this paper.