possible for larger word lengths, with both R_1 and R_3 approaching 1 bit per symbol, the capacity of the binary noiseless channel, as L increases [6].

Acknowledgment

The author gratefully acknowledges several discussions of the comma-free coding problem with Prof. S. W. Golomb and Prof. L. Welch. The original suggestion that comma-free codes might be constructed from synchronizable codes was made to the author by Prof. S. Even. The author also wishes to thank W. L. Eastman for several stimulating comments on this work.

References

[1] S. W. Golomb, B. Gordon, and L. R. Welch, "Comma-free codes," *Can. J. Math.*, vol. 10, pp. 202-209, 1958.
[2] S. W. Golomb and B. Gordon, "Codes with bounded synchronization delay," *Inform. and Control*, vol. 8, pp. 355–376, August 1965.
[3] R. A. Scholtz, "Codes with synchronization capability," *IEEE Trans. Information Theory*, vol. IT-12, pp. 135–142, April 1966.
[4] W. L. Eastman and S. Even, "On synchronizable and PSK-synchronizable block codes," *IEEE Trans. Information Theory*, vol. IT-10, pp. 351–356, October 1964.
[5] W. L. Eastman, "On the construction of comma-free codes," *IEEE Trans. Information Theory*, vol. IT-11, pp. 263–266, April 1965.
[6] R. A. Scholtz, "A noiseless coding theorem for synchronizable codes," *Proc. 1968 Hawaii Internat'l Conf. on System Sciences* (University of Hawaii).

Rank Permutation Group Codes Based on Kendall's Correlation Statistic

HENRY D. CHADWICK, MEMBER, IEEE, AND LUDWIK KURZ, MEMBER, IEEE

***Abstract*—A coding scheme based on the properties of rank vectors is presented. The new codes are based on the theory of permutation groups by introducing a new notation for the group operation that simplifies the generation and decoding of desirable rank codes. The use of group theory is made possible by the introduction of the Kendall correlation coefficient as a measure of the distance between code words. This technique provides a method for the choice of rank vector code words superior to those that have been proposed in the past. Much of the terminology used in block coding can also be used to describe rank vector codes, but the actual quantities involved are quite different. The rank vector codes discussed in the paper offer the advantage of low sensitivity of the probability of error to the noise distribution because of the nonparametric character of rank vector detection schemes. Bounds that have been verified by extensive computer simulation have been derived for the probability of error.**

Introduction

IN COMMUNICATION systems the correlation between a received sample vector and the various possible transmitted signal vectors is often used to determine which signal has been transmitted. If one set of observations of a random process is denoted by $\{x_i\} = X$, $i = 1, 2, \cdots, n$, and a second set of observations by $\{y_i\} = Y$, then a commonly used measure of their correlation is

$$X \cdot Y = \sum_{i=1}^{n} x_i y_i$$

or the scalar product of the two vectors. The solution of the optimum signal detection problem in the Neyman–Pearson sense for a signal embedded in white additive Gaussian noise gives a test statistic that can be reduced to a correlation of this type.

Kendall [7] has investigated the extension of the concept of correlation to cases in which the ranks (appropriate rank vectors) of the observations instead of the values themselves are correlated. The use of rank vectors can be justified in the signal detection problem by considering non-Gaussian noise, particularly impulsive noise [3], [4]. Should an impulsive transient greatly alter the value of a single observation in the sample vector, the effect on the rank correlation will be considerably less than the effect on the time correlation. Thus, rank correlation can give a better picture of the "closeness" of the samples when the noise distribution is severe than the picture given by time correlation

In this paper, the good properties of rank correlation detectors and the theory of permutation groups are used to generate a class of rank permutation codes which preserve the good qualities of block codes and the robustness (low sensitivity to the unknown noise distribution) of the nonparametric detection procedures. The rank correlation detector decides that a signal has been transmitted on the basis of the rank correlation measured between the

Manuscript received April 13, 1967; revised June 19, 1968. This research was sponsored in part by the National Science Foundation under Grant GK-1075.

H. D. Chadwick was with the Department of Electrical Engineering, School of Engineering and Science, New York University, Bronx, N. Y. He is now with the Jet Propulsion Laboratory, Telecommunication Division, Pasadena, Calif. 91103.

L. Kurz is with the Department of Electrical Engineering, School of Engineering and Science, New York University, Bronx, N. Y. 10453.

received sample vector and the possible transmitted sample vectors. Stoll and Kurz [1] have previously investigated a scheme of this type using a metric equivalent to the Spearman coefficient of rank correlation. By using, instead, the Kendall rank correlation coefficient in this paper, it has become possible to develop a simplified scheme for generating rank vector group codes. Here the theory of permutation groups has been useful. Much of the terminology used in the description of these rank vector codes is the same as that for block codes. The actual quantities involved are quite different, however.

The Mathematical Background for the Rank Coding Problem

Definition of Symbols

The coding problem considered in the paper consists of selecting for the transmission of messages one of the subsets of the $n!$ arrangements of the n digits $1, 2, \cdots, n$. The coded message will be specified by the vector $R_0 = \{r_i^0\}$; $i = 1, 2, \cdots, n$, where the coordinate r_i^0 is the ith digit in the arrangement. The coded message is then used to modulate a transmitted signal $S = \{s_i\}$, $i = 1, 2, \cdots, n$, where the coordinate s_i is the value of the transmitted signal at time t_i. At the receiver the signal is

$$x_i = \gamma s_i + n_i \tag{1}$$

where γ is a gain constant and n_i is the value of the additive noise at the sampling instant t_i. The n_i are assumed to be independent of each other and identically distributed with probability density function $f(n)$. The detector calculates the rank vector of the signal X to obtain $R = \{r_i\}$, where r_i is the relative ranking of the component x_i in the vector X.[1] The decoder then decides which message has been transmitted on the basis of this rank vector, or, in a feedback system, may request a repeat transmission.

Transmitted Signal Design

Stoll and Kurz [1] have shown[2] that, in a system of this type, the optimum signal when the transmitter is average power limited and Gaussian noise is present in the channel is

$$\Delta S = C = \text{constant}, \tag{2}$$

where ΔS is the difference between two adjacent signal levels, corresponding to two components differing by 1 in the code vector R_0. The transmitted signal is thus[3]

$$s_i = Cr_i^0 \tag{3}$$

at the sampling intervals. This transmitted signal form will be assumed in the remainder of this paper.

[1] If x_i is the algebraically smallest component, $r_i = 1$. If x_i is the 2nd smallest, $r_i = 2$. If x_i is the largest, $r_i = n$.

[2] In [1], proof of optimality is given for $n = 3$. Extensive computer simulations support the conjecture that it is true for larger n. Here, "optimal" means that the average probability of error is minimized when all the signals are a priori equally likely.

[3] If an average power limitation is of importance, one subtracts from every signal the mean $C\,n(n+1)/2$.

Definition of an Error

A transmission error will be said to occur if the rank vector associated with the corresponding received and transmitted signals are different. The most probable error will be assumed to be an interchange between two digits differing by 1, e.g., the arrangement [1423] is changed into [2413] by interchanging the digits 1 and 2 and [1423] $\rightarrow$ [1324] by interchanging 3 and 4.

Definition of Group Operation

We denote an arrangement of the integers $1, 2, \cdots, n$ by an expression in square brackets, e.g., [2 1 4 3]. There are $n!$ such arrangements. Let A and B be two such arrangements. We can think of obtaining B from A by the application of a substitution or permutation operator to A, $B = aA$. The operator a specifies the substitution that must be made for each integer appearing in A to convert it into B. For example, A = [2 3 1 4] is converted to B = [3 1 4 2] by replacing 2 by 3, replacing 3 by 1, replacing 1 by 4, and replacing 4 by 2. We write this operator as a = (4 3 1 2). In general the operator $(x_1x_2x_3 \cdots x_n)$ means replace the integer i by x_i, $i = 1, 2, \cdots, n$. We have [3 1 4 2] = (4 3 1 2)[2 3 1 4]. Note that the result of applying $(x_1x_2 \cdots x_n)$ to [1 2 3 $\cdots$ n] is $[x_1x_2 \cdots x_n]$.

The result of applying two substitutions (permutation operators) a and b in succession to an arrangement A of the integers is again an arrangement and hence can be regarded as the result of applying a new substitution c to A. If a is applied first to A, and then b applied to the result, we write $c = ba$. For example

$$\begin{aligned}(2413)(4312)[1234] &= (2413)[4312] \\ &= [3124] \\ &= (3124)[1234]\end{aligned}$$

so that

$$(2413)(4312) = (3124)$$

which is independent of the choice of the initial arrangement A = [1 2 3 4].

The collection of $n!$ operations $(x_1x_2 \cdots x_n)$ under the multiplication just defined forms a well-known group S_n, called the symmetric group on n letters. There is another common notation for permutations or substitutions. The operator $\{x_1x_2 \cdots x_i\}$ means replace x_1 by x_2, replace x_2 by $x_3, \cdots$, replace x_{j-1} by x_j, and replace x_j by x_1. $\{x_1x_2 \cdots x_j\}$ is called a cycle of length j. A cycle of length 2 is called a transposition. Any substitution can be written as a product of cycles on disjoint letters, e.g.,

$$(2143) = \{12\}\{34\}.$$

The product of cycles is defined by successive applications of the indicated operators, starting with the rightmost factor.

If a = (3412) and b = (4321) are two operators, then the multiplication of b by a is

$$ab = (3412)(4321) = (2143)$$

where the method of multiplication of permutation operators has been used. In terms of transpositions

$$a = (3412) = \{23\}\{34\}\{12\}\{23\}.$$

The product ab can also be formed by performing a series of transpositions on b, namely,

$$\begin{aligned}(3412)(4321) &= \{23\}\{34\}\{12\}\{23\}(4321)\\ &= \{23\}\{34\}\{12\}(4231)\\ &= \{23\}\{34\}(4132)\\ &= \{23\}(3142)\\ &= (2143).\end{aligned}$$

The Weight of a Permutation

The use of rank permutations as code words can be more easily systematized if the concept of weight is introduced. The weight of a permutation is defined as the Kendall correlation statistic[4] Q between the permutation and the identity element and will hereafter be denoted by w. The weight of the permutation $a = (x_1, x_2, \cdots, x_n)$ is given by

$$w = w(a) = Q = \sum_{j=2}^{n} \sum_{i=1}^{j-1} m_{ij} \tag{4}$$

where

$$m_{ij} = \begin{cases} 1 & \text{if } x_i > x_j \\ 0 & \text{if } x_i < x_j. \end{cases}$$

In addition, for any permutation a, the equality $w(a) = w(a^{-1})$ is satisfied.

The following theorem shows the equality between the weight of a permutation and the minimum number of transpositions involved in its decomposition into transpositions of integers differing by unity.

Theorem 1

The minimum number of interchanges of digits differing by 1 required to convert the identity element **1** to a given permutation is equal to the weight of the permutation.

Proof: Let $\alpha = \alpha_0 = (x_1, \cdots, x_n)$ be a permutation. If $\alpha_0 \neq (1, 2, \cdots, n) = \mathbf{1}$, then for some i and j with $i < j$, $x_i = x_j + 1$. Let $\alpha_1 = \{x_i, x_j\}\alpha_0$. Note that $w(\alpha_1) = w(\alpha_0) - 1$. Repeat the procedure for α_1, thus obtaining α_2 for which $w(\alpha_2) = w(\alpha_1) - 1 = w(\alpha_0) - 2$. Continuing in this way, we obtain α_j $(j = 1, 2, \cdots, w(\alpha_0))$. Since $w(\alpha_{w(\alpha_0)}) = 0$, $\alpha_{w(\alpha_0)} = \mathbf{1}$. Thus we have shown a procedure (not necessarily unique) for transforming α_0 to **1** with $w(\alpha_0)$ cycles of the form $\{i, i-1\}$. Further, since the application of a cycle $\{i, i-1\}$ can reduce the weight of a permutation by at most 1, we must apply at least $w(\alpha_0)$ such cycles to transform α_0 to **1**. Hence, $w(\alpha_0)$ is the *minimum* number of cycles $\{i, i-1\}$ which can transform α to **1**. Q.E.D.

[4] The Kendall correlation statistic is defined by (4). See also Kendall [7].

Distance Between Permutations

The Euclidean distance between rank vectors may be used as a distance metric, but it is of limited usefulness in generating practical rank vector codes.[5] In this paper, the Kendall correlation coefficient Q between two rank vectors will be used as the distance metric[6]. It is reasonable to assume that the most probable error consists of a transposition of digits differing by 1 (this represents a distance of 1), and that errors representing greater distances are less probable. Thus, the error probability will go down in some manner with increased distance among the rank vectors representing the signaling alphabet, where the distance between two rank vectors is defined as the minimum number of transpositions of digits differing by 1 that is necessary to transform one permutation into the other.

Let a, b, and c be any three permutations satisfying $ca = b$. The permutation c can be written as a product of $w(c)$ transpositions of integers differing by unity, and $w(c)$ is the minimal number of factors in any such representation of c. These transpositions convert a into b and, hence,

$$d(a, b) = w(c) = w(ba^{-1}) \tag{5}$$

where

$$ca = b.$$

The following theorem is then analogous to one for binary group codes and is very useful in determining the properties of subgroups as codes.

Theorem 2

The minimum distance between any two members of a subgroup of the symmetric group S_n is equal to the weight of the member with lowest weight, excluding the identity element.

Proof: A subgroup of S_n is any subset of the members of S_n that satisfy the group axioms. From the definitions of weight and distance, the weight of any element is equal to its distance from the identity element. If a, b, and c are members of the subgroup and $ab = c$, then the distance between b and c is equal to the number of transpositions performed on b by a. This, in turn, is equal to the weight of a. Hence, the distance between any two elements is equal to the weight of a third element that must also belong to the subgroup, and the lowest such weight will be equal to the smallest distance. Q.E.D.

From (5) it follows that for any two elements b, c of S_n

$$\begin{aligned}d(ba, ca) &= w[ca(ba)^{-1}] = w[caa^{-1}b^{-1}]\\ &= w(cb^{-1}) = d(b, c).\end{aligned} \tag{6}$$

All properties of the standard array follow from (6).

[5] Stoll and Kurz [1] have investigated the use of the Euclidean metric in connection with this problem.

[6] A metric over a vector space is a distance function that satisfies certain axioms. It can be easily shown that the Q statistic is a metric.

The Formation of Rank Vector Codes[7]

Subgroups of S_n

If a code is selected as a subgroup of S_n with the property that the minimum weight of the elements (except for the identity element) is some value w, then obviously, if $w - 1$ or fewer transpositions occur in the transmission of a code word, it will be possible to detect at the receiver that an error has occurred. An important property of subgroups, which is used extensively for binary group codes, is the following. All of the elements of a subgroup H are written in a row, with the identity element at the left. Select one of the group members not in the subgroup and write it under the identity element. Multiply each of the members of the subgroup by this element, called the coset leader, and place the resulting element under the corresponding subgroup element. If any elements remain in the group, select one of them as a second coset leader and repeat the procedure until all of the elements of the group are written down. A theorem from the theory of finite groups states that each element of the group will appear once and only once in the array [6, p. 17]. Obviously, for this to be true, both the number of coset leaders (including the identity element) and the number of elements in the subgroup must be divisors of $n!$, the number of elements in the group. A typical array is shown below:

	Subgroup			
Coset Leaders	(1234)	(3142)	(2413)	(4321)
	(2134)	(3241)	(1423)	(4312)
	(1324)	(2143)	(3412)	(4231)
	(1243)	(4132)	(2314)	(3421)

The coset leaders can be regarded as possible error patterns, e.g., (2134) is the interchange {12} of the digits 1 and 2. A possible received permutation is then the product of one of these coset leaders with one of the code words, e.g., (3241) = (2134) (3142).

If the coset leaders are chosen to be those elements that have a weight e or less and decoding is performed by selecting the code word directly above the received word in the array, then all errors of weight e or less can be correctly decoded. Since the weight of the coset leader is equal to the number of transpositions caused by the error, it is then possible to correct all errors that involve e transpositions or less.

Cyclic Subgroups

An important class of subgroups of S_n are the cyclic subgroups. These subgroups are generated by taking all of the distinct powers of a single element: $a, a^2, a^3, \cdots$, where $a^2 = aa$ indicates that the permutation is multiplied by itself. Since the number of elements in S_n is finite, the powers of a single element will eventually repeat themselves in a cycle. If a is the generating element, then the members of the subgroup are

$$a, a^2, a^3, \cdots, a^{r-1}, a^r = 1.$$

The subgroup used in the example given above is such a cycle group with (3142) as the generating element and $r = 4$.

$$(3142)(3142) = (4321) = a^2$$

$$(3142)(4321) = (2413) = a^3$$

$$(3142)(4213) = (1234) = a^4 = 1.$$

The minimum weight of the elements in this subgroup is 3.

The cyclic subgroups provide a simple method for generating rank vector codes.

Derivation of Rank Cyclic Codes

A digital computer program has been used to derive cyclic rank codes generated by each of the permutations of 4, 5, and 6 digits and to find the weights of each code word. From the computations made, it is possible to find the largest cyclic subgroups of a given minimum weight. The table below gives the sizes of these cyclic subgroups.

	Number of Code Words in Subgroup		
Minimum Weight	$n = 4$	$n = 5$	$n = 6$
2	4	6	6
3	4	6	6
4	3	5	6
5	2	4	6
6	2	5	6
7		2	6
8		2	6
9		2	2
10		2	3
11			2
12			2
13			2
14			2
15			2

There is no guarantee that these cyclic subgroups are the largest possible subgroups of a given minimum weight, and indeed, for $w = 2$, the alternating subgroup[8] A_n, which is not cyclic, has 12, 60, and 360 elements for $n = 4$, 5, and 6. No consistent method has been determined for finding the largest subgroup of a given minimum weight. An upper bound can be determined, however, and, as will be shown below, for w sufficiently large, the cyclic subgroups approach this bound.

Upper Bound on Size of a Subgroup[9]

For a code to be able to correct all errors of weight e or less, the minimum distance between its code words must be at least $2e + 1$. Given a code with this minimum

[7] Although much of the terminology used in this paper is the same as that used in describing binary group codes, it should be remembered that the quantities are calculated in a different way.

[8] The alternating subgroup A_n consists of the $n!/2$ permutations of even weight. A code comprised of the alternating subgroup has minimum weight 2. Such a code is analogous to a single parity check digit binary code and is capable of detecting a single transposition.

[9] This derivation is similar to the derivation of the Shannon bound [6] (also known as Hamming bound) for linear group codes.

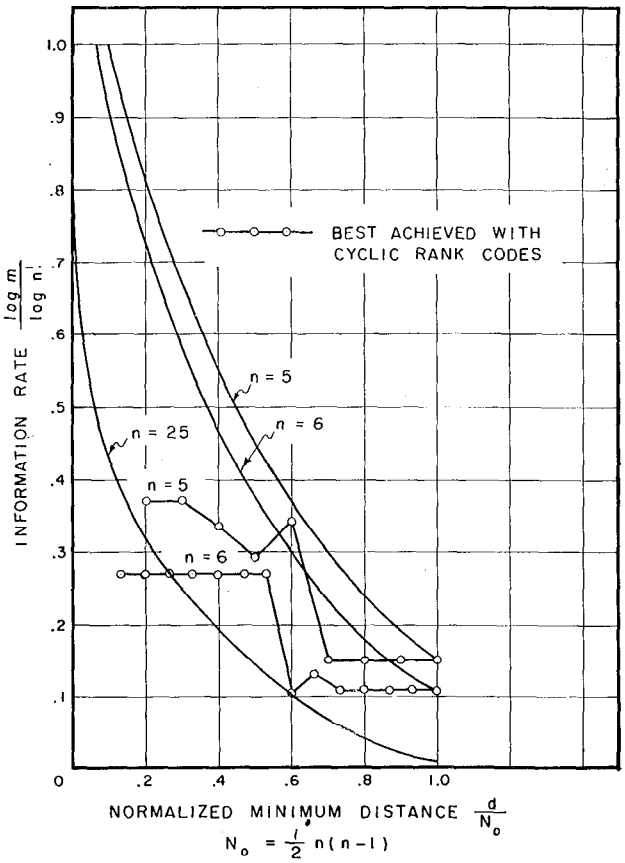

Fig. 1. Upper bound on information rate versus minimum distance.

distance, then all of the permutations of weight e or less must be coset leaders. If there are m code words there will be $n!/m$ coset leaders, including the identity element. For such a code to exist, the number of coset leaders must be greater than or equal to the number of permutations of weight e or less. Hence,

$$m \leq \frac{n!}{\sum_{w=0}^{e} U_n(w)}$$

where $U_n(w)$ is the distribution of weights, as given in the Appendix.

This bound is plotted in Fig. 1 for $n = 5$ and 6. It can be seen that for w larger than half of the maximum of $(\frac{1}{2})n(n-1)$, the bound is closely approached by the cyclic subgroup codes. The curves are plotted with $\log m/\log n!$, referred to as the information rate of the rank permutation code, versus the minimum distance $d = 2e + 1$, normalized by the maximum distance $N_0 = \frac{1}{2}n(n-1)$. If n is large, the normal approximation to the distribution of weights can be used (see Appendix and [7]), namely,

$$U_n(w) \approx n! \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(w-\mu)^2\right] \tag{7}$$

where

$$\mu = (\tfrac{1}{4})n(n-1)$$

and

$$\sigma^2 = (\tfrac{1}{72})n(n-1)(2n+5),$$

then

$$\frac{n!}{m} \geq \int_{-\infty}^{e} U_n(w)\, dw$$

$$\frac{n!}{m} \geq n!\, \Phi\left(\frac{e-\mu}{\sigma}\right),$$

or

$$m \leq \frac{1}{\Phi\left(\frac{e-\mu}{\sigma}\right)} \tag{8}$$

where

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx. \tag{9}$$

Decoding Methods

In order to correct every transmission error of weight e or less, it is necessary to construct a subgroup with an array having all permutations of weight e or less as coset leaders. The decoding method is then to choose that code word which lies directly above the received code word in the array. The implementation of such a scheme can be performed more simply, however, without storing the complete array. A simpler procedure is outlined below.

Assume that having transmitted the rank vector R_0 the rank vector R is received. If the error pattern $E = \{e_i\}$ occurs where the weight of E is less than e, then E will be a coset leader and $R = ER_0$ will appear under R_0 in the array. If $R_0 \neq 1 = (1, 2, \cdots, n)$ then the weight of R must be greater than e, since all permutations of weight less than e are coset leaders.

Theorem 3

If R_1 is a member of the subgroup (code word), the product RR_1^{-1} will have weight e or less if and only if $R_1 = R_0$, where R is the received word and R_0 is the transmitted word.

Proof: To prove the "if" statement assume that $R_0 = R_1$, then the product

$$RR_1^{-1} = ER_0R_1^{-1} = ER_0R_0^{-1} = E,$$

which has weight e or less.

To prove the "only if" statement assume that $R_0 \neq R_1$, then the product

$$RR_1^{-1} = ER_0R_1^{-1} = ER_2$$

where R_2 must be some code word and the product must appear in the array not as a coset leader and have weight greater than e, which results in a contradiction. Q.E.D.

Decoding can, therefore, be performed in the following manner.

1) If R, the received rank vector, is a code word, accept that code word as the decoded output.

2) If R is not a code word, form the products RR_i^{-1}, where the R_i are the possible code words, and calculate the weight of each product. The accepted code word is that R_i for which the weight of the product is e or less.

Communication Systems with Feedback

By choosing a subgroup as a code such that all permutations of weight e or less are coset leaders but some permutations of weight greater than e are also coset leaders, a code can be formed such that all error patterns of weight less than e are correctable and some error patterns of weight greater than e are detectable. In the latter case a feedback communication system can be used in which the transmitter requests a repeat from the receiver when a detectable but not correctable error has occurred.

Error Probabilities

Monte Carlo Computer Simulation

The calculation of error probabilities for the rank vector coding scheme is extremely difficult. To determine the probability that one permutation will be transformed into another by noise it is necessary to assume some signal form and some noise distribution. The signal form used in this paper is probably the most easily analyzed, but exact analytical results have not been obtained. Monte Carlo simulation techniques using the digital computer were used instead to estimate the error probabilities.

The signal was assumed to be of the form $s(t) = Cr_i$; $i = 1, 2, \cdots, n$, where C is an arbitrary constant and the r_i are the components of the transmitted rank vector. A single observation is assumed at each sampling interval, although an improvement in accuracy could easily be obtained by taking more than one observation and averaging over each sample interval.[10] The signal is thus pulse modulated with the pulse levels proportional to the rank of each interval. A smoother signal with considerably lower bandwidth could easily be obtained by rounding the corners of the stairstep function as long as the relative ranking at the sampling intervals is maintained.

Additive noise that is statistically independent and identically distributed from sample to sample is assumed. A transposition will occur at the receiver if the effect of the noise is to change the relative rankings of two of the observations. Because the ranking of the received signal observations is used for decoding rather than the signal levels themselves, the probabilities will be relatively insensitive to changes in noise distribution. For example, the noise mean is not important because it does not change the relative ranking of the observations.

The direct method of calculating the probability that one rank vector is transformed into another by noise would be to integrate the joint probability density function of the sample observations over the region of the sample space in which the transformation occurs. This method requires evaluating an n-fold integral and is discarded as impossibly complex.

A simplification of this method is provided by the extension to the Hoeffding Theorem [8] as was demonstrated by Stoll and Kurz [1]. This theorem allows the calculation of the probability of receiving some rank vector R given that the rank vector R_0 was transmitted. The fundamental weakness of this approach is that the equations are again difficult to evaluate except for cases where the signal-to-noise ratio is small. The evaluation uses the first term of a series approximation, and gives poor results for signal-to-noise ratios above approximately -10 dB.

The procedure adopted in this paper to give approximate results for the error probabilities was to simulate the addition of noise to the transmitted signal and to determine the rank vector of the total signal. The computer program made repeated trials of this procedure and counted the distribution of the distances between the transmitted and received rank vectors. The simulation was performed for various signal-to-noise ratios between $+3$ dB and -6 dB when the signal-to-noise ratio for Gaussian noise was defined by

$$\frac{S}{N} = \frac{C^2}{\sigma^2} \tag{10}$$

where c is the signal gain constant defined in (2) and σ^2 is the variance of the Gaussian distributed noise. The simulations were also performed with Cauchy distributed noise, with probability density function

$$p(x) = \frac{\lambda}{\pi(x^2 + \lambda^2)}. \tag{11}$$

For this case (in which the noise variance is infinite), the signal-to-noise ratio was defined by

$$\frac{S}{N} = \frac{C^2}{\lambda^2}. \tag{12}$$

The signal-to-noise ratio was defined in this manner, rather than as a direct ratio of signal power to noise power, so that it would not depend on n and comparisons could be easily performed between performance for various values of n if the value of C is maintained.[11]

Figs. 2–7 show the results of the computer simulations plotted as the cumulative distribution function $P[d > e]$, where d is the distance between transmitted and received rank vectors, versus e. Thus, if a code can correct all errors up to weight e, these curves represent the probability that the decoding will be in error.

A Bound on Error Probabilities

Using another approach it is possible to determine an upper bound on the error probability that will be close to the actual error probability when the signal-to-noise ratio is high.

[10] In such a system, the assumption of statistical independence between the averaged sample observations will be closer to the truth than would the individual observations used directly.

[11] In the case of Cauchy noise, the standard definition of the signal-to-noise ratio is meaningless.

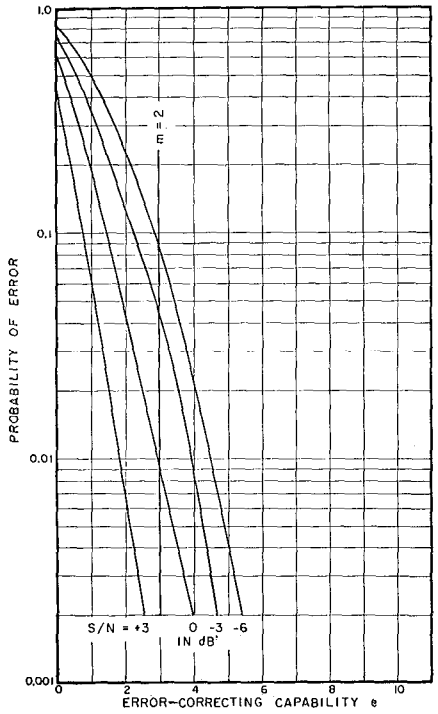

Fig. 2. Probability of error versus error-correcting capability: Gaussian noise $n = 4$.

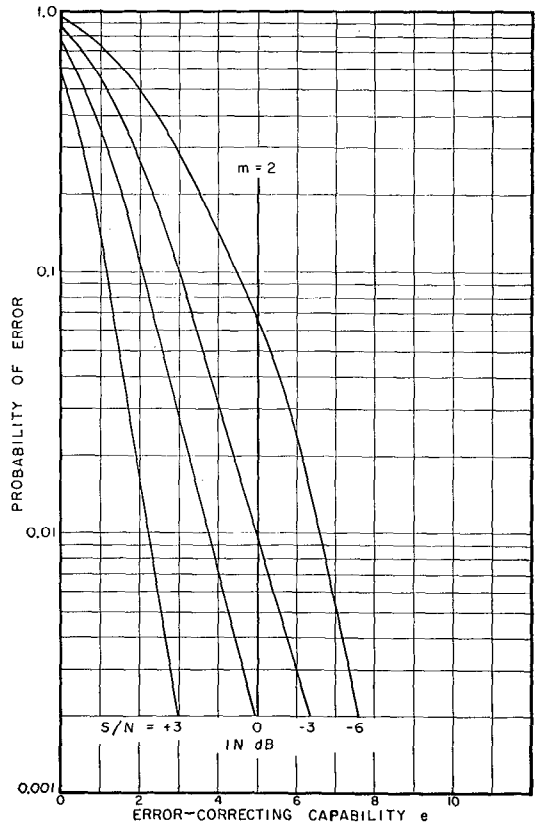

Fig. 3. Probability of error versus error-correcting capability: Gaussian noise $n = 5$.

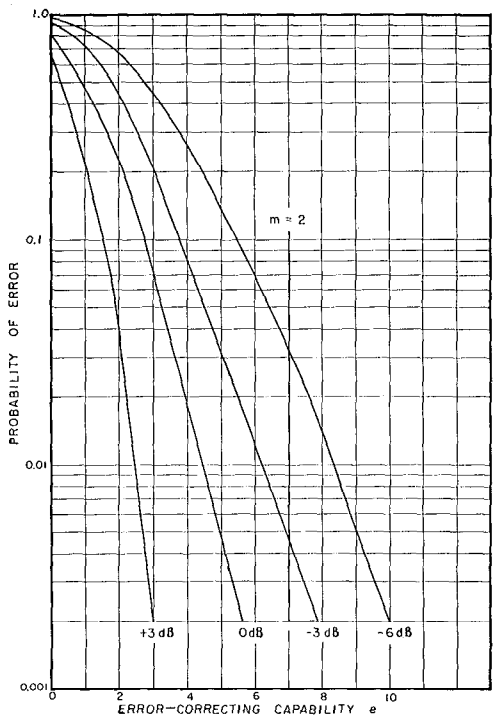

Fig. 4. Probability of error versus error-correcting capability: Gaussian noise $n = 6$.

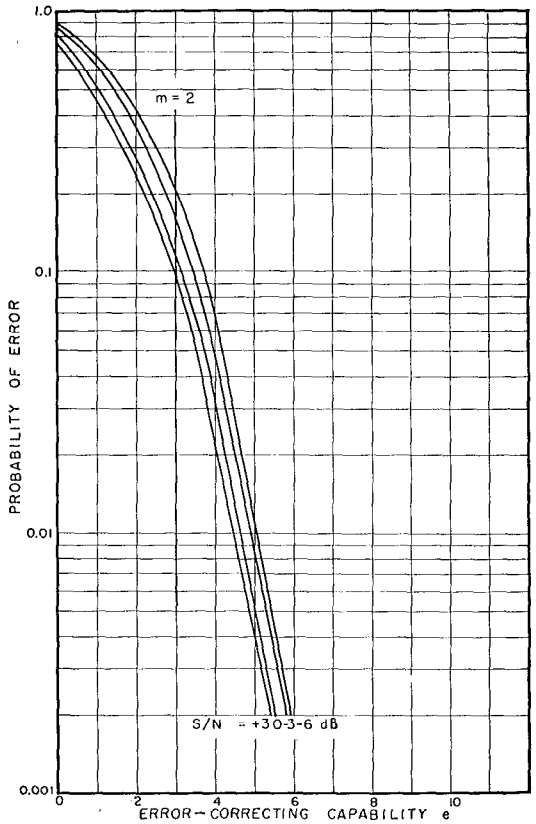

Fig. 5. Probability of error versus error-correcting capability: Cauchy noise $n = 4$.

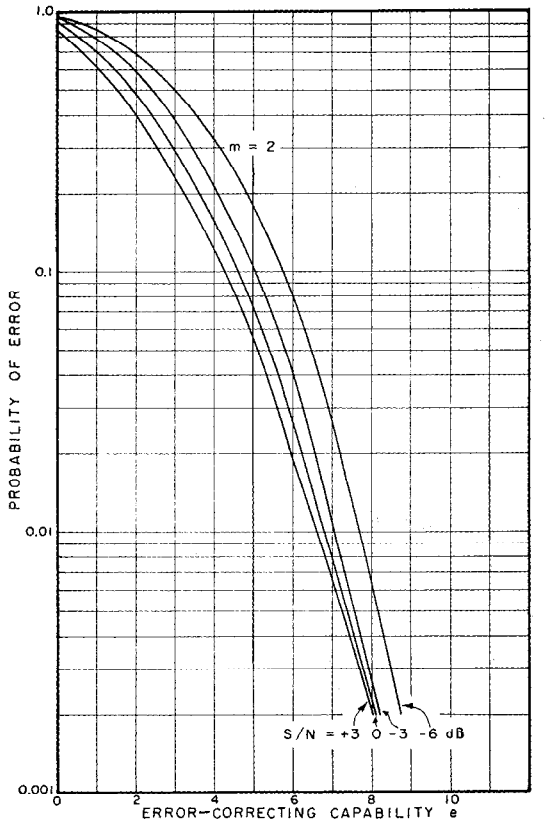

Fig. 6. Probability of error versus error-correcting capability: Cauchy noise $n = 5$.

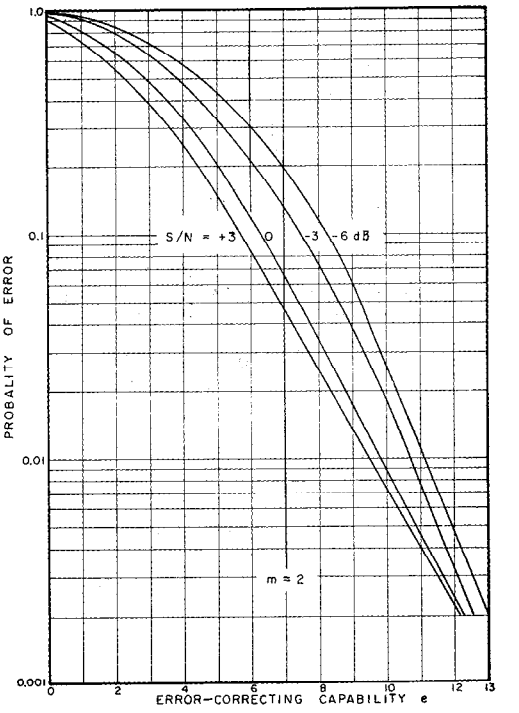

Fig. 7. Probability of error versus correcting capability: Cauchy noise $n = 6$.

The transmitted vector is assumed to be the identity element $\mathbf{1} = (1, 2, \cdots, n)$. This assumption does not lead to a loss of generality, as will become obvious later.

The probability of the interchange (12) occurring is the probability that the observation x_1 is greater than x_2 at the receiver and is given by

$$P_{12} = P[x_2 > x_1] = \int_{-\infty}^{\infty} \int_{x_2}^{\infty} f(x_1, x_2)\, dx_1\, dx_2. \tag{13}$$

This probability can be calculated when the noise density is known and will be a function of the gain constant C and the noise variance σ^2.

The probability P_{12} is the combined probability of all rank vectors occurring at the receiver in which $x_1 > x_2$. There are $n!/2$ such permutations. For example when $n = 4$ the permutations are divided as follows.

$x_2 > x_1$	$x_1 > x_2$
1234	2134
1243	2143
1324	3124
1342	3142
1423	3241
1432	3214
2314	4123
2341	4132
2413	4213
2431	4231
3412	4312
3421	4321

The same probability can be calculated for all other possible transpositions. When the signal form is such that all levels are equally separated, then

$$P_{12} = P_{23} = P_{34}, \cdots, P_{n-1-n} \tag{14}$$

where $P_{i-1,i}$ is the probability that the $(i-1)$st and ith coordinates have been interchanged.

Similarly, the probabilities $P_{i-2,i}$; P_{i-3}, $\cdots$, can also be calculated. Each such transposition will occur for exactly half of the $n!$ permutations.

Let $p(R_i)$ be the probability that the rank vector R_i occurs at the receiver. Then

$$\sum_{i=1}^{n!} P(R_i) = 1 \tag{15}$$

and the probability P_{12} is equal to the sum of the probabilities P_i for those permutations for which $x_1 > x_2$, because the permutations form a mutually exclusive set of events. For $n = 4$

$$P_{12} = P(2134) + P(2143) + P(3124) + \cdots + P(4321). \tag{16}$$

If the probabilities of each possible permutation $P_{i-k,i}$ are written as above, the permutation R_i will occur in a number of expressions equal to its weight. For example, the permutation (4321) has weight 6 and appears in the six expressions for

$$P_{12},\ P_{23},\ P_{34},\ P_{13},\ P_{24},\ P_{14}.$$

Thus, if all of the expressions are summed, the following equation is obtained

$$\sum_{i=2}^{n}\sum_{k=1}^{i-1} P_{i-k,i} = \sum_{j=1}^{n!} w(R_j)P(R_j) \tag{17}$$

where $w(R_j)$ is the weight of the permutation R_j.

This sum is the expected value of the weight w given that the rank vector $\mathbf{1} = (1, 2, \cdots, n)$ was transmitted. It can also be interpreted as the average weight of the noise patterns that will occur in a given noise situation, namely,

$$E(w) = \sum_{w=}^{w_m} wP(w) = \sum_{i=2}^{n}\sum_{k=1}^{i-1} P_{i-k,i} \tag{18}$$

where $w_m = \frac{1}{2}n(n-1)$ is the maximum weight.

The generalized Chebyshev inequality (see Fisz [9, p. 74]) states that if $g(x)$ is a non-negative function of a random variable, then

$$P[g(x) \geq k] \leq \frac{E[g(x)]}{k} \tag{19}$$

where k is a positive constant. Substituting the pertinent values results in

$$P_e = P[w \geq e+1] \leq \frac{E(w)}{e+1} \tag{20}$$

where $E(w)$ is given by (18).

The above relation can be extended to form the Bienayme–Chebyshev inequality for cases in which the variance of w exists, thus obtaining

$$P[|x-\mu| \geq k] \leq \frac{\sigma^2}{k^2} \tag{21}$$

where μ is the mean and σ^2 the variance of the distribution of x.

It can be assumed that the variance of the distribution of w will be no greater when a signal is present than when no signal is present (signal-to-noise ratio equals 0). When no signal is present, the variance is the same as that derived for the distribution of w among the permutations (see Appendix), because, in this case, all permutations are equally probable. Hence

$$\sigma_w^2 \leq \tfrac{1}{72}n(n-1)(2n+5)$$

and, by the Bienayme–Chebyshev inequality,

$$P[|w - E(w)| \geq e+1] \leq \frac{\sigma_w^2}{(e+1)^2}$$

$$P_e = P\left[\left(w - \sum_{i=2}^{n}\sum_{k=1}^{i-1} P_{i-k,i}\right) \geq e+1\right] \leq \frac{n(n-1)(2n+5)}{72(e+1)^2}, \tag{22}$$

which provides a tighter bound on the probability of error for high signal-to-noise ratios. For low signal-to-noise ratios, the extended Hoeffding theorem may be used.

Conclusions

The use of the Kendall correlation statistic as a measure of distance between rank vectors permits the application of the theory of permutation groups to the rank vector permutations. Codes have been derived that are subgroups of the basic permutation group and which have properties similar to those of linear group codes. The rank permutation group codes offer an efficient and easily applicable means of coding transmitted information.

Bounds on the probability of error for a communication system using rank permutation group codes have been obtained and verified by Monte Carlo simulation on a digital computer.

Appendix

The Distribution of Q

Kendall [7] has derived the distribution of the values of Q measured between each of the members of the group S_n and the identity element. If $U_n(Q)$ is the number of elements of S_n that have the particular value Q, then the following recursion formula can be used to determine the distribution.

$$U_{n+1}(Q) = U_n(Q-n) + U_n(Q-n+1) + \cdots + U_n(Q) = \sum_{i=0}^{n} U_n(Q-n+i). \tag{23}$$

Since for $n = 2$ there can be only two permutations in the group S_2, (12) and (21), and since $Q = 0$ and $Q = 1$ for these permutations, the starting values for the recursion formula are known. The distribution for Q up to 6 is given in the following table.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	1	1														
3	1	2	2	1												
4	1	3	5	6	5	3	1									
5	1	4	9	15	20	22	20	15	9	4	1					
6	1	5	14	29	49	71	90	101	101	90	71	49	29	14	5	1

The maximum value of Q is $N_0 = \frac{1}{2}n(n-1)$. Since the total number of elements in the group is $n!$,

$$\sum_{Q=0}^{N_0} U_n(Q) = n!. \tag{24}$$

As n gets large, the distribution of Q tends toward the normal distribution. The mean of the distribution is

$$\mu = \tfrac{1}{4}n(n-1) \tag{25}$$

and the variance is

$$\sigma^2 = \tfrac{1}{72}n(n-1)(2n+5). \tag{26}$$

Acknowledgment

The authors wish to express their gratitude to the anonymous reviewers whose comments helped to improve the paper considerably.

References

[1] E. Stoll and L. Kurz, "Suboptimum rank detection procedures using rank vector codes," *1967 IEEE Internat'l Conv. Rec.*, pt. 11, pp. 22–34; also *IEEE Trans. Communication Technology*, vol. COM-16, pp. 402–410, June 1968.

[2] J. C. Hancock and D. G. Lianiotis, "On learning and distribution-free coincidence detection procedures," *IEEE Trans. Information Theory*, vol. IT-11, pp. 272–280, April 1965.

[3] R. F. Daly and C. K. Rushforth, "Nonparametric detection of a signal of known form in additive noise," *IEEE Trans. Information Theory*, vol. IT-11 pp. 70–76, January 1965.

[4] J. Millard and L. Kurz, "Nonparametric signal detection—an application of the Kolmogorov–Smirnov, Cramér–von Mises tests," Lab. for Electroscience Research, New York University, N. Y., Tech. Rept. 400-127, January 1966; also "The Kolmogorov–Smirnov tests in signal detection," *IEEE Trans. Information Theory (Correspondence)*, vol. IT-13, pp. 341–342, April 1967.

[5] R. D. Carmichael, *Introduction to the Theory of Groups of Finite Order*. Boston: Ginn, 1937.

[6] W. W. Peterson, *Error-Correcting Codes*. Cambridge, Mass.: M.I.T. Press, 1961.

[7] M. Kendall, *Rank Correlation Methods*. London: Griffin, 1958.

[8] D. A. S. Fraser, *Nonparametric Methods in Statistics*. New York: Wiley, 1957.

[9] M. Fisz, *Probability Theory and Mathematical Statistics*. New York: Wiley, 1963.

[10] H. H. Schaefer, *Topological Vector Spaces*. New York: MacMillan, 1966.

Correspondence

Nonstationary Autoregressive Processes

***Abstract*—Let $\mathbf{R}y_t = u_t$ be a stochastic difference equation. Various relations between the input and output covariances and spectral densities are deduced under the hypotheses that $\mathbf{R}$ is time dependent and that u_t is a member of a nonstationary random process.**

Introduction

If $\mathbf{R}$ is a linear difference operator and u_t is a member of a stochastic process, then the equation $\mathbf{R}y_t = u_t$ defines an autoregressive process $\{y_t\}$. Such discrete analogs of the continuous model (that is, the case when $\mathbf{R}$ is a *differential* operator) are of interest in the statistical theory of time series and signal detection [1], [2], [4]. In almost all situations that we have observed in the literature the coefficients of $\mathbf{R}$ are assumed to be constant and u_t is assumed to be a member of a wide-sense stationary stochastic process. Our objective here is to obtain results when the coefficients of $\mathbf{R}$ are functions of time, and the $\{u_t\}$ process is nonstationary. The relations we deduce between covariance functions and spectral densities generally have well-known continuous analogs, although the correspondence is not always one to one.

The Difference Equation

Let $I = \{\cdots, -1, 0, 1, \cdots\}$ be the set of all integers. We shall assume that all functions considered are defined on I. Consider the linear difference operator

$$\mathbf{R} = \alpha_0(t)\mathbf{L}^0 + \alpha_1(t)\mathbf{L} + \cdots + \alpha_q(t)\mathbf{L}^q \tag{1}$$

where $\mathbf{L}$ is the lag operator, $\mathbf{L}y_t = y_{t-1}$, and the $\alpha_i(t)$, $0 \leqq i \leqq q$, are real-valued functions of t defined on I. We assume $\alpha_0(t)\alpha_q(t) \neq 0$ on I so that $\mathbf{R}$ is of the qth order. Let $\{u_t \mid t \in I\}$ be a family of random variables with mean-zero and covariance function $\sigma_u(t, s) = \mathcal{E}u_t u_s$. Then if $G(t, s)$ is the one-sided Green's function for $\mathbf{R}$,

$$y_t = \sum_{r=0}^{\infty} G(t, t-r)u_{t-r} \tag{2}$$

converges in mean-square to a mean-zero random variable, which is a solution of

$$\mathbf{R}y_t = u_t \tag{3}$$

Manuscript received June 7, 1968; revised August 26, 1968. This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and was monitored by Electronic System Division, AF Systems Command under Contract F19-628-68-6-0084.

on I, provided

$$\sum_{s=-\infty}^{t} |G(t, s)| < \infty \tag{4}$$

for all t, and

$$|\sigma_u(t, t)| < M$$

for some constant M and for all $t \in I$ (see [5]).

The covariance function of the autoregressive process $\{y_t \mid t \in I\}$ defined by (3) is

$$\begin{aligned}\sigma_y(t, s) &= \mathcal{E}y_t y_s \\ &= \sum_{r=0}^{\infty}\sum_{\rho=0}^{\infty} G(t, t-r)G(s, s-\rho)\sigma_u(t-r, s-\rho).\end{aligned} \tag{5}$$

The Spectral Density

A real stochastic process $\{u_t \mid t \in I\}$, which is stationary in the wide-sense, defines a sequence of covariances $\sigma_u(0)$, $\sigma_u(1)$, $\cdots$. For any covariance sequence $\{\sigma_u(h)\}$, there exists a unique monotone nondecreasing spectral distribution function $F_u(\lambda)$ with $F_u(-\pi) = 0$ such that

$$\sigma_u(h) = \int_{-\pi}^{\pi} \cos h\lambda \, dF_u(\lambda),$$

(Stieltjes integral representation). In particular, if $\sum_{k=0}^{\infty} \sigma_u(k)$ converges absolutely, then $dF_u(\lambda) = f_u(\lambda)d\lambda$ where $f_u(\lambda)$ is the spectral density of the $\{u_t\}$ process. Futhermore, $f_u(\lambda)$ is a continuous function of λ,

$$f_u(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} \sigma_u(k)e^{-ik\lambda}$$

and

$$\sigma_u(h) = \int_{-\pi}^{\pi} f_u(\lambda)e^{ih\lambda}\, d\lambda, \qquad h \in I, \tag{6}$$

(see [1]).

Now let us specialize the random process $\{u_t \mid t \in I\}$.

Case 1

Let $u_t = x_t\xi_t$ where $\{x_t \mid t \in I\}$ and $\{\xi_t \mid t \in I\}$ are real independent wide-sense stationary mean-zero processes with respective co-