DECODING OF GROUP CODES FOR BINARY SYMMETRIC CHANNELS

by

WALTER IRVINE WELLS

B. S. Massachusetts Institute of Technology
1952

S. M. Massachusetts Institute of Technology
1952

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 1960

Signature of Author _____
Department of Electrical Engineering, February 26, 1960

Certified by _____
Professor R. M. Fano Thesis Supervisor

Accepted by _____
Chairman, Departmental Committee on Graduate Students

DECODING OF GROUP CODES FOR BINARY SYMMETRIC CHANNELS

by

WALTER IRVINE WELLS

Submitted to the Department of Electrical Engineering on February 26, 1960, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

Coding information prior to transmission is a way of inserting redundancy into transmitted sequences so that the effects of noise disturbances during transmission can be minimized at the receiver. Shannon's theorem for noisy channels indicates that with proper encoding and decoding, one can transmit information over a noisy channel with vanishingly small probability of error provided the transmission rate does not exceed the channel capacity. To achieve this goal, the coded sequences must be very long, so that, in effect, they are subjected to the average noisy behavior of the channel rather than to its instantaneous behavior. As a practical matter, the decoding of very long sequences is a severe problem, since for a constant rate of information transmission the number of sequences the receiver must be prepared to decode grows exponentially with the length of the codes.

The investigation of this report considers the decoding problem for an additive (finite field) channel with noise consisting of statistically independent components. The approach is based upon the classification of receivable sequences, such that the complexity of decoding is related more directly to the number of equivalence classes than to the number of receivable sequences. This concept, first studied by Prange, is motivated by the existence of a step by step decoding scheme requiring only a knowledge of the 'distance' a received sequence is from the set of code sequences. Classification is accomplished by a group of distance preserving transformations such that all receivable sequences in a given class are an equal distance from the code sequences. It is shown that these classes have the formal properties of complete conjugate classes of finite groups. The theory of group representations is employed to formulate the entire problem and to prove that there do exist complete invariants for the classes, which can be found, in theory at least, by methods of the report. It is shown also that the group of measure preserving transformations, for the case of noise with statistically independent components, is a group of coordinate permutations. Group codes are developed as a special case of the general 'equivalence class' technique. This report is concerned primarily with the general formulation of the 'equivalence class' approach to decoding rather than with its practical application to specific codes.

Thesis Supervisor: Robert M. Fano

Title: Professor of Electrical Engineering

ACKNOWLEDGMENT

The author expresses his sincere appreciation to those whose help and interest have contributed to the formulation and development of topics in this report. Primary credit, of course, belongs to Professor R. Fano, thesis advisor, and to Professors P. Elias and J. Wozencraft, thesis readers. The author also expresses his thanks to Professor D. Arden and to Dr. M. Schutzenberger for the several informative discussions of group representation theory. Manuscript typing was done in draft form by E. M. Vadeboncoeur and B. Quigley, and the final typing by B. Quigley. The author thanks them sincerely for a job well done.

Without continual support and encouragement from home, this work could not have been finished. It is fitting then that the author dedicates this paper to his wife and family.

The author acknowledges the aid given by the Research Laboratory of Electronics at M. I. T. whose facilities he used while carrying out thesis research, and also the financial support of the Lincoln Laboratory. *

* The work reported in this paper was performed by Lincoln Laboratory, a center for research operated by M. I. T. with the joint support of the U. S. Army, Navy, and Air Force under Air Force Contract AF 19(604)-5200.

TABLE OF CONTENTS

TABLE OF CONTENTS
(continued)

Figures

CHAPTER I

Introduction and Summary

1. Statement of Problem

This study is concerned with the properties of codes that make them suitable for the transmission of information over noisy channels. Encoding information prior to transmission is a way of inserting redundancy into the transmitted sequences so that effects of noise in the channel can be eliminated at the receiver. This insertion of redundancy (coding) can be thought of as a mapping of the message or information space onto a larger signal space. The decoding is a mapping of the signal space back onto the message space. The first is one-to-one while the latter is a many-to-one mapping. Each can be carried out by using a dictionary. However, since the dictionary size grows exponentially with the length of the sequences, code books for accomplishing transmission with very long sequences are too large to be practical. To alleviate this difficulty it is desirable to accomplish encoding and decoding in a more systematic fashion, such that the code book sizes are limited.

2. History of the Problem

Practically all the present impetus for the study of coding theory stems from a theorem by Shannon[1]. This theorem states that it is possible to communicate over a noisy channel at a finite rate, with vanishingly small error, provided the rate is less than the channel capacity. The price to be paid for this error free operation is that one must encode the information so that intersymbol dependencies extend over long periods of time. Heuristically, this insures that any given symbol is subjected to the average noisy behavior of the channel, rather than to its instantaneous behavior.

Methods for creating the desired intersymbol dependencies are called codings. For the binary symmetric channel, two important types of codes have received attention. The first is the block code, in which a set of k information bits is mapped into a block (or sequence) of n bits to be transmitted. Decoding is done for each n bit block independently from all other blocks.

A second type of code has been studied, in which the encoding and decoding process is a sequential, step-by-step operation. These convolution codes, as they are called, have error correcting properties similar to block codes.

One might ask two questions about any code:

1. What is its error correcting capability, and
2. How can it be encoded and decoded?

These questions have been answered for many specific block codes [2, 3, 4]. For the general block code, Elias[5] showed that the error correcting capability approaches the Shannon bound. For the convolution code, Wozencraft[6] showed that the error correcting capability approaches the Shannon bound and that coding and decoding procedures can be kept within quite reasonable limits of complexity.

General results for block codes with respect to ease of decoding have been lagging. One of the most promising approaches was made by Slepian[7] when he considered group codes, a class of block codes. Elias[5] showed that there are group codes that attain Shannon's bound. The beauty of group codes is the simplicity of encoding and the possibility of using the well known theory of groups in their study.

Prange[8] has been studying group codes from the point of view of their relationship to groups of permutations. He reasoned that one might classify all received sequences into a few equivalence classes such that a dictionary may need only one entry per class. The experimental approach he has employed indicates that great strides might be made if we could further understand the underlying theory. Prange set forth the basic ideas being investigated in this report.

3. Approach to the Problem

Prange' s work illustrated the possibility of a very fruitful method of attack; however, several important aspects of his technique needed clarification. The purpose of this report is to formulate in a general fashion the ' equivalence class' approach to the decoding problem.

It is assumed that sequences of length n are to be transmitted and received. They are composed of symbols drawn from a finite field. The received sequences differ from the transmitted ones by the addition of a noise sequence of length n, composed of symbols drawn from the same finite field. Addition of the noise sequence and the transmitted one is component by component addition in the finite field. It is further assumed that the noise sequences occur randomly with the components being statistically independent, and that the symbols in the finite field occur with fixed probability.

The receivable sequences comprise the group G of all sequences of length n, with components in the finite field. The transmittable sequences form a subset A, of G. The problem of decoding is for the receiver to associate some transmittable sequence from A with each received sequence. Consequently, the decoder must be prepared, in some sense, to associate any sequence of G with a sequence of A. A distance function over G is defined such that the distance from one sequence, a, to another, r, is a monotonic function of the probability of occurrence, as noise in the channel, of the difference sequence (r - a). The maximum likelihood method of estimation leads to a decoder whose output, for any received sequence r, is that member a of A, such that the distance from a to r is minimum.

The basis for classifying the receivable sequences, G, is in terms of the defined distance function. As a first step it is shown that if, for any sequence r in G, the minimum distance from A to r is known, then a step-by-step decoding can be effected. Consequently, a classification of the sequences of G, in terms of their minimum distances from A, might be desirable. The actual classification is less direct, for one must provide in addition a way of determining to which class any received sequence belongs, by making some calculation based upon that sequence.

By means of distance preserving transformations, the sequences of G may be arranged into classes of equal minimum distance from A. That is, the distance will be constant over any class, but several classes may have the same distance. It is then shown that these classes may be exhibited as complete conjugate classes of a finite group Z. Consequently, there exists a set of irreducible representations of Z, such that the corresponding character functions form an orthogonal basis for complex valued functions over the classes. In particular, it is always possible to construct a function f(r) of the received sequence r whose numerical value is indicative of the class to which r belongs.

Decoding then proceeds as follows. For a given received sequence r, f(r) is evaluated, giving a numerical value that uniquely determines the class to which r belongs. A dictionary, with one entry per class, is then consulted to obtain the minimum distance from the set of code sequences, A. The first coordinate of r is changed in a specified way and the new minimum distance is determined. By changing successive coordinates and observing the resulting minimum distance from A, one arrives at the correct decoded sequence in a certain maximum number of steps.

It is shown that this procedure applies in general to non-group codes, and as a special case to group codes. It is also shown that the distance preserving transformation, for statistically independent noise components, must always be permutations. Some simple examples are given to illustrate the ideas presented.

4. Scope of the Work

This report is concerned chiefly with formulating the underlying theory of the 'equivalence class' approach to decoding; however, the techniques have been applied to several specific codes by others. Prange[8] gave decoding schemes for the Golay[9] code and for some of a class of cyclic codes, including one of block length, 73; Zierler[10], following a suggestion of the author, gave a different decoding scheme for the Golay code. The simplicity of decoding achieved in these cases is encouraging and it is felt that further work along these lines may lead fairly directly to some very useful and practical techniques.

This report is organized as follows. Chapter II states and proves the theorem on step-by-step decoding which is the basic motivation for the classification method developed in the report. Chapter III introduces the notation and some of the properties of Abelian groups. In Chapter IV, the concepts of the theory of finite groups are employed to formulate the classification problem and to carry it through to a demonstration of the properties of the classes and of the characters of the group representations for the classes. In Chapter V, we show that the group of distance preserving transformations must be a permutation group. Then follows a brief discussion of permutation groups. Examples of several group and non-group codes for the binary symmetric channel are given in Chapter VI. Chapter VII is a discussion of some of the possible variations on the theory, or extensions of it, which may be exploited in further work, or in application to specific codes. The Appendix II gives topics from the theory of finite groups that are assumed beginning with Chapter III. Appendix I gives computations, and the formulae for doing them, for some linear groups in their representation as permutation groups on sequences for the binary channel.

CHAPTER II

Step by Step Decoding for the Additive Channel

1. The Channel

We consider a channel through which the transmitter sends sequences of symbols. The sequences are assumed to be of length n, and the symbols to be drawn from a finite field of $m = p^q$ elements. We digress here to give the definition of a finite field and to introduce the notation.

Definition of a Field[11]

Let F be a set of elements, a, b, c, . . . for which the sum a+b and product a b of any two elements, a and b (distinct or not) of F, are defined. F is a field if the following postulates hold:

1. If a and b are in F, then $a + b$ and $a\,b$ are in F (closure).
2. If $a = a'$ and $b = b'$ are in F, then
   $a + b = a' + b'$ and $a\,b = a'\,b'$ (uniqueness).
3. For all a and b in F
   $a + b = b + a$ and $a\,b = b\,a$ (commutative).
4. For all a, b, and c, in F,
   $a + (b + c) = (a + b) + c$ and $a\,(b\,c) = (a\,b)\,c$
   (associative).
5. For all a, b, and c in F,
   $a\,(b + c) = a\,b + a\,c$ (distributive).
6. F contains an element 0, such that for all a in F,
   $0 + a = a$ (zero).
7. F contains an element $1 \neq 0$, such that for all a in F,
   $a\,1 = a$ (unity).
8. For each a in F, the equation $a + x = 0$, has a solution for x in F.

9. For every a, b, and c, in F, where $c \neq 0$, if $c\,a = c\,b$, then $a = b$.

10. For every a in F, where $a \neq 0$, there exists an element a^{-1} in F, such that $a\,a^{-1} = 1$. (inverse).

On the basis of postulate 8, it is possible to introduce the following notation. If

$$a + b = c, \text{ then } b = c - a, \text{ where } a + (-a) = 0.$$

It can be shown that if the number of elements, m, in the field is finite (finite field), then m is an integral power of a prime, p^q. These fields are called Galois Fields, GF(m). If $q = 1$, the elements of the field can, with no loss of generality, be the integers modulo p, (0, 1, . . . , p - 1). When $q \neq 1$, the elements of the field can be taken in the form of vectors in q space, with the coordinate values the integers modulo p. Thus the elements of the finite fields, ($\alpha_0 = 0$, α_1, α_2, . . . , α_{m-1}), can, for the purposes of this report, be visualized as q - tuples

$$\alpha_i = u_{i1}, u_{i2}, \ldots, u_{iq} \qquad \text{II-1}$$

in which the u_{ij} are the integers modulo p.

We consider a channel in which the transmitter can send one of a set of sequences A (a_1, a_2, . . . , a_s) consisting of n symbols drawn from GF(m). The sequence received may differ from the transmitted one by the addition, in GF(m), of a noise sequence belonging to the noise set N (ζ_1, ζ_2, . . . , ζ_t), similarly constructed. The received sequences, therefore, form a set R (r_1, r_2, . . . , r_u) of the same type, where we may denote

$$r_{ij} = a_i + \zeta_j \ ,$$

the addition being component by component in the GF(m).

It is further assumed that the noise sequences occur randomly with fixed probability. A particular noise sequence ζ_i is composed of components

$$\zeta_i = z_{i1}, z_{i2}, \ldots, z_{in},$$

in which each z_{ij} is an element of GF(m) ($\alpha_0 = 0, \alpha_1, \alpha_2, \ldots, \alpha_{m-1}$).

If b_k of the components, z_{ij}, are equal to α_k, and the probability of occurrence of α_k as a component of the noise sequence is p_k, the probability of the whole noise sequence ζ_i is

$$P(\zeta_i) = \prod_{k=0}^{m-1} p_k^{b_k} \qquad \text{II - 2}$$

This is, of course, the model for an additive channel in which the noise components are statistically independent.

If r_{ij} is received when a_i is transmitted and the noise ζ_j occurs, we define the distance from a_i to r_{ij} in terms of the probability of occurrence of ζ_j. A useful monotonic function of the probability is the so called self information (in arbitrary units)

$$I(\zeta_j) = -K \log P(\zeta_j).$$

The distance from a_i to r_{ij} is then defined as

$$D(a_i, r_{ij}) = I(\zeta_j) = K \sum_{k=0}^{m-1} b_k \log 1/p_k. \qquad \text{II - 3}$$

where K is an arbitrary constant.

It is further assumed that all transmittable sequences are sent with equal probability, i.e. $p(a_1) = p(a_2) = \cdots = p(a_s) = 1/s$. To maximize the probability of choosing the correct transmitted sequence, having received a given sequence r, one computes the conditional probability of a_i, given that r has been received. Then, since

$$r = a_i + \zeta_j$$

$$P(a_i/r) = \frac{P(\zeta_j = r - a_i)}{s \cdot P(r)}.$$

For every a_i in the transmittable set, A, there is an $\zeta_j = (r - a_i)$ whose probability can be computed. The largest one is the one that maximizes $P(a_i/r)$. Hence, a maximum likelihood decoder associates with each received

sequence that transmittable sequence corresponding to the noise with largest probability of occurrence. For the binary channel, this implies that the decoder chooses the transmittable sequence differing in the smallest number of components from the received sequence.

The maximum likelihood decoder output, for a given received sequence r, is that transmittable sequence a in A, for which D (a, r) is a minimum.

2. Requirements for Step by Step Decoding

We now prove a theorem which is the basis for a step by step method of decoding. Let e_j be a sequence whose j' th coordinate is 1, and the other n -1 coordinates are zero.

Theorem* II - 1:

Let the elements $(\alpha_0 = 0, \alpha_1, \alpha_2, \ldots, \alpha_{m-1})$ of the GF(m) be ordered according to their probability of occurrence, i. e.,

$$p_0 > p_i \geq p_{i+1} \qquad i = 1, 2, \ldots, m - 2.$$

If and only if there exists an a in A, such that r - a has α_t $(t \leqq k)$ for its j' th coordinate, and a minimizes D(a, r) then

$$\min_{a \epsilon A} D(a, r - \alpha_k e_j) = \min_{a \epsilon A} D(a, r) + (K \log 1/p_0 - K \log 1/p_k).$$

Proof: The sequences a, that minimize the left side can be one of four types:

a_I minimizes D(a, r) and $r-a_I$ has α_k in the j' th coordinate

a_{II} minimizes D(a, r) and $r-a_{II}$ has $\alpha_t = \alpha_k + \alpha_s$ in the j' th coordinate

a_{III} does not minimize D(a, r) and $r-a_{III}$ has α_k in the j' th coordinate

a_{IV} does not minimize D(a, r) and $r-a_{IV}$ has $\alpha_t = \alpha_k + \alpha_s$ in the j' th coordinate

* A version of this theorem is assumed in Prange's[8] work. He has given a proof in a private communication dated 27 August 1959.

Since $p_k \leqq p_t$

$$\log {}^{1}/_{p_k} \geq \log {}^{1}/_{p_t}$$

and $$\log {}^{1}/_{p_o} < \log {}^{1}/_{p_s}$$

hence $$(K \log {}^{1}/_{p_o} - K \log {}^{1}/_{p_k}) < (K \log {}^{1}/_{p_s} - K \log {}^{1}/_{p_t})$$

Case I.

$$D(a_I, r - \alpha_k e_j) = D(a_I, r) + (K \log {}^{1}/_{p_o} - K \log {}^{1}/_{p_k})$$

hence

$$\min_{a \epsilon A} D(a, r - \alpha_k e_j) = \min_{a \epsilon A} D(a, r) + (K \log {}^{1}/_{p_o} - K \log {}^{1}/_{p_k})$$

Case II.

$$D(a_{II}, r - \alpha_k e_j) = D(a_{II}, r) + (K \log {}^{1}/_{p_s} - K \log {}^{1}/_{p_t})$$

hence

$$\min_{a \epsilon A} D(a, r - \alpha_k e_j) > \min_{a \epsilon A} D(a, r) + (K \log {}^{1}/_{p_o} - K \log {}^{1}/_{p_k})$$

Case III.

$$D(a_{III}, r - \alpha_k e_j) = D(a_{III}, r) + (K \log {}^{1}/_{p_o} - K \log {}^{1}/_{p_k})$$

$$\min_{a \epsilon A} D(a, r - \alpha_k e_j) = D(a_{III}, r) + (K \log {}^{1}/_{p_o} - K \log {}^{1}/_{p_k})$$

hence

$$\min_{a \epsilon A} D(a, r - \alpha_k e_j) > \min_{a \epsilon A} D(a, r) + (K \log {}^{1}/_{p_o} - K \log {}^{1}/_{p_k})$$

Case IV.

$$D(a_{IV},\ r - \alpha_k e_j) = D(a_{IV},\ r) + (K \log {}^{1}/_{p_s} - K \log {}^{1}/_{p_t})$$

$$\min_{a \epsilon A} D(a,\ r - \alpha_k e_j) > D(a_{IV},\ r) + (K \log {}^{1}/_{p_o} - K \log {}^{1}/_{p_k})$$

hence

$$\min_{a \epsilon A} D(a,\ r - \alpha_k e_j) > \min_{a \epsilon A} D(a,\ r) + (K \log {}^{1}/_{p_o} - K \log {}^{1}/_{p_k})$$

Only Case I satisfies the conditions of the Theorem. Q. E. D.

We define the minimum distance function from any a in A to a sequence r,

$$D'(A,\ r) = \min_{a \epsilon A} D(a,\ r) \qquad \text{II - 4}$$

The decoding scheme is as follows.

Let the received sequence be (r). Now set

$$r_o = r$$

and

$$r_j = \begin{cases} r_{j-1} & \text{if } D'(A,\ r_{j-1} - \alpha_k e_j) > D'(A,\ r_{j-1}) - K \log {}^{1}/_{p_k} + K \log {}^{1}/_{p_o} \\ & \text{for all } \alpha_k \neq o \\ r_{j-1} - \alpha_k e_j & \text{if } D'(A,\ r_{j-1} - \alpha_k e_j) = D'(A,\ r_{j-1}) - K \log {}^{1}/_{p_k} + K \log {}^{1}/_{p_o} \\ & \text{for some } \alpha_k \neq o \end{cases}$$

This gives a procedure for "guessing" successively the n components of the noise sequence. If the guessed value α_k, of the j' th component is correct the distance decreases by a specific amount. For any incorrect guess $\alpha_k \neq \alpha_t$, the distance changes by a different amount, so long as $k > t$. The test is applied starting with the $\alpha_k \neq 0$ of lowest probability and working up to those of higher probability. The first α_k satisfying the second condition is used. This is because other α_k of higher probability may satisfy the given condition even if they do not match the coordinate value properly.

The number of steps required is $\leq n(m-1)$.

In order to do this step by step decoding procedure, one must have a method for determining $D'(A, r)$ for every receivable sequence. The remainder of this report is concerned, therefore, with a particular way of evaluating this function. The approach will be to collect together into classes those receivable sequences having equal values of $D'(A, r)$ in such a way that a dictionary with only one entry per class is required, in place of one entry per received sequence. The necessary constructions are given in Chapter IV. Chapter III is devoted to a discussion of some of the properties of Abelian groups used in Chapter IV.

CHAPTER III

Abelian Group Representations

Appendix II, Topics From the Theory of Finite Groups, gives the fundamental definitions and theorems of group theory and the theory of group representations, presented in a form specifically aimed at developing the topics necessary for this report. Parts of this report assume knowledge of these topics. This chapter is concerned with notation and examples of Abelian Groups and the characters of their irreducible representations, especially as they occur in the coding problem.

In the following work, we shall be concerned with certain Abelian Groups. These are composed of sequences of n symbols drawn from a finite field, GF(m), containing $m = p^q$ (p, prime) distinct elements. The group operation is addition of sequences, component by component, in GF(m).

We saw (II - 1) that each element in the GF(m), where $m = p^q$, can be represented as a q - tuple of integers modulo p. Consequently, the sequences of n symbols from GF (p^q) can be represented as sequences of qn integers modulo p. Thus each sequence in the Abelian Group under consideration will be in the form

$$s = b_1, b_2, \ldots, b_{qn} \quad \text{where the } b_i = 0, 1, \ldots, p-1.$$

Let t_k be a sequence whose k' th coordinate is 1, and the other qn-1 coordinates are zero. Then

$$s = \sum_{k=1}^{qn} b_k t_k \qquad \text{III - 1}$$

A representation of the group by diagonal matrices, with matrix multiplication as the group operation, can be formed by letting the basis sequences, t_k, be represented by the diagonal matrices $u_k = \text{diag.}\ (d_{k1}, d_{k2}, \ldots, d_{kqn})$, where

$$d_{ki} = \begin{cases} \sigma & \text{if } i = k \\ 1 & \text{otherwise} \end{cases}.$$

σ is a primitive p' th root of 1. That is, σ^i, for $i = 0, 1, \ldots, p-1$, takes on all values of the p' th roots of 1.

The group element, s, of III - 1 is then represented by the diagonal matrix, S,

$$S = \prod_{k=1}^{qn} (u_k)^{b_k} = \text{diag.}\ (\sigma^{b_1}, \sigma^{b_2}, \ldots, \sigma^{b_{qn}}) \qquad \text{III - 2}$$

Since the group is Abelian all irreducible representations are of first degree and their characters are equal to the single element in the 1 x 1 matrix of the representation. One can devise a useful notation for naming the different representations. Consider any group element, as represented in III - 2. With each group element we associate an irreducible representation Γ_s, in which the qn basis elements t_k are represented by the 1 x 1 matrices v_k, as follows

$$v_1 = \sigma^{b_1}, \; v_2 = \sigma^{b_2}, \ldots, \; v_{qn} = \sigma^{b_{qn}} \qquad \text{III - 3}$$

Then in this representation, any other group element

$$s' = \sum_{k=1}^{qn} c_k t_k \qquad \text{III - 4}$$

has the representation

$$\Gamma_s(s') = \prod_{k=1}^{qn} (v_k)^{c_k} , \qquad \text{III - 5}$$

$$= \prod_{k=1}^{qn} (\sigma^{b_k})^{c_k} = \sigma^{\sum b_k c_k} . \qquad \text{III - 6}$$

Consequently

$$\Gamma_s(s') = \Gamma_{s'}(s)$$

and as s runs through the n^m, $m = p^q$, elements of the group, Γ_s runs through the same number of irreducible representations.

If we regard the group elements s and s' (III - 1 and III - 4) as row vectors and s^t and s'^t as their transposes (column vectors) their inner product

$$s' s^t = s s'^t = \sum_{k=1}^{qn} b_k c_k . \qquad \text{III - 7}$$

Thus, since the character ψ is just equal to the element of the representation matrix, we have

$$\psi_s(s') = \psi_{s'}(s) = \sigma^{(s' s^t)} = \sigma^{(s s'^t)} \qquad \text{III - 8}$$

Example III - 1

Suppose $m = p = 3$. The field elements can be taken as 0, 1, and 2.

Let $n = 2$, and let σ be a primitive cube root of 1.

Two possible group elements are

$s = (1, 2)$

$s' = (1, 1)$

thus $s' s^t = 1 + 2 = 3$

so that $\psi_s(s') = \psi_{s'}(s) = \sigma^3 = 1$

The complete character table for the group of 3^2 elements is:

	00	01	02	10	11	12	20	21	22
00	1	1	1	1	1	1	1	1	1
01	1	μ	μ^2	1	μ	μ^2	1	μ	μ^2
02	1	μ^2	μ	1	μ^2	μ	1	μ^2	μ
10	1	1	1	μ	μ	μ	μ^2	μ^2	μ^2
11	1	μ	μ^2	μ	μ^2	1	μ^2	1	μ
12	1	μ^2	μ	μ	1	μ^2	μ^2	μ	1
20	1	1	1	μ^2	μ^2	μ^2	μ	μ	μ
21	1	μ	μ^2	μ^2	1	μ	μ	μ^2	1
22	1	μ^2	μ	μ^2	μ	1	μ	1	μ^2

The significant result of this chapter is the fact that the characters corresponding to the irreducible representations can be computed easily from the sequences themselves. The notation, in terms of the inner product of two elements viewed as vectors, will be especially useful in the following chapter.

CHAPTER IV

The Block Decoding Problem; m-ary Channel

1. Introduction

When the transmitter sends a block of n m-ary symbols, we consider them as elements of an Abelian group G, of size m^n of all m-ary sequences of length n, the rule of combination being component by component addition in the GF(m). Prior to reception, these sequences may be changed by noise in the channel which we assume has the effect of adding any group element to the transmitted one. In this sense we are speaking of additive noise.

The distinct receivable sequences clearly comprise the total group G; hence in order to decode what has been received, the decoder must be prepared, in some sense, to associate some transmittable sequence with any one of the m^n receivable ones.

One way to approach the problem is to attempt some classification of receivable sequences; to try to discover some characteristic common to whole sets of receivable sequences that simplify the decoding problem. One such technique is based upon the group code. Here one chooses transmittable code sequences A, to be all members of a subgroup of G. Then it can be shown that the cosets of A in G do form sets having desirable decoding properties. In this case, if there are m^k code sequences, there are m^{n-k} cosets and consequently one need only provide a dictionary of size m^{n-k} rather than m^n. This is a simplification, but unfortunately m^{n-k} grows beyond practical bounds of equipment complexity also. We shall return to group codes after developing the principal ideas of this report and show that group codes do possess desirable properties which can be exploited in the decoding problem.

It was shown in Chapter II how the decoding can be done if D'(A, b) can be evaluated for every receivable b. Our next step is based upon the observation that D'(A, b) can have the same value for several different received sequences. In this sense, at least, several of the received sequences are

equivalent. To pursue this idea we shall construct classes of receivable sequences such that D' (A, b) is constant over the class and such that there exists a computational method for determining to which class any receivable sequence belongs.

2. Imposing Structure on G.

It is a fundamental concept of the theory of finite groups that every group can be decomposed into disjoint complete conjugate classes; and that the characters of the irreducible representations of the group, as groups of non-singular transformations in the complex number field, form a complete set of orthonormal bases for the complex valued functions defined over the classes. At first glance one considers the receivable sequences to be members of an Abelian group. For Abelian groups, every element forms a complete conjugate class by itself, hence the number of calsses is m^n, and the number of irreduciable representations and characters is m^n. The conclusion is that these groups possess very little of the structure desired for simplification of the decoding problem.

To impose structure on this Abelian group G, such that the complete conjugate classes become larger and more useful, we imbed G in a larger group, Z, which is not Abelian. The conjugate class structure in then induced on those elements G' of Z which are simply isomorphic to G.

Let $G(s_1 = I, s_2, \ldots, s_g)$ be the group of m-ary sequences, $g = m^n$. Let $T(\Upsilon_1 = I, \Upsilon_2, \ldots, \Upsilon_f)$ be a group of transformations of the sequences of G, closed under the group operation, multiplication $\Upsilon_i \Upsilon_j = \Upsilon_k$. If

$$s_i + s_j = s_k$$

then

$$s_i\Upsilon + s_j\Upsilon = s_k\Upsilon$$

so that every element of G is transformed into another element of G and the group G is simply isomorphic to the group $G\Upsilon_i$, for every Υ_i in T.

We now define a group Z, having the general elements of the form

$$(s_i, \Upsilon_j)$$

in which the group combinatorial operation is

$$(s_i, \Upsilon_j)(s_u, \Upsilon_v) = (s_i + s_u\Upsilon_j, \Upsilon_v\Upsilon_j) \qquad \text{IV - 1}$$

From the multiplication formula it is readily seen that the group, Z, is closed under the group operation. The identity is (I, I) and the inverse of any element is

$$(s_i, \Upsilon_j)^{-1} = (s_i^{-1}\Upsilon_j^{-1}, \Upsilon_j^{-1}) . \qquad \text{IV - 2}$$

Associativity is demonstrated as follows:

$$\begin{aligned}(s_i, \Upsilon_j)\left[(s_u, \Upsilon_v)(s_w, \Upsilon_x)\right] &= (s_i, \Upsilon_j)(s_u + s_w\Upsilon_v, \Upsilon_x\Upsilon_v) \\ &= (s_i + s_u\Upsilon_j + s_w\Upsilon_v\Upsilon_j, \Upsilon_x\Upsilon_v\Upsilon_j) \\ &= (s_i + s_u\Upsilon_j, \Upsilon_v\Upsilon_j)(s_w, \Upsilon_x) \\ &= \left[(s_i, \Upsilon_j)(s_u, \Upsilon_v)\right](s_w, \Upsilon_x) .\end{aligned}$$

The elements so defined, thus form a group, Z, which is easily seen to be of order fg. Now form the conjugate of (s_u, Υ_v) with respect to (s_i, Υ_j)

$$\begin{aligned}&(s_i, \Upsilon_j)(s_u, \Upsilon_v)(s_i, \Upsilon_j)^{-1} \\ &= (s_i + s_u\Upsilon_j, \Upsilon_v\Upsilon_j)(s_i^{-1}\Upsilon_j^{-1}, \Upsilon_j^{-1}) \\ &= (s_i + s_u\Upsilon_j + s_i^{-1}\Upsilon_j^{-1}\Upsilon_v\Upsilon_j, \Upsilon_j^{-1}\Upsilon_v\Upsilon_j)\end{aligned} \qquad \text{IV - 3}$$

If $\Upsilon_v = 1$,

$$(s_i, \Upsilon_j)(s_u, I)(s_i, \Upsilon_j)^{-1} = (s_u\Upsilon_j, I) \qquad \text{IV - 4}$$

That is, elements of the form (s_u, I), which clearly form a subgroup G', of Z, simply isomorphic to the group G, also form a self-conjugate subgroup of Z.

Furthermore (s_u, I) and $s_j \Upsilon_j, I)$ are in the same complete conjugate class of Z, for every Υ_j in T Note that the complete conjugate classes of G' are not composed of just a single element as were the classes of G, but have several elements, depending on the group of transformations, T. This construction of Z is called a Holomorph[12].

3. Properties of the Holomorph

We have defined, previously (II-3), a measure or distance function on the elements of G. We now make use of this function to specialize the Holomorph just constructed. Assume from here on that T is measure preserving; that is

$$I(s_u \Upsilon_j) = I(s_u) \text{ for every } s_u \epsilon G \text{ and } \Upsilon_j \epsilon T.$$

If we define this same distance function for the corresponding elements of G'

$$I(s_u, I) \equiv I(s_u)$$

it is seen that the complete conjugate classes of G' consist of elements, all of which have the same value of $I(s_u, I)$. Consequently we make the following

Definition: If the class C of G, contains the element (s_u, I)

$$I(C) \equiv I(s_u, I) \equiv I(s_u) \qquad \text{IV - 5}$$

Theorem IV - 1: Given any two classes C_i and C_j of G', there is a minimum distance from class C_i to C_j such that

$$D'(C_i, C_j) = D'(C_i, b) \text{ for every element, b, in } C_j.$$

Proof: If $a \epsilon C_i$ and $b \epsilon C_j$, we define (see II-3)

$$D(a, b) = I(\zeta)$$

where

$$\zeta = a^{-1} b^*$$

and

$$D'(C_i, b) = \min_{a \epsilon C_i} D(a, b)$$

*Recall that for elements in G, for which distance was originally defined, addition was the group operation. In the group Z, corresponding elements combine by multiplication, so that the definition of distance is altered accordingly.

Let $C_{i'}$ be the class of a^{-1} if $a \epsilon C_i$. There is a theorem (theorem 2, appendix II) which says:

> If $b \epsilon C_j$ and the set $C_i b$ contains $d_{i',jk}$ elements from the complete conjugate class C_k, then for every $b \epsilon C_j$, the set $C_i b$ contains $d_{i',jk}$ elements from C_k.

ζ, above, is in one of the C_k for which $d_{i',jk} \neq 0$. Hence, choose a $\bar{C}_k$ for which $d_{i',jk} \neq 0$ which minimizes

$$I(\bar{C}_k)$$

Then

$$D'(C_i, b) = \min_{a \epsilon C_i} D(a, b) = I(\bar{C}_k) \text{ for all } b \epsilon C_j$$

and

$$D'(C_i, C_j) = D'(C_i, b) \text{ for all } b \epsilon C_j .$$

Q. E. D.

4. The Use of Class Characters

It will be recalled that the characters of the distinct irreducible representations form a complete ortho-normal basis for functions over the classes of a group. Before proceeding in general let us work a small example. Let G be the group of 2^4 binary sequences of length 4. Let $T(1, \Upsilon, \Upsilon^2, \Upsilon^3)$ be the group of four cyclic permutations on the four coordinates. The group, T, collects G into orbits. An orbit is a set of sequences having the form $\Upsilon_i b$, for some $b \epsilon G$, and every $\Upsilon_i \epsilon T$. For this example they are displayed as follows:

0	1 0 0 0	1 0 0 1	1 0	1 0 1 1	1
0	0 1 0 0	1 1 0 0	0 1	1 1 0 1	1
0	0 0 1 0	0 1 1 0	1 0	1 1 1 0	1
0	0 0 0 1	0 0 1 1	0 1	0 1 1 1	1
d_0	d_1	d_2	d_3	d_4	d_5

The conjugate classes of Z can be arrayed as follows:

$c_0 = (d_0, 1)$	$c_6 = \{(d_0, \tau), (d_5, \tau), (d_3, \tau), (d_2, \tau)\}$	$c_8 = \{(d_0, \tau^2), (d_5, \tau^2), (d_3, \tau^2)\}$	$c_{11} = \{(d_0, \tau^3), (d_5, \tau^3), (d_3, \tau^3), (d_2, \tau^3)\}$
$c_1 = (d_5, 1)$			
$c_2 = (d_3, 1)$			
$c_3 = (d_2, 1)$		$c_9 = (d_2, \tau^2)$	
$c_4 = (d_1, 1)$	$c_7 = \{(d_1, \tau), (d_4, \tau)\}$	$c_{10} = \{(d_1, \tau^2), (d_4, \tau^2)\}$	$c_{12} = \{(d_1, \tau^3), (d_4, \tau^3)\}$
$c_5 = (d_4, 1)$			

Table IV -1

Conjugate Class Structure

The characters for the irreducible representations are given in Table IV-2 below. (Here ω is a primitive fourth root of 1.) (See ref. 13 for methods for computing characters.)

h		Γ_0	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5	Γ_6	Γ_7	Γ_8	Γ_9	Γ_{10}	Γ_{11}	Γ_{12}
1	c_0	1	1	2	4	4	4	1	1	1	1	1	1	2
1	c_1	1	1	2	4	-4	-4	1	1	1	1	1	1	2
2	c_2	1	1	2	-4	0	0	1	1	1	1	1	1	2
4	c_3	1	1	-2	0	0	0	1	1	1	1	1	1	-2
4	c_4	1	-1	0	0	+2	-2	1	1	1	-1	-1	-1	0
4	c_5	1	-1	0	0	-2	2	1	1	1	-1	-1	-1	0
8	c_6	1	-1	0	0	0	0	-1	ω	ω^3	ω	ω^3	1	0
8	c_7	1	1	0	0	0	0	-1	ω	ω^3	ω^3	ω	-1	0
4	c_8	1	1	2	0	0	0	1	-1	-1	-1	-1	-1	-2
4	c_9	1	1	-2	0	0	0	1	-1	-1	-1	-1	-1	2
8	c_{10}	1	-1	0	0	0	0	1	-1	-1	+1	+1	1	0
8	c_{11}	1	-1	0	0	0	0	-1	ω^3	ω	ω^3	ω	1	0
8	c_{12}		1	0	0	0	0	-1	ω^3	ω	ω	ω^3	-1	0

Table IV - 2

Character Table for Holomorph Example

Now, of course, we are only interested in the classes $c_0, \ldots, c_5$ which comprise the self-conjugate subgroup G'. In the example we have partitioned off these 6 classes and, with them, 6 particular irreducible representations. It will be noted, in this example, that the characters of the first

six irreducible representations are actually orthogonal over the classes of G' only, that is

$$\sum_{j=0}^{5} h_i \Psi_i(c_j) \Psi_k(c_j) = M_i \delta_{ik} \quad (i, k = 0, 1, \ldots, 5)$$

$$\text{where } \delta_{ik} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{otherwise} \end{cases}$$

It will be shown in the following section that this orthogonality of a certain set of characters over the self-conjugate subgroup G', is a general property, which leads to the proof of the existence of a complete basis for all complex functions over the classes of m-ary sequences.

5. Induced Characters of the Abelian Self-Conjugate Subgroup

In our example we imbedded the Abelian group of binary sequences in the holomorph and saw that it was divided up into self-conjugate classes. Furthermore, in the example we noticed a partitioning of the array of characters where certain characters formed an orthogonal set of functions over the classes of the Abelian group only. These representations, we say, are induced onto the group G by the group of transformations T. G' has been shown to be a self-conjugate subgroup of Z (IV-4). We now investigate the class structure and consequent characters induced upon it.

It was shown that the group, G, of m-ary sequences was simply isomorphic to the subgroup, G', of Z. As a matter of terminology, we shall use G, to mean the group of m-ary sequences, which is of course Abelian, whose irreducible representations are all of first degree, and may be given in the form III-8. G' will be the group abstractly identical to G, which is a subgroup of Z. Representations of G', as a subgroup of Z, are actually representations of Z, for a particular element occurring in G'. As a further simplification of notation, since to every element s, in G, there corresponds an element (s, I) in G', we shall refer to the element s, as occurring in either group interchangeably. Let the i'th irreducible character of Z for some element s, in G', be $\Psi_i(s)$, and let the j'th irreducible character of G

for the corresponding element be $\psi_j(s)$. In any matrix representation of Z, the subset of matrices representing the elements of G' give a representation for G. Hence for any element s, in G, the irreducible characters of Z are either irreducible or compound characters of G. Thus we write

$$\Psi_i(s) = \sum_j b_{ij}\, \psi_j(s) \,. \qquad \text{IV - 6}$$

But from III-8, the irreducible characters of G can be substituted to give

$$\Psi_i(s) = \sum_{s_j \epsilon G} b_{ij}\, \sigma^{(s\, s_j^t)} \,.$$

Now since any two elements in the same complete conjugate class of Z, say s, and $s\,\Upsilon$, have the same character

$$\Psi_i(s) = \Psi_i(s\,\Upsilon) = \sum_{s_j \epsilon G} b_{ij}\, \sigma^{s s_j^t} = \sum_{s_j \epsilon G} b_{ij}\, \sigma^{s \Upsilon s_j^t}$$

for all Υ in T,

$$= \sum_{s_j \epsilon G} b_{ij}\, \sigma^{s(s_j \Upsilon^t)^t} \,.$$

for all Υ in T,

So the coefficient b_{ij} has the same value for s_j and $s_j \Upsilon^t$ for all Υ in T.

Letting

$$b_{ij} = \begin{cases} 1 & \text{if } s_j \Upsilon^t \text{ is in the i' th orbit of G induced by } T^t \\ 0 & \text{otherwise} \end{cases}$$

we obtain one character of Z for each orbit of G induced by T^t, namely

$$\Psi_i(s) = \Psi_{s_i}(s) = \frac{h'_i}{f} \sum_{\Upsilon \epsilon T} \psi_{(s_i \Upsilon^t)}(s) = \frac{h'_i}{f} \sum_{\Upsilon \epsilon T} \sigma^{s(s_i \Upsilon^t)^t} \qquad \text{IV - 7}$$

where h'_i is the number of elements in the i' th orbit of G induced by T^t, and f is the order of the group T or T^t.

Alternatively, suppose the holomorph had been formed with G and T^t, the the above derivations would have given the corresponding characters of Z', as

$$\Psi'_{s_j}(s) = \frac{h_j}{f} \sum_{\Upsilon \epsilon T} \sigma^{s(s_j \Upsilon)^t}$$

where h_j is the number of elements in the j' th orbit of G induced by T. Thus we have the relationship

$$\frac{\Psi'_{s_j}(s_i)}{h_j} = \frac{\Psi_{s_i}(s_j)}{h'_i} \qquad \text{IV - 8}$$

The characters IV-7 of Z, one for each orbit of G induced by T^t may be shown to be orthogonal over the classes of Z which occur in G' . These classes of course correspond to the orbits of G induced by T.

$$\sum_{s \epsilon G} h_s \Psi_{s_i}(s) \Psi_{s_k}(s) = \sum_{s \epsilon G} h_s \frac{h'_i}{f} \sum_{\Upsilon \epsilon T} \sigma^{s(s_i \Upsilon^t)^t} \frac{h'_k}{f} \sum_{\Upsilon \epsilon T} \sigma^{s(s_k \Upsilon^t)^t}$$

If s_i and s_k are in different orbits of G induced by T^t, the last two sums are summed over different sets of irreducible representations of G, and hence the sum over $s \epsilon G$ is zero. If s_i and s_k are in the same orbit, there are $h'_i = h'_k$ contributions to the sum, so we have the result

$$\sum_{s \epsilon G} h_s \Psi_{s_i}(s) \Psi_{s_k}(s) = h'_i \, g \, \delta_{ik} \qquad \text{IV - 9}$$

where

$$\delta_{ik} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{otherwise} \end{cases}$$

$g =$ order of G.

We have shown that the irreducible characters of Z, corresponding to the orbits of G induced by T^t, are all orthogonal to each other over the classes

of G' only. Consequently, the number of representations is less than or equal to the number of classes. From the reasoning leading to IV-8, however, the converse holds. Namely, that the number of orbits of G induced by T is less than or equal to the number of orbits of G induced by T^t. Thus there is an equal number and the orthogonal functions corresponding to the orbits of G induced by T^t (T) form a complete basis for complex valued functions over the classes of G' induced by T (T^t).

6. Self-Conjugate Subgroups of G', and Group Codes

An m-ary Group Code (n, k) is a set of m-ary sequences of length n, which form a group, H, of order m^k. If G' contains a proper subgroup H which is self-conjugate in Z, H consists of complete conjugate classes and may thus comprise the set of transmittable (code) sequences. In this case, we say H is a Group code, invariant to the group of transformations T. This assures that a minimum distance from any class in G, to H, may be defined.

Let the elements of G be arrayed with the elements $s_1 = I, s_2, \ldots, s_h$ of the self-conjugate subgroup H in the first line and cosets of H in subsequent lines, as follows:

$$
\begin{array}{lllll}
s_1 = I, & s_2, & s_3, & \ldots, & s_h \\
t_2 & t_2 s_2, & t_2 s_3, & \ldots, & t_2 s_h \\
t_3 & t_3 s_2, & t_3 s_3, & \ldots, & t_3 s_h \\
\cdot & & & & \\
\cdot & & & & \\
\cdot & & & & \\
t_\lambda & t_\lambda s_2, & t_\lambda s_3, & \ldots, & t_\lambda s_h
\end{array}
$$

If a received sequence b, belongs to say line j, then any element of line j added to b gives a code sequence. Hence, decoding consists of selecting that element of line j, of minimum self information I( ) and adding this to b. For simplicity we let the t_j in the above array be these elements of minimum I( ), and call them coset leaders.

If H is the group of code sequences of order m^k, there is in G' another group U, of order m^{n-k}, such that every element of U is orthogonal to every element of H. That is, for every $b \epsilon H$ and $d \epsilon U$, we have

$$b\, d^t = d\, b^t = 0$$

From IV-7, if $d \epsilon U$ and $b \epsilon H$, the character of Z corresponding to d for the element b, is

$$\Psi_d(b) = \frac{h'_d}{f} \sum_{\Upsilon \epsilon T} \sigma^{b(d\Upsilon^t)^t} = \frac{h'_d}{f} \sum_{\Upsilon \epsilon T} \sigma^{0} = h'_d$$

Consequently, for every d in U and b in H,

$$\Psi_d(b) = \Psi_d(I) \qquad \text{IV - 10}$$

But in an irreducible representation of degree n, of any element of order q, the n characteristic roots, $r_1, r_2, \ldots, r_n$, are q'th roots of 1. If the sum of these,

$$\sum_{i}^{n} r_i = n ,$$

then $r_i = 1$ for all i. Consequently, any element, whose character is equal to the character of the identity, is represented by the identity transformation in this representation. From IV-10, then, every element of H, is represented by the identity in the representations $\Psi_d(\,)$, for all d in U. It follows then that every member of a coset, $t_i H$, is represented by the same transformation, and has the same character, in these representations.

If elements of a class of Z belong to two or more cosets, these cosets have the same characters, in the representations corresponding to U. Suppose coset $t_i H$ contains elements from the complete conjugate class, C, of Z. We define

Definition: The class of cosets to which $t_i H$ belongs comprises all those cosets which contain elements from C.

It is clear that the irreducible characters of Z, corresponding to the elements of U, are invariant over any class of cosets. Further, just as in IV-9, these characters are orthogonal over the classes of cosets because they are disjoint sums of the irreducible characters of the group G. We show next that the number of such characters (number of orbits of U induced by T^t) is equal to the number of classes of cosets. We note immediately that because of the orthogonality of the characters, there are equal to or fewer characters than classes of cosets. The converse, namely that there are equal to or fewer classes of cosets than characters, will be illustrated next.

The code group, H, can be taken as the null space of some singular transformation, which without loss of generality can be put in the standard form.

$$D = \begin{array}{|c|}\hline \overleftarrow{\text{n-k}} \\ \hline \\ \text{M} \\ \\ \hline \\ \text{I} \\ \\ \hline \end{array} \begin{array}{l} \\ \} \text{ k} \\ \\ \} \text{ n-k} \end{array}$$

where I is an n-k by n-k identity matrix and M is an n-k by k matrix with elements in the GF(m), such that for any element b, of H, we have $bD = 0$. The columns of D are seen to be the basis elements for the group U, orthogonal to H. Note now that the m^{n-k} elements with all zeros in the first k coordinates form a group κ, such that for every distinct element, b, in κ, bD is distinct, so that each of the elements of κ occurs in a distinct coset of H. Consequently, since the cosets of H are grouped into classes by T, there is a T' which groups the elements of κ into corresponding classes. We then have that the number of classes of κ induced by T' is equal to the number of classes of cosets of H induced by T.

Now also, the orthogonal group U can be taken as the null space of a singular transformation, namely

$$D = \begin{array}{|c|}\hline \overleftarrow{\text{k}} \\ \hline \\ \text{I} \\ \\ \hline \\ \text{M}^t \\ \\ \hline \end{array} \begin{array}{l} \\ \} \text{ k} \\ \\ \} \text{ n-k} \end{array}$$

where I is a k by k identity matrix. It is easily seen that the group θ, of elements with the last n-k coordinates equal to zero occur in distinct cosets of U, and is invariant to the transformations of T^t.

We now reverse the construction of the holomorph by starting with the group G of m-ary sequences and the group of transformations T'^t. In this case it is seen that θ is a group of size m^k invariant to T'^t, having cosets such that distinct elements of U occur in distinct cosets. Then since clearly κ is the group of sequences orthogonal to θ, the representations corresponding to κ are orthogonal over the cosets of U, or what is the same, over the elements of θ. Consequently we are led to conclude that the number of orbits of κ induced by T' is equal to or less than the number of orbits of U induced by T^t. This, together with the previous result shows that the number of classes of U is equal to the number of classes of cosets of H. Since for every orbit of U, there is a character, orthogonal to all the others, there is a complete basis for all complex valued functions over the classes of cosets.

CHAPTER V

The Group of Permutations

1. The Group of Permutations

In the construction of the holomorph Z it was assumed that a group T existed which gave a measure preserving transformation of the m-ary group G. We recall that if

$$r_{ij} = a_i + \zeta_j$$

then $$D(a_i, r_{ij}) = I(\zeta_j) = K \sum_{k=0}^{m-1} b_k \log 1/p_k \qquad V - 1$$

where b_k is the number of coordinates of ζ_j equal to α_k, where α_k occurs with a probability p_k.

Since this measure is based upon only the number, but not the order, of coordinates with a given value, any permutation of the coordinates is measure preserving. We now prove the following theorem.

Theorem: V-1:

If p_0 is the probability of a 0 coordinate and p_α is the probability of an α coordinate such that $p_0 > p_\alpha > p_\beta$ where β is any other field element GF(m), then the only measure-preserving linear transformation group T is a group of coordinate permutations.

Proof:

Let a general non-singular linear transformation Υ, on the n coordinates, in the field GF(m), be

$$\begin{bmatrix} a_{11}, a_{12}, a_{13}, \cdots, a_{1n} \\ a_{21}, a_{22}, a_{23}, \cdots, a_{2n} \\ \vdots \\ a_{n1}, a_{n2}, a_{n3}, \cdots, a_{nn} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = \begin{bmatrix} s'_1 \\ s'_2 \\ \vdots \\ s'_n \end{bmatrix}$$

Let $S = (s_1 = \alpha,\ s_2 = s_3 = \ldots = s_n = 0)$,

then $P_S = p_0^{n-1} p_\alpha$, where P_S is clearly the highest probability sequence except for the sequence of all zero which would have $P_{S_0} = p_0^n$. Any other sequence except for another with just one coordinate $= \alpha$ would have a smaller P_S. Hence, any measure-preserving transformation must yield an S' with just one α, and the remaining coordinates $= 0$. Since

$$S' = s'_1 = a_{11}\alpha$$
$$s'_2 = a_{21}\alpha$$
$$\cdot$$
$$\cdot$$
$$\cdot$$
$$s'_n = a_{n1}\alpha$$

clearly, only one element of the first column = 1, the remainder zero. Choosing $S = (s_1 = 0,\ s_2 = \alpha,\ s_3 = s_4 = \ldots = s_n = 0)$, we show that the same is true of the second column. Hence, by induction, each column has only one element = 1, the rest = 0. Since the transformation is non-singular, these 1's must also be spread one to each row, hence the only permissible transformations are permutations.

Q. E. D.

For any permutation Υ, the inverse is just the transpose, hence if a group T contains Υ, it also contains Υ^t. Therefore, the references of the previous chapter to the group of Υ^t can be replaced by Υ, when T is the measure-preserving group of permutations.

The group T is a group of permutations on the n symbols from $GF(m = p^q)$. If $m = p$, prime, these symbols can be taken as the integers mod. p. We saw that if $q \neq 1$, one could represent each symbol as a sequence

of length q of integers mod. p. It was convenient to take this view when calculating characters for the group.

It should be borne in mind, however, that the permutation group T is still a group on n symbols. Hence, viewed as a group on the nq symbols, it permutes these nq symbols in n sets of q symbols each, without permuting within any set. A cyclic permutation on n = 3, q = 2 represented as a group on 6 symbols would be:

0	0	0	0	1	0
0	0	0	0	0	1
1	0	0	0	0	0
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0

2. Choice of Permutation Groups

We have seen that, in general, the number of distinct classes of m-ary sequences is reduced when the order of T is large. This would imply that one should use the largest permutation groups possible. We show now that this is not the unrestricted best answer.

The largest permutation group on n symbols is the Symmetric Group, containing every possible permutation of n symbols, of order n'. In particular, this group will permute every sequence with b_k symbols equal to α_k, $\alpha_k = (0, 1, \ldots, p-1)$ into every other sequence with the same numbers b_k of symbols. While this indeed does minimize the number of classes of p-ary sequences, it always results in code sequences which do not differ from each other sufficiently for good error correction. For instance, if the sequence

$$B = b_1, b_2, b_3, \ldots, b_n$$

is in a class of code sequences, then so is the sequence

$$B' = b_2, b_1, b_3, \ldots, b_n$$

so that a noise sequence having only one coordinate equal to $b_1 - b_2$, and the other n - 1 coordinates equal to zero, would be sufficient to confuse the two possible transmitted sequences, B and B'. For the binary case this is equivalent to a code which detects, but does not correct, single errors.

Similarly, if one considers the alternating permutation group of order $n!/2$, consisting of all permutations composed of an even number of alternations, the difference between code sequences is small. For example, consider B above to be in a class of code sequences. Then the permutation which interchanges the first and second components and then the second and third, belongs to the alternating permutation group, hence the sequence

$$B'' = b_2, b_3, b_1, \ldots, b_n$$

is also in the class. Note that this differs in three coordinates from B. If the sequences are binary, however, at most two of these differences can be a 1, so that there is still no error correcting capability.

In the search for large permutation groups, one finds two classes commonly studied. One class, the Mathieu groups[12], is unusual because of its high degree of transitivity. However, there are only 5 such groups:

degree	order	transitivity
11	$11 \cdot 10 \cdot 9 \cdot 8$	4-fold
12	$12 \cdot 11 \cdot 10 \cdot 9 \cdot 8$	5-fold
22	$22 \cdot 21 \cdot 20 \cdot 16 \cdot 3$	3-fold
23	$23 \cdot 22 \cdot 21 \cdot 20 \cdot 16 \cdot 3$	4-fold
24	$24 \cdot 23 \cdot 22 \cdot 21 \cdot 20 \cdot 16 \cdot 3$	5-fold

As an example, the Golay[9] 23-bit binary code can be gotten from the Mathieu groups (the 4-fold group of degree 23)[14]. Prange[8] gave one decoding algorithm based on equivalence classes and Zierler[10] gave another, for this code.

3. Linear Groups

The second, and more extensive, class of permutation groups consists of Linear Groups[15]. These are formed by the permutation of points, lines, etc.,

in a finite space when the coordinates of the space are subjected to the full group of linear homogeneous transformations.

Assume a space of d dimensions, over a finite field GF(m). Every point in the space is represented as a d-tuple of symbols from GF(m), and every linear homogeneous transformation of the space can be exhibited as a d x d non-singular matrix, whose elements belong to GF(m). Clearly, the m^d points of the space are permuted among themselves by any such transformation. Also, of course, so are the lines, since in a linear transformation lines go into lines.

If one considers all those lines of the space which pass through the origin, these are all permuted into one another, since the origin is unchanged in a homogeneous transformation. These lines are referred to as the "points" of a projective geometry. When we consider the Linear Group as a permutation group on these "points", we are led to a class of codes studied by Prange[8].

Consider a finite 3-space over GF(q). The number of points in the space is q^3. Each non-singular linear homogeneous transformation is represented as a 3 x 3 matrix. The first row can be chosen $q^3 - 1$ ways; the second row, $q^3 - q$ ways not multiples of the first; and the third row, $q^3 - q^2$ ways not linear combinations of the first two. Hence, the order of the linear group GL(3, q) is

$$v = (q^3 - 1)(q^3 - q)(q^3 - q^2) = q^3(q - 1)^3(q + 1)(q^2 + q + 1)$$

There are q points on every line, and the number of lines through any one point is $\frac{q^3 - 1}{q - 1} = q^2 + q + 1$. Hence, this is the number of lines through the origin, and these lines are permuted among themselves. Consequently, we say that GL(3, q) has a representation as a permutation group on $q^2 + q + 1$ symbols. If we let $n = q^2 + q + 1$ be the length of m-ary sequences, then GL(3, q) is a measure-preserving group of transformations, and hence, possibly suitable as a starting place for the construction of error-correcting codes.

CHAPTER VI

Analysis of Binary Codes

In this chapter we shall work some simple examples to illustrate the general ideas developed so far. Because of the widespread interest in codes for the binary channel and because they are the easiest to handle, numerically, we shall restrict our attention to these. The first example is based on the holomorph constructed in the example of Section IV-3.

Example

Consider the holomorph constructed in Section IV-3, in particular, the table of characters, Table IV-1. We reproduce here for convenience, that portion of the table relating to the characters of the representations $\Gamma_0, \ldots, \Gamma_5$, of classes $d_0, \ldots, d_5$.

	Γ_0	Γ_1	Γ_2	Γ_3	Γ_4	Γ_5
d_0	1	1	2	4	4	4
d_5	1	1	2	4	-4	-4
d_3	1	1	2	-4	0	0
d_2	1	1	-2	0	0	0
d_1	1	-1	0	0	+2	-2
d_4	1	-1	0	0	-2	2

Table VI - 1

Segment of Table IV - 1

In addition, we construct the multiplication table of classes by inspection of the classes set out in Section IV-3. The table is clearly symmetric for the binary case since if $a + \eta = b$, $a = b + \eta$, thus only one half the table is filled in.

	d_0	d_1	d_2	d_3	d_4	d_5
d_0	d_0	d_1	d_2	d_3	d_4	d_5
d_1		$4d_0 + 2d_2 + 2d_3$	$2d_1 + 2d_4$	$d_1 + d_4$	$2d_2 + 2d_3 + 4d_5$	d_4
d_2			$4d_0 + 4d_3 + 4d_5$	$2d_2$	$2d_1 + 2d_4$	d_2
d_3				$2d_0 + 2d_5$	$d_1 + d_4$	d_3
d_4					$4d_0 + 2d_2 + 2d_3$	d_1
d_5						d_0

Table VI - 2

Class Multiplication Table

In the binary channel it is customary to use ' weight' (wt.), the number of ones in a binary sequence in place of I, the self information, and to use in place of the defined distance function, the Hamming[12] distance, equal to the number of coordinates of one sequence that differ from those of another sequence. It is easy to see that for the binary channel these functions possess the property of being monotonic functions of probability and hence theorems II - 1 and V - 1 hold, with appropriate modifications in notation. Throughout the examples of this chapter we shall use weight and Hamming distance.

Referring to the above table, it is seen that the minimum distance from any given class to any other is just the weight of the class (with minimum weight) appearing in the appropriate box. For instance, the distance from class d_1 to d_3 is the minimum of the weights of the classes d_1 and d_4 namely 1, the weight of class d_1. A similar table can be formed giving the minimum distance from any given class to any other.

	d_0	d_1	d_2	d_3	d_4	d_5
d_0	0	1	2	2	3	4
d_1		0	1	1	2	3
d_2			0	2	1	2
d_3				0	1	2
d_4					0	1
d_5						0

Table VI - 3

Class Minimum Distances

Now if one chooses certain classes to comprise the set of code sequences, a dictionary of minimum distance to a code sequence for every class can be set up. For instance, suppose the code sequences were to be the members of classes, d_3, d_4, namely

1 0 1 0 1 1
0 1 1 1 0 1
1 0 1 1 1 0
0 1 0 1 1 1

These do not form a group, hence the group is not a group code. (It is not a good error corrector either, of course, but this is a consequence of choosing a very simple example.) The dictionary to be used in decoding would look as follows:

Class	Min. distance to code sequence D'	Ψ
d_0	2	8
d_1	1	2
d_2	1	-4
d_3	0	4
d_4	0	-2
d_5	1	0

Table VI - 4

Decoding Dictionary

One last step remains. Namely to recognize, for every received sequence, into which class it falls. To do this we construct a compound character which is sufficient to distinguish at least those classes with different D'. For instance, suppose we use the representation

$$\Gamma = 2\,\Gamma_2 + \Gamma_4$$

having the character shown at the right of Table VI-4. This character distinguishes every class, giving a unique entry into the dictionary. The actual mechanics of computing the characters of Γ_2, Γ_4 can be done several ways; however, straightforward use of equation III-8 will serve for illustration.

Let the sequences of d_3 corresponding to Γ_2, and d_1 corresponding to Γ_4, be written as matrices A_2, A_4.

$$A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad A_4 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

Now any received sequence b can premultiply these matrices to give bA_2 and bA_4. Each component of either of these vectors is the inner product of b with a vector in d_2 or d_4 and hence raising (-1) to that power gives the character of b in the representation corresponding to that vector.

If $\quad bA_2 = x_1, x_2, \ldots, x_n$

$\quad bA_4 = y_1, y_2, \ldots, y_n$

then the characters of $\Gamma_2, \Gamma_4,$ are

$$\Psi_2 = \sum_{i=1}^{n} (-1)^{x_i} \qquad \Psi_4 = \sum_{i=1}^{n} (-1)^{y_i}$$

and: $\quad \Psi = 2\Psi_2 + \Psi_4$

For instance, if the sequence b = 1100 (in class d_2) is received

$$bA_2 = 11$$

$$bA_4 = 0110$$

$$\Psi_2 = (-1)^1 + (-1)^1 = -2$$

$$\Psi_4 = (-1)^0 + (-1)^1 + (-1)^1 + (-1)^0 = 0$$

$$\Psi(b) = 2(-2) + 0 = -4$$

hence, (b) has a min. distance 1 from a code sequence.

Example:

Now consider the previous example in the case of a group code. Notice that Γ_2 and Γ_3 have the same value for d_5 as for d_0, hence there is a self-conjugate subgroup. Written out in the coset array we have:

0	1	0	1	A
0	0	1	1	
0	1	0	1	
0	0	1	1	
1	0	1	0	coset 1
0	0	1	1	
0	1	0	1	
0	0	1	1	
0	1	0	1	coset 2
1	1	0	0	
0	1	0	1	
0	0	1	1	
1	0	1	0	coset 3
1	1	0	0	
0	1	0	1	
0	0	1	1	

Now the sequences orthogonal to A are:

0	1	0	1
0	0	1	1
0	1	0	1
0	0	1	1

hence, $A^{\perp} = A$. $A^{\perp}$ comprises classes d_0, d_3, d_5, hence the corresponding representations Γ_0, Γ_1, Γ_2 must form a complete set of orthogonal functions over the classes of cosets. The following table illustrates this:

	Γ_0	Γ_1	Γ_2
A	1	1	2
coset 1, 2	1	-1	0
coset 3	1	1	-2

where coset 1 and coset 2 are in the same coset class, as can be seen, since each coset contains only parts of classes d_1 and d_4. The character Ψ_2 corresponding to Γ_2 distinguishes all classes of cosets, hence the dictionary would be set up as:

Ψ_2	D'
2	0
0	1
-2	2

Ψ_2 is evaluated as before. Define A_2, the matrix of sequences in class d_3 corresponding to Γ_2

$$A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

then if the received sequences is b, and $bA_2 = x_1, x_2,$

$$\Psi_2(b) = \sum_{i=1}^{2} (-1)^{x_i}$$

for instance, if $b = 1110$ $\quad bA_2 = 01 \quad \Psi_2(b) = (-1)^0 + (-1)^1 = 0$

hence, b is in coset 1 or 2, at a distance 1 from the closest code sequence.

In the previous two examples, we began essentially with a knowledge of the permutation group and derived the classes and characters from that. Actually, in the end, all we really needed were the classes of sequences themselves, since the characters can be gotten directly from them. In fact, for group codes one could start with a knowledge of the classes of the orthogonal group only and work up from there. This is often what is required when the group code is given in terms of generator sequences.

Consider a set of generator sequences given in the standard form as rows of an encoding matrix.

$$k \begin{bmatrix} \overset{k}{\overbrace{1 \quad\quad}} & \vdots & \overset{n-k}{\overbrace{1\,1\,0\,\cdots}} \\ \quad 1 \quad\quad & \vdots & \cdots\cdots \\ \quad\quad \cdot \quad\quad & \vdots & \cdots\cdots \\ \quad\quad\quad \cdot & \vdots & \cdots\cdots \\ \quad\quad\quad\quad 1 & \vdots & 0\,0\,1\,\cdots \end{bmatrix} = \left[\begin{array}{c:c} I & M \end{array}\right] = E$$

then clearly a matrix of the form

$$\left[\begin{array}{c} M \\ \hdashline I \end{array}\right] = D$$

contains n-k linearly independent columns each orthogonal to the rows of the encloding matrix, since

$$E \cdot D \equiv 0 .$$

Hence, since the rows of E are the generators of the group of code sequences A, the columns of D are the generators of the group of orthogonal sequences $A^{\perp}$. We see further that if b is a received sequence

$$bD = \begin{cases} \overbrace{0\,0\,0\cdots 0}^{n-k} & \text{if } b \in A \\ 1\,0\,1\cdots 0 & \text{if } b \notin A \end{cases}$$

where the sequences bD are different for every coset. Hence, every b of the form

$$b = \overbrace{0\,0\,0\cdots 0}^{k} \quad \overbrace{1\,0\,1\,0\cdots 0}^{n-k}$$

is in some coset, and the 2^{n-k} distinct b' s of this form can be taken as representatives of the cosets. We let the last n-k bits of these sequences be called $v_1, v_2, \ldots, v_{2^{n-k}}$.

The representations corresponding to the classes of the orthogonal group $A^{\perp}$ must provide a complete set of characters over the classes of cosets. Their different values on these coset classes can be determined from their values on the v_i, since one v_i belongs to each coset.

Example:

Consider a Slepian[7] code (n = 9, k = 5) with generators.

$$E = \begin{bmatrix} 1\,0\,0\,0\,0\,1\,1\,1\,1 \\ 0\,1\,0\,0\,0\,1\,1\,1\,0 \\ 0\,0\,1\,0\,0\,1\,1\,0\,1 \\ 0\,0\,0\,1\,0\,1\,0\,1\,1 \\ 0\,0\,0\,0\,1\,0\,1\,1\,1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1\,1\,1\,1 \\ 1\,1\,1\,0 \\ 1\,1\,0\,1 \\ 1\,0\,1\,1 \\ 0\,1\,1\,1 \\ 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{bmatrix}$$

The orthogonal group $A^{\perp}$, of which the columns of D are generators can be written out, grouped according to class.

0	1 1 1 1 1 1 1 1	0 0 0 0 0 0	0
0	1 1 1 0 1 0 0 0	0 0 1 1 0 1	1
0	1 1 0 1 0 1 0 0	0 1 1 0 1 0	1
0	1 0 1 1 0 0 1 0	1 1 0 0 0 1	1
0	0 1 1 1 0 0 0 1	1 0 0 1 1 0	1
0	1 0 0 0 1 1 1 0	1 0 0 1 1 0	1
0	0 1 0 0 1 1 0 1	1 1 0 0 0 1	1
0	0 0 1 0 1 0 1 1	0 1 1 0 1 0	1
0	0 0 0 1 0 1 1 1	0 0 1 1 0 1	1
C_0	C_1	C_2	C_3

It is easily seen that there do exist permutations which take say the first element of C_2 into every other element of C_2 while leaving the group $A^{\perp}$ unchanged, etc., for the elements of all the classes.

We now construct the table of characters for the v_i using the representations corresponding to C_0, C_1, C_2, C_3.

v	Ψ_0	Ψ_1	Ψ_2	Ψ_3	
0 0 0 0	1	8	6	1	A
1 0 0 0	1	0	0	-1	
0 1 0 0	1	0	0	-1	
0 0 1 0	1	0	0	-1	
0 0 0 1	1	0	0	-1	C_a
1 1 1 0	1	0	0	-1	
1 1 0 1	1	0	0	-1	
1 0 1 1	1	0	0	-1	
0 1 1 1	1	0	0	-1	
1 1 0 0	1	0	-2	1	
0 1 1 0	1	0	-2	1	
0 0 1 1	1	0	-2	1	C_b
1 0 0 1	1	0	-2	1	
1 0 1 0	1	0	-2	1	
0 1 0 1	1	0	-2	1	
1 1 1 1	1	-8	6	1	C_d

Hence, the 16 cosets break into 4 classes; A is clearly the code sequences. Classes C_a, C_d are the rows of D, hence are cosets with leaders of wt. 1. Class C_b has coset leaders of wt. 2. This is a close-packed, single error-correcting code. The above table is written in terms of the classes of cosets.

Class	Ψ_0	Ψ_1	Ψ_2	Ψ_3	
A	1	8	6	1	code sequence
C_a	1	0	0	-1	wt 1
C_d	1	-8	6	1	wt 1
C_b	1	0	-2	1	wt 2

If the sets C_1, C_3 are taken as matrices, then a received sequence, b, has a character

$$\Psi = \Psi_1 + \Psi_3 = \sum_{i=1}^{8} (-1)^{x_i} + (-1)^{y}$$

where

$$b\,C_1 = x_1, x_2, \ldots, x_8$$

$$b\,C_3 = y$$

and Ψ distinguishes the classes.

Example:

As a final example, consider the Bose-Chaudhuri[4] (n = 12, k = 7) code. If the non-zero elements of a $GF(2^4)$ are written $1 = \sigma^{15}, \sigma^1, \sigma^2, \ldots, \sigma^{14}$, they can be put in one-to-one correspondence with the non-zero binary sequences of length 4, so that addition of the binary sequences mod. 2, component by component corresponds to addition of the field elements. The code then is determined by specification of the decoding matrix D, which may be exhibited either in terms of the σ^i or in terms of binary sequences, as follows.

$$D = \begin{vmatrix} 1 & , 1 \\ \sigma & , \sigma^3 \\ \sigma^2 & , \sigma^6 \\ \sigma^3 & , \sigma^9 \\ \sigma^4 & , \sigma^{12} \\ \sigma^5 & , 1 \\ \sigma^6 & , \sigma^3 \\ \sigma^7 & , \sigma^6 \\ \sigma^8 & , \sigma^9 \\ \sigma^9 & , \sigma^{12} \\ \sigma^{10} & , 1 \\ \sigma^{11} & , \sigma^3 \\ \sigma^{12} & , \sigma^6 \\ \sigma^{13} & , \sigma^9 \\ \sigma^{14} & , \sigma^{12} \end{vmatrix} = \begin{vmatrix} 1000, 1000 \\ 0100, 0001 \\ 0010, 0011 \\ 0001, 0101 \\ 1100, 1111 \\ 0110, 1000 \\ 0011, 0001 \\ 1101, 0011 \\ 1010, 0101 \\ 0101, 1111 \\ 1110, 1000 \\ 0111, 0001 \\ 1111, 0011 \\ 1011, 0101 \\ 1001, 1111 \end{vmatrix}$$

It may be verified that the code corrects all single and double errors. If a code sequence b occurs, we have $bD = 0$. If a single error occurs in the received sequence, b, then bD is of the form

$$bD = \sigma^i, \sigma^{3i}$$

whereas, if a double error occurs in the received sequence, b, we have

$$bD = \sigma^i + \sigma^j , \sigma^{3i} + \sigma^{3j}$$

If it is desired to decode only these possibilities, and accept possible errors whenever more than two errors occur, a simple decoding scheme can be devised. Notice that all elements of the form σ^{3i} are 1, σ^3, σ^6, σ^9, or σ^{12}, while all the elements of the form $\sigma^{3i} + \sigma^{3j}$ $(i \neq j)$ are the remaining 10 non-zero elements of the field. Consequently, one need examine only the last 4 bits of bD, to decide between the three possibilities of none, one, or two errors, (ignoring the possibility of more than two errors).

Consideration of these classes and the associated characters leads one to a character which indeed does discriminate between the three different weight classes of errors.

Let

$$M = \begin{vmatrix} 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0 \\ 0\,1\,1\,1\,1 \\ 0\,1\,1\,0\,0 \\ 0\,0\,1\,1\,0 \\ 1\,1\,0\,1\,1 \end{vmatrix}$$

then bDM is of the form

$$bDM = x_1, x_2, x_3, x_4, x_5,$$

We define the character $\psi(b)$ as

$$\psi(b) = \sum_{i=1}^{5} (-1)^{x_i} .$$

Then $\psi(b)$ differentiates the classes by number of errors as follows,

$\psi(b)$	5	-3	1
number of errors	0	1	2

.

Even more simply, these classes are differentiated by $\sum x_i$.

$\sum_{i=1}^{5} x_i$	0	4	2
number of errors	0	1	2

Now suppose two errors actually do occur and one begins the step by step decoding process. If, on a particular trial, the number of errors is increased to three, we must be sure that the above indicators do not indicate that only one error remains. This may be a possibility, since no specific provision was made for the case of more than two errors. For this code it is easily verified that weight three errors do indeed appear the same as weight two errors so far as the last four bits of bD are concerned. Thus one would never make the mistake of increasing the number of errors during the step-by-step decoding process.

This is a point to consider, however, anytime a decoder is being designed to decode step-by-step up to only r errors. Care must be taken that all weight $r + 1$ errors are distinguishable from the $r - 1$ weight errors.

CHAPTER VII

Conclusions

Application of the techniques in this paper to large codes presents practical difficulties. Consider group codes for the binary symmetric channel. To design a decoding scheme it is first necessary to specify, in some manner, the coset leader weights for every coset. For an arbitrarily chosen code, this, in effect, amounts to enumerating the cosets - a task that grows exponentially with the length of the code sequences. For special codes, such as the Reed-Muller[3] or the Bose-Chaudhuri[4] codes, the make-up of the cosets is specified by certain algebraic relationships. One might suspect, in these cases, that the technique would be easier to apply. Unfortunately, so far, this has not been the case because the next step in the process also depends upon enumeration. This is the step of determining the invariant function which specifies to which class a given received sequence belongs. It was shown that this invariant function depends upon the group characters corresponding to the orbits of the group of sequences orthogonal to the code group. While there is a possibility that finding the invariant function can be done more systematically for these special codes, as yet the only sure method is based on strict enumeration. Clearly, here is one place where some further careful investigation may lead to fruitful results.

In the special cases of the cyclic codes studied by Prange[8], it was rather straightforward, from geometric reasoning, to find one invariant function. In the terminology of this paper, the function he used was one of the irreducible characters of the group Z. There is no guarantee, however, that any one irreducible character will distinguish all classes, and indeed this was the case with the character used by Prange. It did turn out, however, that almost all classes could be distinguished, so that a variation on the technique yielded a complete decoding scheme. He found that if one computed the character of the received sequence, and the characters of all its neighbors, that is all sequences differing in only one coordinate, the set of characters was often

sufficient to determine the weight of the coset leader uniquely. This leads to his concept of a 'road map' decoding procedure, in which one visualizes a road map with most of the forks clearly marked, but occasionally one is not. In this case one proceeds down one of the unmarked roads until the direction becomes clear and then returns to take the other direction if necessary.

Such variations on the basic procedure of 'equivalence class decoding' are interesting for at least two reasons. First, it is often quite easy to 'find' one or more invariants without exhausting enumeration, from geometric or other considerations of the code make-up. If these invariants, though not a complete set, can be worked into a reasonable decoding scheme requiring not too many extra steps, the solution may be quite practical. Second, even if one assumes that all the characters have been found by some procedure, it may require a large number of them to actually differentiate all classes. In this case, it may be more economical to use a simpler invariant function, which does not differentiate all the classes, coupled with some sort of road map technique which allows trial and error decoding at those places where ambiguities arise.

One interesting observation can be made in the step-by-step decoding. We state it here for the binary case only. One need never know the weight of the coset leader; it is only necessary to know whether or not the weight has decreased when a given coordinate is changed. For this reason, one requires only three classes of cosets, those whose weights are of the form m, $m + 1$, $m - 1$, for any m. (Decoding ends when a code sequence is reached, and this can be detected by the parity check sequence being identically zero.) A question of theoretical interest arises. How can one most easily divide all sequences into two classes? (Two such divisions are sufficient to give the three classes above.) This is the kind of classification problem that has been studied for some time in the design of logical combinational networks[15]. One wishes to provide a combinational network, which, for every possible n - bit input sequence, the output is either 0 or 1. Experimental work [17] indicates that the

magnitude of the combinational network is often reasonably small. Unfortunately, so far, practical solutions have had to be obtained by an exhaustive enumeration of all possibilities, followed by a well defined, but laborious, simplification of the logical network.

As a last suggestion, one might consider a decoder which decodes correctly only those cosets whose leaders are of weight less than or equal to r, where the code corrects all errors of weight r or less. Present decoding schemes for the Reed-Muller and Bose-Chaudhuri codes are of this type. The author has tried, without much success, to capitalize upon the fact that the coset leaders of weight m (for $m \leq r$) are all the $\binom{n}{m}$ sequences of weight m. While this does make the task of enumerating coset leaders very easy, so far it has not reduced the complexity of computing invariant functions which distinguish the classes. Further work along these lines seems desirable, especially in view of the fact that these known codes do have rather simple decoding schemes for errors of weight r or less, tempting one to suspect that there may be a simple key somewhere if we could only find it.

At the present time, the formal application of techniques in this paper is very difficult. It is hoped, therefore, that this work will be regarded as a framework, which may inspire other ideas and constructions, eventually leading to a unified and practical theory of coding and decoding.

APPENDIX I

Linear Groups

This appendix is concerned with the structure of Linear Groups represented as permutation groups. In Section 1 we shall give certain equations (due to Slepian[13] for the binary case) for the enumeration of classes of sequences formed by the permutation groups. Section 2 is concerned with the group characters and actual computation of class sizes. Section 3 contains tables for various cases.

1. Enumeration Equations

If a permutation group P of order v permutes n symbols, the character of each permutation is just the number of 1's on the main diagonal of the matrix, hence is just the number of symbols not permuted. If these symbols are taken to be the n symbols in a m-ary sequence, the m^n sequences are also permuted among themselves. Here again P has a representation on m^n symbols. If these m^n symbols (sequences) are permuted in N disjoint sets, then this particular representation is reducible into N irreducible representations.

In this representation of $P(1, p_2, p_3, \cdots, p_v)$ let the classes of permutations be $c_1, c_2, \cdots, c_t$ where the character Ψ of each class is just the number of sequences left invariant by the permutations, p_i of the class. Let n_j be the order of class c_j, then

$$\sum_{j=1}^{t} n_j \Psi(c_j) = \sum_{j=1}^{v} R(p_j) \qquad \text{A-I-1}$$

where $R(p_j)$ is the number of sequences left invariant by permutation p_j.

However, for the identity representation which occurs exactly once in each irreducible permutation representation

$$\sum_{j=1}^{t} n_j \Psi_0(c_j) = v \qquad \text{A-I-2}$$

and hence the number of irreducible permutation representation is

$$N = \frac{1}{v} \sum_{j=1}^{t} n_j \Psi(c_j) = \frac{1}{v} \sum_{j=1}^{v} R(p_j) \qquad \text{A-I-3}$$

which is the number of disjoint sets of m-ary sequences.

Now suppose a particular permutation p_j permutes the n sequences in K cycles of length $\lambda_i (i = 1, 2, \cdots, K)$, so that

$$\sum_{i=1}^{K} \lambda_i = n. \qquad \text{A-I-4}$$

We display the m-ary sequences as

	s_0	s_1	s_2	$\cdots$	s_n
f_0	0	0	0		0
f_1	α	0	0		0
.					
.					
.					
f_{n^m}	α^{m-1}	α^{m-1}	α^{m-1}	$\cdots$	α^{m-1}

Now if a cycle of length λ_i permutes the s_i, those sequences f_j with all the same symbol in these λ_i positions will be the only ones left invariant. The fraction of rows (f_j's) with all the same symbol in a given set of λ_i positions is m/m^{λ_i}. A second cycle leaves invariant a fractional number of these, etc. Hence, the total number of sequences left invariant by a permutation with cycles $\lambda_i (i, = 1, 2, \cdots, K)$ is

$$m^n \prod_{i=1}^{K} m/m^{\lambda_i} = m^n \prod_{i=1}^{K} m \cdot m^{-\lambda_i} = m^n m^K m^{-\Sigma \lambda_i}$$

$$= m^n m^K m^{-n} = m^K.$$

Thus, since $\Psi(c_i) = m^{K_i}$

$$N = \frac{1}{v} \sum_{i=1}^{t} n_i m^{K_i}. \qquad \text{A-I-5}$$

This result says the number of sets of sequences is related only to the number K_i of cycles in the permutations of the classes c_i of the group of permutations.

Slepian[13] gives also a generalized result for m = 2. (His development of the above is for m = 2, also, but is easily extended. The generalized result is also extendable to any m, but its notation becomes clumsy.) Of the 2^n binary sequences ($\binom{2^n}{r}$) of them have exactly r, 1's and n-r, 0's (weight = wt. = r). If c_j is a class of permutations with $\lambda_i^{(j)}$ ($i = 1, 2, \cdots, K_j$) cycles, then the number, N_r, of wt. r sequences left invariant by p_j is the number of ways of forming r as a sum of the λ_i (each λ_i occurring only once in a sum). Thus $\Psi_r(p)$ is the coefficient of y^r in

$$\prod_{i=1}^{K_j} \left(1 + y^{\lambda_i^{(j)}}\right) \qquad \text{A-I-6}$$

and thus the number of classes of wt. r sequences is

$$N_r = \text{coefficient of } y^r \text{ in } \frac{1}{N} \sum_{j=1}^{t} n_j \prod_{i=1}^{K_j} \left(1 + y^{\lambda_i^{(j)}}\right) \qquad \text{A-I-7}$$

2. Linear Group Characters and Classes of Binary Sequences

To use the formula (A-I-5) we need to know the class structure of the linear group. We need n_i, the number of permutations in each class, and K_i, the number of cycles of the permutations in each class, when represented as a permutation group on n symbols. L. E. Dickson[15] gave canonical forms for the GL(2, q), GL(3, q) and GL(4, q) and quite recently R. Steinberg[16] determined the characters for these groups. From these analyses, the values of n_i and K_i can be determined.

The canonical forms for the transformations of LG(3, q) are:

$$A_1 = \begin{bmatrix} \rho^a & & \\ & \rho^a & \\ & & \rho^a \end{bmatrix} \quad A_2 = \begin{bmatrix} \rho^a & & \\ 1 & \rho^a & \\ & & \rho^a \end{bmatrix} \quad A_3 = \begin{bmatrix} \rho^a & & \\ 1 & \rho^a & \\ & 1 & \rho^a \end{bmatrix} \quad A_4 = \begin{bmatrix} \rho^a & & \\ & \rho^a & \\ & & \rho^b \end{bmatrix}$$

$$A_5 = \begin{bmatrix} \rho^a & & \\ 1 & \rho^a & \\ & & \rho^b \end{bmatrix} \quad A_6 = \begin{bmatrix} \rho^a & & \\ & \rho^b & \\ & & \rho^c \end{bmatrix} \quad B_1 = \begin{bmatrix} \rho^a & & \\ & \sigma^b & \\ & & \sigma^{bq} \end{bmatrix} \quad C_1 = \begin{bmatrix} \tau^a & & \\ & \tau^{aq} & \\ & & \tau^{aq^2} \end{bmatrix}$$

where ρ, σ, τ are primitive elements of GF(q), GF(q^2), GF(q^3), respectively, such that $\rho = \sigma^{q+1} = \tau^{q^2+q+1}$, $\sigma = \tau^{q^2+1}$. a, b, c are distinct integers; $a \neq$ mult $(q^2 + q + 1)$ in C_1.

The number of classes of each type are:

Element	No. of classes	Elements in each class
A_1	$q - 1$	1
A_2	$q - 1$	$(q - 1)(q + 1)(q^2 + q + 1)$
A_3	$q - 1$	$q(q - 1)^2 (q + 1)(q^2 + q + 1)$
A_4	$(q - 1)(q - 2)$	$q^2(q^2 + q + 1)$
A_5	$(q - 1)(q - 2)$	$q^2(q - 1)(q + 1)(q^2 + q + 1)$
A_6	$\frac{1}{6}(q - 1)(q - 2)(q - 3)$	$q^3(q + 1)(q^2 + q + 1)$
B_1	$1/2\ q(q - 1)^2$	$q^3(q - 1)(q^2 + q + 1)$
C_1	$1/3\ q(q - 1)(q + 1)$	$q^3(q - 1)^2(q + 1)$

TABLE A-I-1

Class Sizes for Linear Groups

The characters of these classes in the representation as a permutation group on $q^2 + q + 1$ symbols is:

Class	Character Ψ
A_1	$\epsilon^{3\alpha a}(q^2 + q + 1)$
A_2	$\epsilon^{3\alpha a}(q + 1)$
A_3	$\epsilon^{3\alpha a}$
A_4	$\epsilon^{\alpha(2a + b)}(q + 2)$
A_5	$\epsilon^{\alpha(2a + b)}\ 2$
A_6	$\epsilon^{\alpha(a + b + c)}\ 3$
B_1	$\epsilon^{\alpha(a + b)}$
C_1	0

$\alpha = 1, 2, \cdots, q - 1$

$\epsilon^{q-1} = 1$

TABLE A-I-2

Class Characters

With this information one can find K_i, the number of cycles in any class, hence the number of sets into which the m-ary sequences fall, as a result of the operation of the group GL(3, q).

Example:

Consider: GL(3, 3).

Table A-I-1 becomes:

Element	No. of classes	No. elements/class	Total elements
A_1	2	1	2
A_2	2	104	208
A_3	2	624	1, 248
A_4	2	117	234
A_5	2	936	1, 872
A_6	0		
B_1	6	702	4, 212
C_1	8	432	3, 456
		Total	11, 232

TABLE A-I-3

Class Sizes for GL(3, 3)

The characters of these classes are:

Class	Ψ	cycles = K
A_{11}	13	13
A_{12}	13	13
A_{21}	4	7
A_{22}	4	7
A_{31}	1	5
A_{32}	1	5
A_{41}	5	9
A_{42}	5	9
A_{51}	2	5
A_{52}	2	5
B_{11}	1	5
B_{12}	1	5

Class	Ψ	cycles = K
B_{13}	1	3
B_{14}	1	3
B_{15}	1	3
B_{16}	1	3
C_{11}	0	1
C_{12}	0	1
C_{13}	0	1
C_{14}	0	1
C_{15}	0	1
C_{16}	0	1
C_{17}	0	1
C_{18}	0	1

TABLE A-I-4

Class Characters and Cycle Structure

Now the cycles of any class can be determined by taking say p in some class and forming p^2, $p^3, \cdots$, 1, in terms of their canonical form. Each time that say p^j changes canonical form to a new class, the character changes and the difference in this and the previous character indicates the number of symbols in the cycle just finished. To continue the example assume p is in class A_{51}, with character 2, indicating 2 symbols are not cycled. Now raise the canonical form A_5 to the j^{th} power

$$(A_5)^j = \begin{bmatrix} \rho^a & 0 & 0 \\ 1 & \rho^a & 0 \\ 0 & 0 & \rho^b \end{bmatrix}^j = \begin{vmatrix} \rho^{ja} & 0 & 0 \\ j\rho^{(j-1)a} & \rho^{ja} & 0 \\ 0 & 0 & \rho^{ja} \end{vmatrix}$$

when $j = 2$ $\quad (A_5)^2 \rightarrow A_2$ character 4

$j = 3$ $\quad (A_5)^3 \rightarrow A_4$ character 5

$j = 6$ $\quad (A_5)^6 \rightarrow A_1$ character 13

Hence, there are 2 cycles of length 1, 1 of length 2, 1 of length 3, 1 of length 6, which we denote as

$$1^2\, 2^1\, 3^1\, 6^1$$

having $K = 2 + 1 + 1 + 1 = 5$ cycles.

The remaining cycles are also listed in Table A-I-4. Use of Equation A-I-5 gives the number of classes of binary sequences

$$N = \frac{1}{11,232} \sum_{i=1}^{t} n_i 2^{K_i}$$

$$= \frac{1}{11,232} \left[2 \cdot 2^{13} + 208 \cdot 2^7 + 1248 \cdot 2^5 + 234 \cdot 2^9 + 1872 \cdot 2^5 + 2 \cdot 702 \cdot 2^5 + 4 \cdot 702 \cdot 2^3 + 3456 \cdot 2^1 \right]$$

$$= 336,960/11,232 = 30 \text{ classes.}$$

The use of Equation A-I-7 can be put in tabular form where entries in each row are to be multiplied by the fraction of elements belonging to the class, $n_i/11,232$, appearing in the right column before adding columns to get the coefficients of y^j in the j^{th} column.

Class	1	y	y^2	y^3	y^4	y^5	y^6	$n_i/11,232$
A_{11}, A_{12}	1	13	78	286	715	1287	1716	2/11,232
A_{21}, A_{12}	1	4	6	7	13	18	15	208/11,232
A_{31}, A_{32}	1	1	0	4	4	0	6	1248/11,232
A_{41}, A_{42}	1	5	14	30	51	71	84	234/11,232
A_{51}, A_{52}	1	2	2	3	3	2	3	1872/11,232
B_{11}, B_{12}	1	1	2	2	3	3	4	1404/11,232
$B_{13} \rightarrow B_{16}$	1	1	0	0	1	1	0	2808/11,232
$C_{11} \rightarrow C_{18}$	1	0	0	0	0	0	0	3456/11,232
	1	1	1	2	3	3	4	

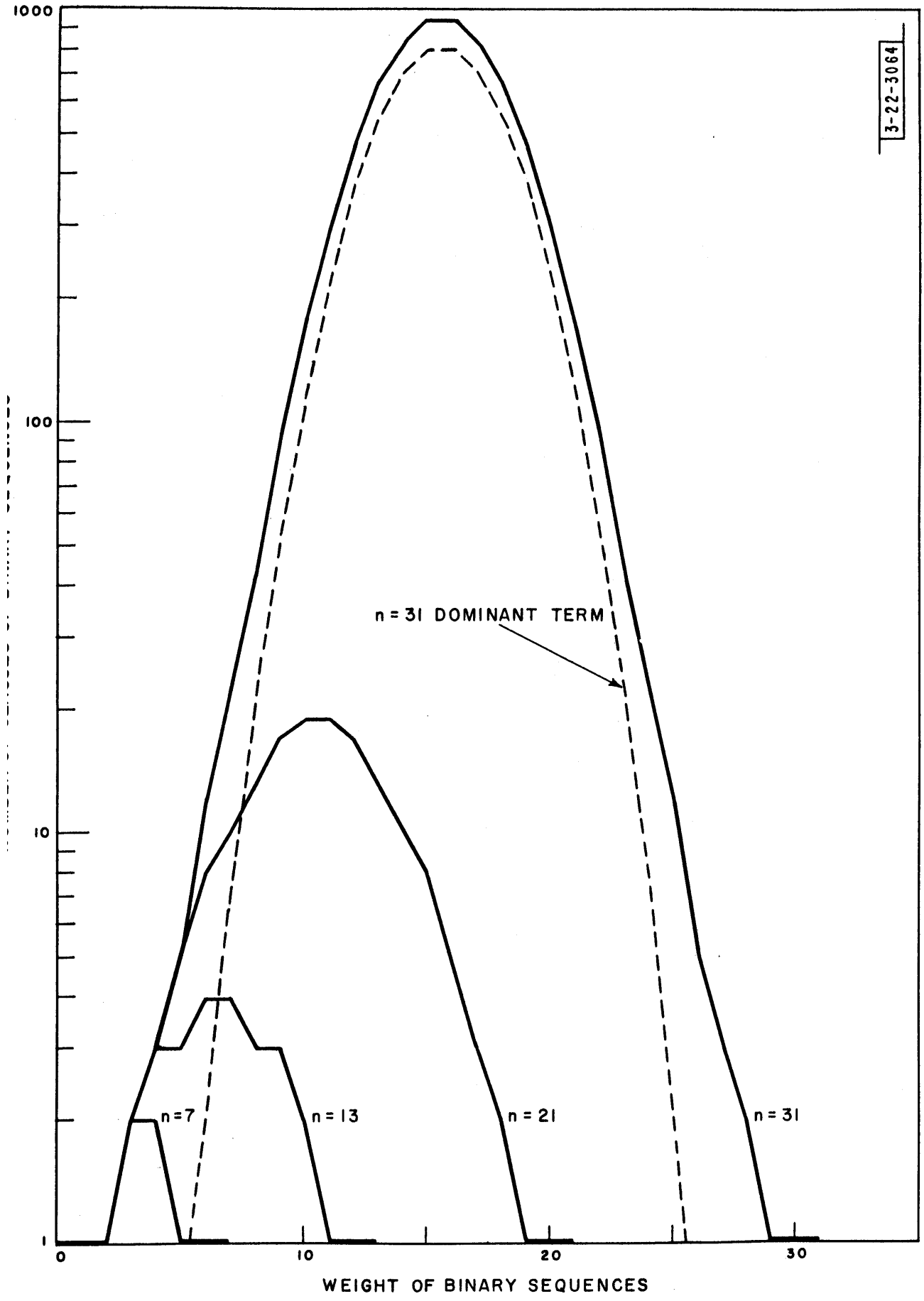
Only classes for binary sequences of wt. $\leq n/2$ are computed since complementory sequences fall in the same sized classes. For other cases of binary sequences arising from

LG(2)	n = 7
LG(3)	n = 13
$LG(2^2)$	n = 21
LG(5)	n = 31

the corresponding tables are given in Section 3.

An interesting graph of number of classes vs wt. of binary sequences can be made for the cases above. Note that these all coincide for the first few lowest wts. The total number of classes is just the sum of all the ordinates under the appropriate curve. If one considers group codes, then the classes of interest are the classes of cosets. If a code corrects no errors of wt. > k, then all coset leaders are of wt. $\leq$ k, and hence the classes of cosets are just those

Figure V-16. Number of Classes of Binary Sequences



under the curve to the left of the abscissa k. If one considers, say the case $n = 31$, and assumes a group code of say $k = 16$ information bits and 15 check bits, there are 2^{15} cosets. Assume further that the code corrects all errors up to say k, and none $> k + 1$. (This would be a close-packed code which does not exist, and is used here just as an example.) In this case, the number of errors ≤ 4 is $> 2^{15}$, hence every coset would be of wt. ≤ 4, hence the number of classes of cosets would be $\leq 1 + 1 + 1 + 2 + 3 = 8$ compared to 7152 classes of received sequences. Of course, such a code specifically does not exist, but the idea illustrated is that the coset leader wts. for good codes are small compared to n, roughly speaking, and hence the number of coset classes is very much smaller than the number of classes of receivable sequences--at least for codes based on these linear groups.

To by-pass the complexity of computing the values plotted in these curves, an approximation may be had by noticing that one term dominates other terms in the sum as q gets large. For the case $q = 5$, $n = 31$, the curve of the dominant term is also plotted. This is simply the contribution to the sum (A-I-7) of the class containing the identity. There are $E = q - 1$ elements in the class, having $n = q^2 + q + 1$ cycles of length $\lambda_i = 1$, the group order is $v = q^3(q - 1)^3(q + 1) \cdot (q^2 + q + 1)$. The dominant term is thus

$$D_q = \frac{E}{v} \prod_{j=1}^{n} (1 + y^{\lambda_i}) = \frac{E}{v}(1 + y)^n \qquad \text{A-I-8}$$

hence, the elements of this term are just a fraction of the binomial coefficients. The total number of classes, due to this term, is

$$N_D = \frac{E}{v} 2^n = \frac{2^{q^2 + q + 1}}{q^3(q - 1)^2(q + 1)(q^2 + q + 1)} . \qquad \text{A-I-9}$$

Whether for the whole number of classes of sequences, or the number of classes by wt., these are lower bounds since each term in the series is positive.

3. Tables of Group Characters and Classes of Binary Sequences

Following the methods of Section 2, the cycle structure of the classes of permutations in some of the linear groups has been worked out and tabulated. Also tabulated are the numbers of classes of binary sequences of each weight which are formed by the permutations of the linear groups. The four cases given are for GL(3, 2), GL(3, 3), GL(3, 4), and GL(3, 5), leading to permutation groups on 7, 13, 21, and 31 symbols, respectively.

GL(3, 2)

Class	Number of Classes	Elements per Class	Total Elements = E	Cycle Structure	Number of cycles = C
A_1	1	1	1	1^7	7
A_2	1	21	21	$1^3 2^2$	5
A_3	1	42	42	$1^1 2^1 4^1$	3
B_1	1	56	56	$1^1 3^2$	3
C_1	2	24	48	7^1	1
Total	6		168		

$$\frac{\sum E\, 2^C}{168} = \text{10 Classes of binary sequences}$$

The classes of binary sequences by weight are:

Weight	0, 7	1, 6	2, 5	3, 4
Classes	1	1	1	2

GL(3, 3)

Class	Number of Classes	Elements per Class	Total Elements = E	Cycle Structure	Number of cycles = C
A_1	2	1	2	1^{13}	13
A_2	2	104	208	$1^4 3^3$	7
A_3	2	624	1, 248	$1^1 3^4$	5
A_4	2	117	234	$1^5 2^4$	9
A_5	2	936	1, 872	$1^2 2^1 3^1 6^1$	5
B_{11}	2	702	1, 404	$1^1 2^2 4^2$	5
B_{12}	4	702	2, 808	$1^1 4^1 8^1$	3
C_1	8	432	3, 456	$(13)^1$	1
Total	24		11, 232		

$$\frac{\sum E\, 2^C}{11,232} = 30 \text{ Classes of binary sequences}$$

The classes of binary sequences by weight are:

Weight	0, 13	1, 12	2, 11	3, 10	4, 9	5, 8	6, 7
Classes	1	1	1	2	3	3	4

GL(3, 4)

Class	Number of Classes	Elements per Class	Total Elements = E	Cycle Structure	Number of cycles = C
A_1	3	1	3	1^{21}	21
A_2	3	315	945	$1^5 2^8$	13
A_3	3	3780	11, 340	$1^1 2^2 4^4$	7
A_4	6	336	2, 016	$1^6 3^5$	11
A_5	6	5040	30, 240	$1^2 2^2 3^1 6^2$	7
A_6	1	6720	6, 720	$1^3 3^6$	9
B_{11}	6	4032	24, 192	$1^1 5^4$	5
B_{12}	12	4032	48, 384	$1^1 5^1 (15)^1$	3
C_{11}	12	2880	34, 560	$(21)^1$	1
C_{12}	6	2880	17, 280	7^3	3
C_{13}	2	2880	5, 760	3^7	7
Total	60		181, 440		

$$\frac{\sum E\, 2^C}{181,440} = 160 \text{ Classes of binary sequences}$$

The classes of binary sequences by weight are:

Weight	0, 21	1, 20	2, 19	3, 18	4, 17	5, 16	6, 15	7, 14	8, 13	9, 12	10, 11
Classes	1	1	1	2	3	5	8	10	13	17	19

GL(3, 5)

Class	Number of Classes	Elements per Class	Total Elements = E	Cycle Structure	Number of cycles = C
A_1	4	1	4	1^{31}	31
A_2	4	744	2,976	$1^6 5^5$	11
A_3	4	14,880	59,520	$1^1 5^6$	7
A_{41}	8	775	6,200	$1^7 4^6$	13
A_{42}	4	775	3,100	$1^7 2^{12}$	19
A_{51}	8	18,600	148,800	$1^2 4^1 5^1 (20)^1$	5
A_{52}	4	18,600	74,400	$1^2 2^2 5^1 (10)^2$	7
A_6	4	23,250	93,000	$1^3 2^2 4^6$	11
B_{11}	16	15,500	248,000	$1^1 6^1 (24)^1$	3
B_{12}	8	15,500	124,000	$1^1 3^2 (12)^2$	5
B_{13}	8	15,500	124,000	$1^1 2^3 8^3$	7
B_{14}	4	15,500	62,000	$1^1 3^2 6^4$	7
B_{15}	4	15,500	62,000	$1^1 3^{10}$	11
C_1	40	12,000	480,000	$(31)^1$	1
Total	120		1,488,000		

$$\frac{\sum E\, 2^C}{1,488,000} = 7152 \text{ Classes of binary sequences}$$

The classes of binary sequences by weight are:

Weight	0, 31	1, 30	2, 29	3, 28	4, 27	5, 26	6, 25	7, 24	8, 23	9, 22	10, 21	11, 20
Classes	1	1	1	2	3	5	12	22	42	92	174	296

Weight	12, 19	13, 18	14, 17	15, 16
Classes	476	669	832	948

APPENDIX II

Topics from the Theory of Finite Groups

1. Introduction

The topics included in this appendix are well known results from the theory of finite groups and hence any good book on the subject can serve as a reference[1, 2, 3*]. It is the purpose of this appendix to present those topics which are of direct use in this report. They are included here only for the sake of convenience of reference. Consequently, no attempt is made to exhaust the subject or to prove theorems not immediately concerned with the topics being presented.

The Definition of a Group[4]

Let G be a system consisting of

1. A set of distinct elements a, b, c, . . .
2. One rule for combining any ordered pair of these elements uniquely. The combination of a and b is indicated by ab (or ba depending upon the order).

The system G is said to form a group if the following conditions are satisfied:

1. If a and b are elements of G, then ab is an element of G (closure).
2. If a, b, and c are elements of G, then a(bc) = (ab)c (associativity).
3. The set G contains a single element I, called the identity, such that for every element a, of G, aI = Ia = a.
4. If a is an element of G, then there is a unique inverse element a', such that aa' = a' a = I.

---

* References for this Appendix are listed separately from those for the remainder of the report because this Appendix is wholly self contained and because the references are specified as to page number in most cases.

Further Definitions

The number of distinct elements in G is called the order of G. If the order is finite, G is called a finite group. These are the only ones with which this report is concerned.

The rule for combining any two elements a, b, in G is called, for simplicity, multiplication. Multiplication of group elements is defined for ordered pairs since $ab \neq ba$ in general. If, however, for every pair of elements a, b, in G, $ab = ba$ then the rule of multiplication is commutative and the group is called Abelian.

If a is an element of a group G, so is $a \cdot a = a^2$. Clearly it is necessary that for finite groups, for every element a there exists an integer i such that $a^i = I$, for otherwise G could not be finite. The smallest positive integer value of i, such that $a^i = I$, is called the order of the element a.

If a set of elements, H, is contained in the group G, and H forms a group subject to the same rules of combination as G, then H is a subgroup of G. H is a proper subgroup of G if its order is greater than 1, but less than the order of G.

Theorem[5] 1: The order of a subgroup of a finite group G, is a factor of the order of G.

Proof: Let G, of order g, have a proper subgroup H, of order h. Arrange the elements of H ($s_1 = I, s_2, \ldots, s_h$) as the first line of an array:

$$
\begin{array}{llll}
s_1 = I & s_2, & s_3, & \ldots, \; s_h \\
t_2 & t_2 s_2, & t_2 s_3, & \ldots, \; t_2 s_h \\
\cdot & & & \\
\cdot & & & \\
\cdot & & & \\
t_\sigma & t_\sigma s_2, & t_\sigma s_3, & \ldots, \; t_\sigma s_h
\end{array}
\qquad \text{A-II - 1}
$$

then choose any other element of G, not in H, say t_2, and form the second line of the array, continuing until all elements of G are exhausted, so that every element of G is in the array at least once. Every element in the array belongs to G and we now show that no element occurs more than once. Assume first that two elements of the same row are equal, i.e.,

$$t_i s_y = t_i s_k$$

but then

$$t_i^{-1} t_i s_y = t_i^{-1} t_i s_k$$

implies

$$s_y = s_k$$

which is impossible, hence all elements in any row are distinct. We now show that all rows are distinct. Take $t_i s_k$ from row i, and $t_j s_m$ from row j, where $j > i$. Now if

$$t_j s_m = t_i s_k \quad \text{then} \quad t_j = t_i s_k s_m^{-1}.$$

Since $s_k s_m^{-1}$ is an element of G, this implies that t_j is in row i. But this is impossible since, by construction, t_j is an element not already used in row i. Therefore, the array contains every element once and only once, thus the number of rows times the number of elements per row must equal the order of G, i.e.,

$$h\sigma = g$$

and thus h and σ are factors of g.

Q. E. D.

In the above array, the first line is the subgroup H, and subsequent lines of elements are called cosets. In this case they are left cosets, since each is of the form $t_i H$. Clearly a construction involving right cosets of the form Ht_i would lead to the same conclusions. Note that for Abelian groups $t_i H = Ht_i$ and there is no distinction between left and right cosets. One other

property of the cosets should be stressed. The coset t_iH is a set of elements formed by left multiplying each element of H by an element of G not in H. Note that the set t_iH is unchanged if any other element of t_iH is used in place of t_i. For we have

$$t_iI, \quad t_is_2, \quad \ldots, \quad t_is_m, \quad \ldots, \quad t_is_m^{-1}, \ldots, \quad t_is_h$$

$$t_is_m, \quad t_is_2s_m, \quad \ldots, \quad t_is_ms_m, \ldots, \quad t_iI, \quad \ldots, \quad t_is_hs_m$$

where the second row contains the same elements as the first. For this reason the choice of t_i is not fixed but can be any element of the coset. When the array is actually written out, the first element is called ' coset leader' . We stress again that choice of the leader of each coset is not dictated by the groups G or H but is arbitrary, within the coset.

Conjugate Elements, Subgroups, and Classes

If a and b are elements of a group G, then (a) and $(b^{-1}ab)$ are said to be conjugate elements. If every conjugate of a in G is equal to a, then a is said to be a self-conjugate element of G.

The elements of G may be divided into sets as follows. Choose an element a of G and form all its conjugates. This is the first set. Choose another element in G not in the first set and form all its conjugates. This is the second set. Continue until G is exhausted. By construction these sets are disjoint and their union comprises G. They are called complete conjugate classes of G.

Suppose the elements of the group fall into complete conjugate classes $C_0 = I, C_1, C_2, \ldots, C_r$. The elements of the class C_i we denote by $s_{i1}, s_{i2}, \ldots, s_{im}$. The product of all the elements of two classes C_i and C_j, is the set of all elements of the form $s_{iu}s_{jv}$. The elements of this set can be arrayed as follows:

$s_{i1}s_{j1}$, $s_{i1}s_{j2}$, . . . , $s_{i1}s_{jn}$

$s_{i2}s_{j1}$, $s_{i2}s_{j2}$, . . . , $s_{i2}s_{jn}$

.

.

.

$s_{im}s_{j1}$, $s_{im}s_{j2}$, . . . , $s_{im}s_{jn}$

Letting these elements be called the set C_iC_j, we now prove:

Theorem[6] 2: The set C_iC_j consists of certain complete conjugate classes C_k, where every element of a class C_k occurs the same number of times, d_{ijk}. Further, each row (or column) of the above array contains an equal number of elements from any given class.

Proof: Form the conjugate of one of the elements in the above array, say

$$s_m^{-1} s_{iu}s_{jv}s_m = (s_m^{-1} s_{iu}s_m)(s_m^{-1} s_{jv}s_m).$$

Now clearly the first factor is in the class C_i and the second factor is in the class C_j. Hence every element of this form is found in the array at least once. Suppose now that some element $s_{iu}s_{jv}$, occurs in the array exactly twice, say $s_{iu}s_{jv} = s_{iw}s_{jx}$. Then since s_{iw} is distinct from s_{iu} and s_{jv} is distinct from s_{jx}, every conjugate of the element $s_{iu}s_{jv}$ can be expressed in two forms.

$$(s_m^{-1}s_{iu}s_m)(s_m^{-1}s_{jv}s_m) \text{ and } (s_m^{-1}s_{iw}s_m)(s_m^{-1}s_{jx}s_m)$$

which both appear in the array distinctly. Conversely, suppose a conjugate of $s_{iu}s_{jv}$ appears in the array more than twice. The same reasoning shows that all of its conjugates must appear more than twice, which implies that $s_{iu}s_{jv}$ occurs more

than twice, contrary to assumption. It follows immediately that if any element occurs d times in the array, each of its conjugates appear d times. To prove the second part of the theorem assume that the first line of the array contains the following elements from the class C_k

$$s_{i1}s_{ja},\ s_{i1}s_{jb},\ \ldots,\ s_{i1}s_{jr}.$$

Now if

$$s_m^{-1}s_{i1}s_m = s_{it}$$

then

$$(s_m^{-1}s_{i1}s_m)(s_m^{-1}s_{ja}s_m) = s_m^{-1}s_{i1}s_{ja}s_m$$

and thus line t contains

$$s_{it}s_m^{-1}s_{ja}s_m,\ s_{it}s_m^{-1}s_{jb}s_m,\ \ldots,\ s_{it}s_m^{-1}s_{jr}s_m.$$

If these are distinct in line number 1, as assumed, they are distinct in line t also, for assume two are equal, then

$$s_{it}s_m^{-1}s_{ja}s_m = s_{it}s_m^{-1}s_{jb}s_m$$

implies $s_{ja} = s_{jb}$ which is a contradiction.

Each row then has the same number of elements from a given class as every other row. A similar consideration proves the same is true for the columns.

Q. E. D.

2. Representation as a Group of Linear Homogeneous Transformations

The properties of a group are often divorced from the elements of a group as such. When this is done we say we are dealing with an abstract group. The properties of an abstract group are given entirely by its multiplication table, showing how any two elements combine to form a third. We no longer consider what the group elements stand for, such as real numbers, vectors, permutations, matrices, etc., only that the elements obey a given multiplication table. One common representation, useful for many purposes, is to

associate each element of a group with an element of a permutation group such that the multiplication table of the permutation group is identical to that of the original group. This brings us to:

Theorem[7] 3: Every finite group G of order g can be represented as a regular permutation group on g symbols, the latter being simply isomorphic to G.

First we define some terms:

A group of permutations on n symbols which contains permutations replacing any given symbol with any other given one is called transitive.

A transitive permutation group on n symbols whose order is also n, is called a regular permutation group.

If G_1 and G_2 are two groups each of order g, and the elements of G_1 and G_2 can be put into one-to-one correspondence such that the product of two elements in G_1 corresponds to the product of the two corresponding elements of G_2 and vice versa, then G_1 and G_2 are said to be simply isomorphic[8].

Proof of theorem 3:

Let $s_1 = I, s_2, \ldots, s_g$ be the elements of G in some given order. Then the g elements $s_i s_1, s_i s_2, \ldots, s_i s_g$ are the same elements arrayed in a different order. Let this permutation of the g elements of G, effected by left multiplication by s_i, be denoted by the symbol

$$S_i = \begin{pmatrix} s \\ s_i s \end{pmatrix} .$$

It is seen that every element s_i of G can be represented by a permutation S_i, and further, that since

$$S_i S_j = \begin{pmatrix} s \\ s_i s \end{pmatrix} \begin{pmatrix} s \\ s_j s \end{pmatrix} = \begin{pmatrix} s \\ s_i s_j s \end{pmatrix}$$

the group of permutations is exhibited as being simply isomorphic to G. The group is clearly transitive since any element s_i can be taken into any other element s_j by one of the permutations, i.e., $S_j S_i^{-1}$. The group is also of order g, and permutes g symbols (the elements of G), hence is a regular permutation group.

Q. E. D.

The permutation group, of course, can be visualized as a group of g x g matrices with one 1 in each row and column, such that matrix multiplication is the combinatorial rule for the group. Thus one has a representation of the abstract group G as a concrete group of g degree matrices. Note that $S_1 = I$ is the only permutation leaving any element of G fixed. Consequently the matrix $S_1 = I$, has all 1's on its principal diagonal while all other matrices S_i, have no 1's on their principal diagonals.

We proceed to consider representations by general homogeneous, non-singular transformations. Consider a linear homogeneous transformation T on the vectors $X = (x_1, x_2, \ldots, x_n)$ of an n-dimensional space S_n, where the x_i are, in general, complex numbers.

$$T : x'_i = \sum_{j=1}^{n} a_{ij} x_j$$

or

$$X' = T X$$

where T is the matrix of the coefficients, a_{ij}. T is said to be non-singular if its n columns (rows) are linearly independent. The determinant of T is zero if and only if T is singular, and T possesses an inverse T^{-1}, if and only if T is non-singular.

Consider a second non-singular transformation B, on the vectors of S_n such that

$$Y = B X \quad \text{and} \quad Y' = B X'$$

Then

$$Y' = BTX = BTB^{-1}Y$$

so that BTB^{-1} is also a transformation on the vectors of S_n.

In a non-singular transformation T, the n linearly independent columns viewed as vectors, form a basis for S_n, in that any vector of S_n can be expressed as a linear combination of the columns. If only $p < n$ columns are linearly independent they form a basis for a p-dimensional subspace, S_p of S_n. Then it is easily seen that T transforms every vector of S_n into a vector of S_p, and in particular, transforms all distinct vectors of S_p into distinct vectors of S_p, so that we say S_p is invariant to the transformation T.

Let a group G have elements $(s_1 = I, s_2, \ldots, s_g)$, and let Σ be a finite group of non-singular transformations, with matrix multiplication as the rule of combination. If the transformations are on the vectors of an n space, and if to each element s_i of G, there corresponds a single element (not necessarily distinct) S_i of Σ, such that when

$$s_i s_j = s_k \quad \text{then} \quad S_i S_j = S_k,$$

Σ is said to be a representation[9] of G of degree n. A particular representation occurs when $S_i = I$ for all S_i in Σ. This is called the identity representation, in which each element of G is represented by the identity transformation.

If another group of non-singular transformations of degree n, Σ' $(S'_1 = I, S'_2, \ldots, S'_g)$ exists such that if T is a non-singular transformation of degree n, and

$$T^{-1} S_k T = S'_k \text{ for every } S_k \text{ in } \Sigma,$$

then also

$$S_k = T S'_k T^{-1}$$

and Σ and Σ' are said to be equivalent representations of G. Otherwise they are distinct.

Reducible and Irreducible Representations[10]

Let $\Sigma(S_1 = I, S_2, \ldots, S_g)$ be a finite group of non-singular transformations of degree n, representing the group G. We display each element S_i in the form of an $n \times n$ matrix of the coefficients of its transformation. If, in addition, there exists a non-singular linear transformation T, such that for <u>every</u> S_i in Σ, the matrix of $T^{-1} S_i T$ is in the form

$$\begin{vmatrix} A_1 & A_3 \\ 0 & A_2 \end{vmatrix}, \qquad \text{A-II - 3}$$

where A_1 is square of degree p, A_2 square of degree q, and $p + q = n$, then we say Σ is reducible. If, in addition, $A_3 = 0$, then Σ is said to be completely reducible. We show later that for finite groups, reducibility implies complete reducibility. That is, if a representation is reducible (or completely reducible) it is equivalent to a representation in which each matrix is in the reduced (or completely reduced) form A-II-3.

3. Properties of Group Representations

It is convenient to prove many of the properties of group representations in terms of representations by unitary transformation matrices (unitary representations). First we give some general properties of unitary and Hermitian matrices, then prove that every representation is equivalent to a unitary representation, and finally prove certain theorems about representations by using these unitary representations.

Let $B^* = \bar{B}^t$, the conjugate complex transpose of B.

Definition: B is unitary if and only if $B^* = B^{-1}$.

It may be shown[11] that unitary matrices have the following properties:

1. The product of two unitary matrices is unitary.
2. The characteristic roots r_i of a unitary matrix all satisfy

$$|r_i|^2 = r_i \bar{r}_i = 1$$

3. A unitary matrix can be transformed into diagonal form by another unitary matrix.

By property 2, it follows that the magnitude of the sum of the characteristic roots for an n x n unitary matrix

$$\left| \sum_{i=1}^{n} r_i \right| \leqq n \ . \qquad \text{A II - 4}$$

Definition: B is Hermitian if and only if $B = B^*$.

It may be shown that [12]

1. A Hermitian matrix is transformed into a Hermitian matrix by any unitary transformation.
2. A Hermitian matrix can be transformed into diagonal form by a unitary matrix.
3. All characteristic roots of a Hermitian matrix are real.

Theorem[13] 4: Any linear representation of a finite group G has an equivalent representation by unitary matrices.

Proof: Let the group G be represented by matrices $A_1 = I, A_2, \ldots, A_g$ of degree n. Form

$$V = \sum_s A_s^* A_s = \sum_s (A_s^* A_s)^* = V^*$$

which is Hermitian. Then since for any non-zero vector X

$$X^*VX = \sum_s X^* A_s^* A_s X$$

is necessarily positive, V is positive definite and its characteristic roots are all positive.

One can form

$$A_j^* V A_j = \sum_s A_j^* A_s^* A_s A_j = \sum_s (A_s A_j)^* (A_s A_j) = V$$

Now let B be a unitary matrix which diagonalizes V, so that

$$B^*VB = \text{diag.}\ (v_1, v_2, \ldots, v_n) \ .$$

Define

$$D = \text{diag.}\ (\sqrt{v}_1,\ \sqrt{v}_2,\ \ldots,\ \sqrt{v}_n)$$

and

$$T = BDB^* \ .$$

Then clearly

$$T^*T = (BDB^*)^*(BDB^*) = B(D^*D)B^* = V,$$

hence

$$A_j^* T^*T A_j = T^*T$$

and

$$T^{*-1}A_j^*T^*T A_jT^{-1} = I$$

$$(T A_jT^{-1})^* (T A_jT^{-1}) = I$$

thus $(T A_jT^{-1})$ is unitary for every A_j in the representation.

Q. E. D.

Theorem[14] 5: A necessary and sufficient condition that a representation be reducible is that the transformations of the representation leave a subspace of the total space on which they operate invariant.

Proof: Let M_i be the transformations of the representations, of degree n. Then by definition there exists a non-singular matrix B, such that $B^{-1}M_iB$ has the following form, for <u>every</u> M_i in the representation

$$B^{-1}M_iB = \begin{vmatrix} A'_i & A'''_i \\ 0 & A''_i \end{vmatrix} \ , \qquad \text{A - II - 5}$$

where A'_i is square of degree p, and A''_i is square of degree q, and $p + q = n$. The first p columns of B, being

linearly independent, form a basis for a p dimensional subspace S_p. Exhibit the rows β_i, of B^{-1} and the columns b_i, of B as follows:

$$B^{-1} = \begin{vmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \end{vmatrix} \qquad B = \begin{vmatrix} b_1 & b_2 & \cdots & b_n \end{vmatrix} .$$

Since $B^{-1}B = I$ the inner product of row β_j and column b_k is 1, if $j = k$, otherwise zero. Now consider the matrix M_iB. All of its columns are linear combinations of the columns b_k of B. It follows then that the only way in which, say, element a_{jk} of $B^{-1}M_iB$ can be zero is that the k' th column of M_iB should not contain the column b_k in its composition. It then follows that the first p columns of M_iB are linear combinations of only the first p columns of B, for otherwise $B^{-1}M_iB$ could not have all elements $a_{jk}(j = p + 1, p + 2, \ldots, n,$ $k = 1, 2, \ldots, p)$ identically equal to zero. Consequently, M_i maps any element of S_p onto another element of S_p, and since M_i is non-singular, the mapping is one-to-one. Conversely, if the M_i leave a subspace S_p invariant, then form a matrix B, whose first p columns form a basis for S_p, and then $B^{-1}M_iB$ will have the form A-II-5.

Q. E. D.

Theorem[13] 6: If a representation by unitary matrices is reducible it is completely reducible.

Proof: Let M_i be the unitary matrices of degree n, of the representation. Then by definition of reducible, there exists a non-singular matrix B, such that for all M_i, $B^{-1}M_iB$ has the following form

$$B^{-1}M_iB = \begin{vmatrix} A_i' & A_i''' \\ 0 & A_i'' \end{vmatrix} .$$

where A_i' is square of degree p, and A_i'' is square of degree q, and $p + q = n$. This implies that there exists a subspace S_p, whose basis elements are the first p rows of B, which is left invariant by M_i. Suppose these p columns are $b_1, b_2, \ldots, b_p$. The complementary subspace S_q, consists of all elements d_k such that

$$\bar{b}_j^t d_k = 0 \quad \text{for} \quad j = 1, 2, \ldots, p.$$

Since $M_i b_j$ is in S_p, then

$$\bar{b}_j^t M_i^* d_k = 0 \quad \text{for all } d_k \text{ in } S_q$$

and thus

$$\bar{b}_j^t (M_i^* d_k) = 0$$

implies that

$(M_i^* d_k)$ is in S_q for all d_k in S_q.

Consequently, if

$$M_i^* d_k = d_v \qquad d_k, d_v \text{ in } S_q$$

then

$$d_k = M_i d_v \qquad d_k, d_v \text{ in } S_q$$

which implies that M_i also leaves the complementary subspace S_q invariant. Now construct a unitary matrix U, with its first p columns from S_p, and its last q columns from S_q. Then clearly M_iU also has its first p columns in S_p and its last q columns in S_q. Thus $U^{-1}M_iU$ is completely reduced to the form

$$U^{-1}M_iU = \begin{vmatrix} X'_i & 0 \\ 0 & X''_i \end{vmatrix}$$

where X'_i is square of degree p, and X''_i is square of degree q.

Q. E. D.

Lemma 6.1: If any representation is reducible it is completely reducible.

Proof: Let A_i be the matrices of degree n of the representation. Then by the definition of reducible, there exists a B such that $B^{-1}A_iB$, has the following form, for all A_i in the representation:

$$B^{-1}A_iB = \begin{vmatrix} X'_i & X'''_i \\ 0 & X''_i \end{vmatrix}$$

Now by theorem 4, there exists a matrix T such that $T^{-1}A_iT = U_i$ is unitary for all A_i. But then

$$(B^{-1}T)\, U\, (T^{-1}B) = \begin{vmatrix} X'_i & X'''_i \\ 0 & X''_i \end{vmatrix}$$

hence, U is reducible, which by theorem 6 means U is completely reducible. Consequently the representation A_i is equivalent to a unitary representation which is completely reducible, hence A_i is completely reducible.

Q. E. D.

Theorem[15] 7: A necessary and sufficient condition for the complete reducibility of a linear representation is the existence of a matrix not of the form kI which commutes with every matrix A_i in the group.

Proof:

Proof of necessity:

If the representation is completely reducible it is equivalent to a representation in which every matrix $B_i = T^{-1}A_iT$ is of the form

$$B_i = \begin{vmatrix} B' & 0 \\ 0 & B'' \end{vmatrix} .$$

where B' is square of degree p, and B'' square of degree q and $p + q = n$.

But then any matrix of the form

$$D = \begin{vmatrix} uI & 0 \\ 0 & vI \end{vmatrix} \begin{matrix} \} p \\ \} q \end{matrix}$$

$$\underbrace{}_{p} \quad \underbrace{}_{q}$$

commutes with B_i, since

$$DB_i = \begin{vmatrix} uB' & 0 \\ 0 & vB'' \end{vmatrix} = B_iD$$

Consequently

$$DT^{-1}A_iT = T^{-1}A_iTD$$

so that

$$TDT^{-1}A_i = A_iTDT^{-1}$$

hence $TDT^{-1} \neq kI$ commutes with every A_i.

Proof of sufficiency:

Assume that for all A_i there exists a $D \neq kI$ such that

$$DA_i = A_iD$$

If D commutes, so does $D' = (D - \zeta I)$ where ζ is an eigenvalue of D. D' is singular which means it maps all of the vectors of the space on which it operates onto a subspace. Denote the whole space as $\tilde{S}$, and the subspace $\widetilde{D'S}$. Then the equation

$$D' A_i = A_i D'$$

implies

$$\widetilde{D' A_i S} = \widetilde{A_i D' S}$$

so that

$$\widetilde{D' S} = A_i \widetilde{(D' S)}$$

Thus A_i maps the subspace $\widetilde{D' S}$ onto itself. We have seen by theorem 5 that the existence of such an invariant subspace implies that the representation is reducible. Consequently the condition of the theorem is also sufficient. Q. E. D.

Theorem[16] 8: If $A_1 = I, A_2, \ldots, A_g$, of degree p, and $B_1 = I, B_2, \ldots, B_g$, of degree q, are two distinct irreducible representations of the group G, and if there exists a rectangular matrix V, with p rows and q columns such that

$$A_i V = V B_i \quad \text{for every } i = 1, 2, \ldots, g$$

then $V = 0$.

Proof: Case 1. $p = q$

If $\det V \neq 0$ we have $A_i = V B_i V^{-1}$ which is a contradiction since the two representations are distinct.

If $\det V = 0$, V defines a subspace which is left invariant by A_i (see the sufficiency proof of theorem 7). By theorem 5, A_i would be reducible, which is a contradiction.

Case 2. $p > q$

V defines a subspace of dimension $\leq q$ smaller than p, implying that A_i leaves a subspace invariant, hence is reducible, which is a contradiction.

Case 3. $p < q$

Consider $V^t A_i^t = B_i^t V^t$, and the argument of case 2 shows B_i^t to be reducible, which in turn implies B_i to be reducible, which is a contradiction. Q. E. D.

We now develop an orthogonality property for the elements of matrices belonging to distinct representations. Let U be any matrix with p rows and q columns. Let A_i be the matrices of degree p of one irreducible representation, and B_i the matrices of degree q, of a distinct irreducible representation, of a finite group G. Assume further that the A_i and B_i are all unitary. Then form[17]

$$V = \sum_i A_i U B_i^{-1}$$

then

$$A_j V B_j^{-1} = \sum_i A_j A_i U (B_j B_i)^{-1} = V \quad .$$

Then by theorem 8, $V = 0$,

so that

$$\sum_i A_i U B_i^{-1} = 0 \quad .$$

Denote the matrix components as: $A_i = a_{st}^{(i)}$; $B_i^{-1} = B_i^* = \beta_{vm}^{(i)}$; $U = u_{tv}$.

The above equation becomes

$$\sum_i \sum_t \sum_v a_{st}^{(i)} u_{tv} \beta_{vm}^{(i)} = 0 \, .$$

Specialization of the components of U leads to the first orthogonality result.

Let

$$u_{tv} = \begin{cases} 1 & \text{if } t = j \text{ and } v = k, \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\sum_i a_{sj}^{(i)} \beta_{km}^{(i)} = 0 \quad \text{for every } s, j, k, \text{ and } m. \qquad \text{A - II - 6}$$

Similarly, if we take

$$V_1 = \sum_i A_i U A_i^{-1}$$

$$A_j V_1 A_j^{-1} = \sum_i A_j A_i U (A_j A_i)^{-1} = V_1$$

so that, by theorem 7, V_1 is a multiple of a unit matrix, $V_1 = KI$.

Denote the matrix components of $A_i^{-1} = A_i^* = \alpha_{vm}^{(i)}$ and we have

$$(KI)_{sm} = \sum_i \sum_t \sum_v a_{st}^{(i)} u_{tv} \alpha_{vm}^{(i)} .$$

To determine the value of K, set $s = m$ and sum over all s.

$$\sum_{s=m} (KI)_{sm} = \sum_i \sum_t \sum_v u_{tv} \sum_s \alpha_{vs}^{(i)} a_{st}^{(i)} .$$

The matrices are of degree p, and the last sum on the right is over the elements of the identity $A^{-1}A$, hence

$$Kp = \sum_i \sum_t \sum_v u_{tv} \delta_{tv} = g \sum_t \sum_v u_{tv} \delta_{tv}$$

where g is the order of the group G, and

$$\delta_{tv} = \begin{cases} 1 & \text{if } t = v \\ 0 & \text{otherwise} \end{cases} .$$

Thus

$$K = g/p \sum_t \sum_v u_{tv} \delta_{tv} .$$

Now if the components of U are

$$u_{tv} = \begin{cases} 1 & \text{if } t = j \text{ and } v = k \\ 0 & \text{otherwise} \end{cases}$$

we have

$$\sum_i a_{sj}^{(i)} \alpha_{km}^{(i)} = g/p \; \delta_{jk} \delta_{sm} \qquad \text{A - II - 7}$$

4. Group Characters

We now define a group character to be the sum of the diagonal elements of the matrix representation. That is

$$\psi_a(s) = \sum_j a_{jj}^{(s)} \; . \qquad \text{A - II - 8}$$

This sum, it will be recalled, is the sum of the characteristic roots of the matrix and thus is unchanged by any non-singular transformation. From the definition of complete conjugate class it follows that the character is the same for every member of a complete conjugate class. We now use the orthogonality results A-II-6 and A-II-7, to prove the first orthogonality property for the group characters. If $\psi_a(s)$ is the character of the group element s, in the a' th irreducible representation we have directly

$$\sum_s \psi_a(s)\psi_b^*(s) = \sum_s \sum_j a_{jj}^{(s)} \sum_i \beta_{ii}^{(s)} = 0 \qquad \text{A - II - 9}$$

Similarly

$$\sum_s \psi_a(s)\psi_a^*(s) = \sum_s \sum_j a_{jj}^{(s)} \sum_i \alpha_{ii}^{(s)} = \sum_j \sum_i g/p \; \delta_{ji} = g \qquad \text{A - II - 10}$$

Theorem[18] 9: A necessary and sufficient condition that two irreducible representations be equivalent is that they have the same character for every element (or class).

Proof: The condition is necessary since non-singular transformations leave the character unchanged.

For sufficiency, assume that the two representations are not equivalent but that the characters are the same. Then, from

A-II-9 and A-II-10 we have

$$\sum \psi_a(s)\, \psi_b^*(s) \;=\; \sum \psi_a(s)\, \psi_a^*(s) \;=\; g$$

which is a contradiction of A-II-9. Q. E. D.

We now introduce the concept of a compound representation which is the sum of several irreducible representations. Let $\Gamma_1, \Gamma_2, \ldots, \Gamma_r$ be distinct irreducible representations and $\psi_1, \psi_2, \ldots, \psi_r$ the respective characters. A compound representation Γ, with character ψ may be exhibited as a sum

$$\Gamma \;=\; \sum_{i=1}^{r} \dot{}\; a_i \Gamma_i \;=\; a_1 \Gamma_1 \dot{+} a_2 \Gamma_2 \dot{+} \ldots \dot{+} a_r \Gamma_r .$$

By definition, the matrix representation of Γ is as follows

$$\Gamma(s) = \begin{vmatrix} \Gamma_1(s) & & & \\ & \Gamma_1(s) & & \\ & & \square & \\ & & & \ddots \\ & & & \Gamma_r \end{vmatrix}$$

where the sub-matrices $\Gamma_i(s)$ occur (in any order) on the principal diagonal a_i times. Clearly, also

$$\psi \;=\; \sum_{i=1}^{r} a_i \psi_i$$

Theorem[19] 10: A given reducible representation may be decomposed uniquely into a sum of its irreducible representations.

Proof: Suppose the character of the given representation is $\psi(s)$, and that it is composed of m irreducible characters $\psi_1(s), \psi_2(s), \ldots, \psi_m(s)$; then

$$\psi(s) = \sum_{i=1}^{m} a_i \psi_i(s)$$

where, by A-II-9 and A-II-10 the a_i are uniquely determined to be

$$a_i = 1 / g \sum_{s} \psi(s) \psi_i^*(s)$$

Q. E. D.

By this theorem on unique decomposability and the previous theorem on equivalent representations, we have shown

Lemma 10: A necessary and sufficient condition for two representations (not necessarily irreducible) to be equivalent is that they have the same characters.

Theorem 11: If a group G has r complete conjugate classes, the number of distinct irreducible representations is less than or equal to r.

Proof: The character of an irreducible representation is a function defined on the r classes. Distinct representations have orthogonal characters, and since the number of orthogonal characters defined on r classes is at most r, the theorem is proved.

Q. E. D.

Let some class of the group be called C_k, having h_k elements, which in a given irreducible representation are called $A_{k1}, A_{k2}, \ldots, A_{kh_k}$. Now form[20] the sum of all these matrices for the class C_k

$$A(C_k) = \sum_{i=1}^{h_k} A_{ki} \quad . \qquad \text{A - II - 11}$$

Then for any A_s we have

$$A_s A(C_k) A_s^{-1} = A(C_k)$$

so that $A(C_k)$ commutes with every matrix A_s in the group. Therefore, by theorem 7, $A(C_k)$ must be a multiple of an identity matrix

$$A(C_k) = m_k I \quad . \qquad \text{A - II - 12}$$

From theorem 2, we recall that if a set is formed of all the products of elements from two classes, C_i and C_j, that set is composed of complete conjugate classes C_k, where every element of a class C_k occurs the same integral number of times d_{ijk}. Consequently

$$A(C_i) A(C_j) = \sum_k d_{ijk} A(C_k)$$

and thus

$$m_i m_j = \sum_k d_{ijk} m_k \quad . \qquad \text{A - II - 13}$$

Now the trace of $A(C_k)$, i.e., the sum of the elements on the principal diagonal, is just the sum of the characters of the A_{ki} in class C_k. There are h_k of these elements all with the same character, $\psi(C_k)$, hence

$$\text{trace } A(C_k) = \text{tr } A(C_k) = h_k \psi(C_k)$$

and

$$\text{tr } A(C_k) = \text{tr } m_k I = n m_k = h_k \psi(C_k)$$

so that

$$m_k = h_k / n \; \psi(C_k) \qquad \text{A - II - 14}$$

where n is the degree of the representation.

Substituting this result into A-II-13, we have

$$h_i \psi(C_i)\, h_j \psi(C_j) = n \sum_k d_{ijk}\, h_k \psi(C_k) \qquad \text{A - II - 15}$$

One particular d_{ijk} will be of importance in the following. If the class C_i contains the inverses of the elements in class C_j, then it is clear that in the set of all products of elements of C_i with those of C_j, the identity element will occur as many times as there are elements in C_i. Otherwise, the identity does not occur at all. If C_j contains the inverses of C_i we denote $C_j = C_{i'}$ and have the result

$$d_{ij1} = h_j \delta_{ji'} \qquad \text{A - II - 16}$$

where h_j is the order of the class C_j and

$$\delta_{ji'} = \begin{cases} 1 & \text{if } j = i' \\ 0 & \text{otherwise} \end{cases} .$$

We are now in a position to show that the regular permutation representation of theorem 3 is composed of r or more irreducible representations if the group has r complete conjugate classes. This result, together with theorem 11 proves that a group with r complete conjugate classes has exactly r distinct irreducible representations. Let G be any finite group with elements $s_1 = I, s_2, \ldots, s_g$ which fall into r complete conjugate classes. In the regular permutation representation, the identity is the only permutation which leaves <u>any</u> element fixed, hence we have for characters

$$\psi(s_1 = I) = g$$

$$\psi(s_i \neq I) = 0 \; . \qquad \text{A - II - 17}$$

Assume that $\psi(s)$ is made up of m distinct irreducible representations $\Psi_1, \Psi_2, \ldots, \Psi_m$, of degrees $n_1, n_2, \ldots, n_m$. Then we have

$$\psi(s) = \sum_{i=1}^{m} a_i \Psi_i(s)$$

where

$$a_i = 1 / g \sum_{s} \psi(s) \Psi_i^*(s) = \Psi_i^*(s_1 = I) = n_i$$

thus:

$$\psi(s) = \sum_{i=1}^{m} n_i \Psi_i(s) \quad . \qquad \text{A - II - 18}$$

Using result A-II-15, we have, for any one of the m irreducible representations $\Psi_{t'}$

$$h_i h_j \Psi_t(C_i) \Psi_t(C_j) = n_t \sum_{k} d_{ijk} h_k \Psi_t(C_k) \; .$$

Summing over t from 1 to m,

$$h_i h_j \sum_{t=1}^{m} \Psi_t(C_i) \Psi_t(C_j) = \sum_{k} d_{ijk} h_k \sum_{t=1}^{m} n_t \Psi_t(C_k)$$

which by A-II-18,

$$= \sum_{k} d_{ijk} h_k \psi(C_k)$$

and from A-II-17 and A-II-16

$$= d_{ij1} g = g h_j \delta_{ji'}$$

giving the final result:

$$\sum_{t=1}^{m} \Psi_t(C_i) \Psi_t(C_j) = g/h_j \delta_{ji'} \qquad \text{A - II - 19}$$

We will now form a set of m homogeneous equations in r unknowns y_j, and show that the only solution to these equations is $y_j = 0$ for all j. This will show that the m equations completely determine the r unknowns so that $m \geq r$, a fact which together with theorem 11, proves m = r. That is the number of distinct irreducible representations equals the number of complete conjugate classes.

Form

$$\sum_{j=1}^{r} y_j \Psi_t(C_j) = 0 \qquad t = 1, 2, \ldots, m$$

and multiply by $\Psi_t(C_i)$ and sum over the m representations.

$$\sum_{t=1}^{m} \Psi_t(C_i) \sum_{j=1}^{r} y_j \Psi_t(C_j) = 0$$

Then from A-II-19, this becomes

$$\sum_{j=1}^{r} y_j \, g/h_j \, \delta_{ji'} = g/h_{i'} \, y_{i'} = 0$$

where the only solution is $y_{i'} = 0$, for every i'. Thus $m \geq r$, for if $m < r$, there would be too few equations and some of the $y_{i'}$ could be set independently. This along with theorem 11, proves the following

Theorem[21] 12: If a finite group has r complete conjugate classes, it has exactly r distinct irreducible representations.

Example A - II - 1

Consider a group of order 10 having the following elements:

$$I, A, A^2, A^3, A^4, B, AB, A^2B, A^3B, A^4B$$

where $A^5 = B^2 = I$ and $AB = BA^4$.

We then have 4 conjugate classes

$$C_0 = I$$

$$C_1 = A, A^4$$

$$C_2 = A^2, A^3$$

$$C_3 = B, AB, A^2B, A^3B, A^4B$$

The multiplication table for these classes (theorem 2) is

	C_0	C_1	C_2	C_3
C_0	C_0	C_1	C_2	C_3
C_1	C_1	$2C_0 + C_2$	$C_1 + C_2$	$2C_3$
C_2	C_2	$C_1 + C_2$	$2C_0 + C_1$	$2C_3$
C_3	C_3	$2C_3$	$2C_3$	$5C_0 + 5C_1 + 5C_2$

The table of Group characters can be written as

	h_i	ψ_0	ψ_1	ψ_2	ψ_3
C_0	1	1	1	2	2
C_1	2	1	1	$-1/2(1+\sqrt{5})$	$-1/2(1-\sqrt{5})$
C_2	2	1	1	$-1/2(1-\sqrt{5})$	$-1/2(1+\sqrt{5})$
C_3	5	1	-1	0	0

We can exhibit the elements A and B as matrices of a permutation group on 5 symbols.

$$A = \begin{vmatrix} 0&1&0&0&0 \\ 0&0&1&0&0 \\ 0&0&0&1&0 \\ 0&0&0&0&1 \\ 1&0&0&0&0 \end{vmatrix} \qquad B = \begin{vmatrix} 0&0&0&0&1 \\ 0&0&0&1&0 \\ 0&0&1&0&0 \\ 0&1&0&0&0 \\ 1&0&0&0&0 \end{vmatrix}$$

Now if we take T of the form

$$T = \begin{vmatrix} a & b & c & d & e \\ a & bw^2 & cw^3 & dw & ew^4 \\ a & bw^4 & cw & dw^2 & ew^3 \\ a & bw & cw^4 & dw^3 & ew^2 \\ a & bw^3 & cw^2 & dw^4 & ew \end{vmatrix} \qquad (w^5 = 1)$$

then $T^{-1} A T =$

$$\left|\begin{array}{c|c|cc} 1 & 0 & 0 & 0 \\ \hline 0 & w^2 & 0 & 0 \\ 0 & 0 & w^3 & 0 \\ \hline 0 & 0 & 0 & w \end{array}\right| \quad \left|\begin{array}{c|cc|c} 0 & 0 & 0 & 0 \\ \hline 0 & 0 & c/bw^2 & 0 \\ 0 & b/cw^3 & 0 & 0 \\ \hline 0 & 0 & 0 & e/dw \end{array}\right|$$

and $T^{-1} B T =$

$$\left|\begin{array}{c|cc|c} 0 & 0 & 0 & 0 \\ \hline 0 & 0 & c/bw^2 & 0 \\ 0 & b/cw^3 & 0 & 0 \\ \hline 0 & 0 & 0 & e/dw \\ 0 & 0 & 0 & d/ew^4 \end{array}\right| \quad \left|\begin{array}{c|c} 0 & 0 \\ \hline 0 & 0 \\ 0 & 0 \\ \hline 0 & 0 \\ d/ew^4 & 0 \end{array}\right|$$

which exhibit the fact that the permutation representation was composed of three of the irreducible representations corresponding to the characters ψ_0, ψ_2, and ψ_3.

The importance of the properties just proved and exemplified can be summarized as follows. We have shown that a group G of order g, made up of say r conjugate classes of order h_i, possess r distinct irreducible representations and corresponding to each there is a character, the set of which form a complete ortho-normal basis for all functions defined over the r classes. These orthogonality relationships are summarized here for ready reference.

$$\sum_{i=1}^{r} h_i \psi_j(C_i) \psi_k^*(C_i) = g \, \delta_{jk} \qquad \text{A - II - 20}$$

$$\sum_{t=1}^{r} \psi_t(C_i) \psi_t(C_j) = g/h_i \, \delta_{i'j} \qquad \text{A - II - 21}$$

$$h_i h_j \psi_t(C_i) \psi_t(C_j) = \psi_t(C_0) \sum_{k} d_{ijk} \psi_t(C_k) \qquad \text{A - II - 22}$$

Biographical Note

Walter I. Wells was born in Kansas City, Missouri, August 16, 1925. The family moved to Oak Park, Illinois when he was fourteen and he graduated from high school there. After working for a year and a half for the Federal Reserve Bank of Chicago he entered the U. S. Army Signal Corps Reserves and attended radio and electronics schools in Ashland, Wisconsin, and Chicago. Shortly after his eighteenth birthday he went on active duty, had basic training at Fort Benning, Georgia, further electronics schooling at Fort Monmouth, New Jersey, and went overseas as a radio technician with the 102 Division Signal Co. He participated in four major battles in Europe and was discharged from service in 1946.

College training began at the Kansas City Missouri Junior College from which he was graduated with an Associate of Science Degree in 1948. Mrs. Wells, whom he married in 1947, is the former Marjorie Merrifield of Breckenridge, Missouri. She accompanied him to Massachusetts where he entered the VI-A cooperative course at M. I. T. Mrs. Wells served as Vail Library accountant while Mr. Wells was an undergraduate. In 1952 he graduated from M. I. T. with a B. S. and S. M. in Electrical Engineering. His Master' s Thesis is entitled The Synthesis of Linear Networks for Prescribed Transient Response.

Mr. Wells then joined the M. I. T. Digital Computer Laboratory as a research assistant. Shortly after this laboratory became part of Lincoln Laboratory, he took on a full time assignment as Section Leader doing analysis and simulation of defense systems problems. The section' s responsibilities were extended to include test planning and evaluation of the Cape Cod bread board model of SAGE. In 1955 he became Associate Group Leader of the Analysis and Evaluation Group at Lincoln responsible for analysis, simulation, test planning, test conduct, and evaluation of the Experimental SAGE Subsector. As this specific task was nearing completion he set up the simulation system for study of BMEWS. His other professional experience includes analysis and design of a signal mixing and distribution panel for T. V. production testing, development of radar receiver circuits, and analysis of color T. V. electron optics, all at Philco Corp. as part of the VI-A cooperative course. He also consulted with Ultrasonics Corp. on noise problems in sampled-data control systems, and for Edgerton, Germeshausen and Grier on design of transmission line equalizers and hydrogen thratron circuitry.

The Wells have two children, Cynthia aged 9, and Gregory aged 3. They have lived in Lexington, Massachusetts, for five years and are active in church work there. Mr. Wells is a senior member of Sigma Xi, and a member of Tau Beta Pi, Eta Kappa Nu, Pi Tau Pi Sigma, and the I. R. E. He was given the gold medal award of the Armed Forces Communications Association as the ' outstanding senior ROTC student majoring in communications and electronics at M. I. T. ' in 1951. He was also given the Pi Tau Pi Sigma Award and designated as a Distinguished Military Graduate of M. I. T. the same year.

References

Appendix II only

1.	Burnside, W.	Theory of Groups of Finite Order, second edition 1911, republication by Dover, 1955
2.	Carmichael, R. D.	Introduction to the Theory of Groups of Finite Order, Copyright, 1937, Dover edition, 1956
3.	Littlewood, D. E.	The Theory of Group Characters, first published, 1940, reprinted Oxford, 1950
4.	Carmichael, op. cit.,	Chapter 1
5.	_____, op. cit.,	Paragraph 10
6.	Burnside, op. cit.,	Paragraph 41
7.	Carmichael, op. cit.,	Paragraph 12
8.	_____, op. cit.,	Paragraph 9
9.	Burnside, op. cit.,	Paragraph 174
10.	Littlewood, op. cit.,	Paragraph 4.3
11.	_____, op. cit.,	Paragraph 1.9
12.	Hildebrand, F. B.	Methods of Applied Mathematics, Copyright, 1952, Prentice-Hall, Inc. Paragraph 1.16
13.	Littlewood, op. cit.,	Paragraph 11.2
14.	Carmichael, op. cit.,	Paragraph 53
15.	Smirnov, W. I. von	Lehrgang der Holeren Mathematik, Vol. III, Pt. 1, First published 1951, German translation published, Berlin, 1954 Paragraph 66
16.	_____, op. cit.,	Paragraph 76
17.	_____, op. cit.,	Paragraph 76
18.	_____, op. cit.,	Paragraph 77
19.	_____, op. cit.,	Paragraph 77
20.	_____, op. cit.,	Paragraph 77
21.	_____, op. cit.,	Paragraph 77

References

1. Shannon, C. E. — The Mathematical Theory of Communication, University of Illinois Press, Urbana, Illinois, 1949

2. Hamming, R. W. — Error Detecting and Error Correcting Codes, B. S. T. J. 26, No. 2, 147(April 1950)

3. Reed, I. S. — A Class of Multiple Error Correcting Codes and the Decoding Scheme, M. I. T. Lincoln Lab. Tech Report No. 44, October, 1953

4. Bose, R. C. and Ray-Chaudhuri, D. K. — On a Class of Error-Correcting Binary Group Codes, to appear in Information and Control.

   Peterson, W. W. — Encoding and Error Correction Procedures for the Bose-Chaudhuri Codes, to appear in I. R. E. Transactions on Information Theory.

5. Elias, P. — Coding for Noisy Channels, I. R. E. Convention Record Part IV, 1955

6. Wozencraft, J. — Sequential Decoding for Reliable Communications, Tech. Report 325, M. I. T. R. L. E., Aug. 1957

7. Slepian, D. — A Class of Binary Signalling Alphabets, B. S. T. J. 35 (1956), pp 203-234

8. Prange, E. — The Use of Coset Equivalence in the Analysis and Decoding of Group Codes, Communication Sciences Laboratory, AFCRC, Bedford, Mass., June 1959

9. Golay, M. J. E. — Notes on Digital Coding, Proc. I. R. E., 37(1949), 657

10. Zierler, N. — Decoding Linear Error-Correcting Codes, Group Report, 55-15, M. I. T. Lincoln Lab., Oct. 1959

11. Birkoff, MacLane — A Survey of Modern Algebra, The Macmillan Co., New York, 1935

12. Carmichael, R. D. — Introduction to the Theory of Groups of Finite Order, Copyright, 1937, Dover edition, 1956

13. Burnside, W. — Theory of Groups of Finite Order, second edition, 1911, republication by Dover, 1955

References
(cont' d)

14. Paige, J. L. — A Note on Mathieu Groups, Canadian Journal of Mathematics, 9: 15-18, 1957

15. Dickson, L. E. — Linear Groups, with an Exposition of the Galois Field Theory, (1901), Dover edition 1958

16. McCluskey, E. J., Jr. — Minimization of Boolean Functions, B. S. T. J. 35, 1417(1956)

17. Bartee, T. C. — The Automatic Design of Logical Networks, M. I. T. Lincoln Lab. Tech. Report No. 191, Dec. 1958

18. Slepian, D. — On the Number of Symmetry Types of Boolean Functions of n Variables, Canadian Journal of Mathematics, Vol. 5, No. 2, 1954

19. Steinberg, R. — The Representation of GL(3, q), GL(4, q), PGL(3, q) and PGL(4, q), Canadian Journal of Mathematics, Vol. III, 1951, pp. 225-235.