THE UNIVERSITY OF MICHIGAN

THE GENERAL PROPERTIES OF FINITE WEIGHTED NUMBER SYSTEMS

Thammavarapu R. N. Rao

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the
University of Michigan
Department of Electrical Engineering
1963

IP-654

Doctoral Committee:

Professor Harvey L. Garner, Chairman
Professor William M. Brown
Associate Professor Bernard A. Galler
Assistant Professor John H. Holland
Professor Norman R. Scott

ACKNOWLEDGEMENTS

I wish to acknowledge my indebtedness and gratitude to Professor Harvey L. Garner, whose patient guidance and interest resulted in many improvements in a number of areas and made this work possible.

I thank all the members of the doctoral committee for the discussions, and the useful criticisms and corrections of the manuscript.

I also wish to thank Dr. R. F. Arnold for the many stimulating discussions in the area of modules and the structure of non-redundant weighted systems.

The stimulating environment of the Information Systems laboratory and the discussions with its members R. Gonzalez, P. Dauber, T. F. Piatkowski have made an important contribution to this effort.

Last but not the least, I thank my wife Rajyalaxmi, for her patience and understanding during the course of this work.

This dissertation was supported by the United States Air Force under Contract AF-33-657-7811.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT'D)

LIST OF TABLES

NOMENCLATURE

Z	ring of integers
Z_M	integers modulo M
$\lvert x \rvert_M$	least non-negative residue of x modulo M
$\lvert x \rvert$	absolute value of x
$a \in \xi$	a is an element of ξ
$a \notin \xi$	a is not an element of ξ
$\ni$	such that
$C \subseteq K$	C is a subset of K
ξ/C	If C is a subgroup of ξ, ξ/C is a quotient group
$A \cong B$	A and B are isomorphic
$\forall x \in \xi$	for any x in ξ
$\exists x \in \xi$	there exists an x in ξ
$K = \{x \in \xi \mid \phi(x) = 0\}$	K is the totality of elements in ξ, which are mapped by ϕ to 0
$\phi: \xi \rightarrow Z_M$	ϕ is a mapping of ξ into or onto Z_M
$\langle m_1, m_2, \ldots, m_n \rangle$	least common multiple of the integers $m_1, m_2, \ldots, m_n$
$(m_1, m_2, \ldots, m_n)$	greatest common divisor of the integers $m_1, m_2, \ldots, m_n$
$x \mid y$	x divides y
$\Longleftrightarrow$	if and only if
$x \nmid y$	x does not divide y
$\left[\frac{x}{y}\right]$	the greatest integer less than or equal to $\frac{x}{y}$
$\begin{bmatrix} c_{11} & \cdot & \cdot & c_{1n} \\ \cdot & & & \\ \cdot & & & \\ c_{n1} & \cdot & \cdot & c_{nn} \end{bmatrix}$	matrix (c_{ij})
$\begin{vmatrix} c_{11} & \cdot & \cdot & c_{1n} \\ \cdot & & & \\ \cdot & & & \\ c_{n1} & \cdot & \cdot & c_{nn} \end{vmatrix}$	determinent C

I. INTRODUCTION AND BACKGROUND

1.1 Consistently Based and Mixed Based Number Systems

A finite number system is a set of n-tuples of integers. These elements of the n-tuples are called the digits and they correspond to the n-moduli of the system. The term modulus as used in this dissertation refers to the cardinality of the digit set. The cardinality of the number system with moduli $m_1, m_2, \ldots, m_n$ is equal to $\prod_{i=1}^{n} m_i$. If there is a mapping of this system onto the integers $0, 1, 2, \ldots, M-1$ (denoted hereafter as Z_M), then the system is said to represent Z_M, and M is said to be the range of the system. Clearly the range M must be less than or equal to $\prod_{i=1}^{n} m_i$, in order that all integers in Z_M may have a representation in the system. Such systems are said to be complete.

For a consistently based system, since $m_1 = m_2 = \ldots = m_n = r$, the range M is equal to r^n. If the digit weights $\rho_1, \rho_2, \ldots, \rho_n$ are such that $\rho_i = r^{i-1}$, then the system can represent Z_M for $M = r^n$ non-redundantly. This is not the only possible set of digit weights, but it is the set normally encountered in practice.

In contrast, a residue number system must have moduli $m_1, m_2, \ldots, m_n$ that are pairwise relatively prime in order that the system can have a range $M = \prod_{i=1}^{n} m_i$. The residue number systems are classified as weighted, since weights $\rho_1, \rho_2, \ldots, \rho_n$ can be attached to the corresponding moduli in such a way that $(x_1, x_2, \ldots, x_n)$ represents an integer $x \epsilon Z_M$ if and only if

$$x = \left| \sum_{i=1}^{n} x_i \rho_i \right|_M \qquad (1.1)$$

Since $1 \epsilon Z_M$ must have a representation of the form $(c_1, c_2, \ldots, c_n)$, $c_i \epsilon Z_{m_i}$ in the system, $\left|\Sigma c_i \rho_i\right|_M = 1$, which implies $(\rho_1, \rho_2, \ldots, \rho_n, M) = 1$. This is a necessary condition for all weighted systems. If $1 \epsilon Z_M$ has a representation $(1, 1, \ldots, 1)$, as in the system of residue classes $\left|\sum_{i=1}^{n} \rho_i\right|_M = 1$, then any $x \epsilon Z_M$ is represented by

$$(x_1, x_2, \ldots, x_n), \tag{1.2}$$

where $x_i = |x|_{m_i}$.

1.2 ρ-Matrix or Weight Matrix

Rozenberg[(3)], in his work on the "Algebraic Properties of Residue Number Systems," has shown that the residue system is a pseudo-vector space or an R-space, since the system obeys the axioms of a vector space except for the uniqueness of representation with respect to the generator elements. Also the scalars here are integers instead of field elements as required for a vector space. For a weighted system with moduli $m_1, m_2, \ldots, m_n$ which are pairwise relatively prime, and $\rho_1, \rho_2, \ldots, \rho_n$ the corresponding weights, the nxn array or matrix of the form

$$\begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{22} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & & & \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{bmatrix}$$

where $\rho_{ij} = |\rho_i|_{m_j}$, is called the ρ-matrix or weight matrix. It is shown that the row elements of the ρ-matrix are the generators of the weighted system and also span the R-space. In a non-redundant system of residue classes, (or in the residue system in which (1, 1, ... 1) represents 1), we have

$$\left|\rho_i\right|_{m_j} = 1 \;, \quad i = j$$
$$= 0 \;, \quad i \neq j \;,$$

and the ρ-matrix is an Identity Matrix. For a general residue system, the ρ-matrix is diagonal with the diagonal element satisfying $(\rho_{ii}, m_i) = 1$.

It is shown in Rozenberg's paper that a sufficient condition for a non-redundant system is that the ρ-matrix be triangular (or can be made triangular by row and column permutation) and the diagonal elements satisfy $(\rho_{ii}, m_i) = 1$, for $i = 1, 2, \ldots, n$. However, the necessary condition was not established. Such systems with triangular forms are residue related systems. An important one in that category is a mixed base system with conventional carry propogation.

1.3 Finite, Non-redundant Number System Weights

For a general non-redundant weighted system using any set of n moduli, the necessary and sufficient conditions are given by Garner's theorem[6]. These conditions are

$$\left.\begin{aligned} (\rho_1, M) &= \frac{M}{m_1} \;, \\ (\rho_2, \frac{M}{m_1}) &= \frac{M}{m_2 m_1} \;, \\ &\;\vdots \\ \text{and} \quad (\rho_n, m_n) &= 1 \;, \end{aligned}\right\} \qquad (1.3)$$

for some ordering of the moduli $m_1, m_2, \ldots, m_n$.

Using these conditions for a special case of pairwise relatively prime moduli, it is shown that the ρ-matrix is triangular, and $(\rho_{ii}, m_i) = 1$. Unfortunately, the ρ-matrix cannot be constructed similarly when the moduli are not relatively prime.

1.4 Relation Between Digit Weights and Triangular Forms

For a system with any set of n moduli, the structure could be expressed by the relations between the digit weights $\rho_1, \rho_2, \ldots, \rho_n$ and the range M . There exist n independent relations between the variables $\rho_1, \rho_2, \ldots, \rho_n$ and M . This set of relations gives rise to the rules for carry generation. As an example, consider a consistently based system, having all $m_i = r$ and weights $\rho_1, \rho_2, \ldots, \rho_n$ where $\rho_i = r^{i-1}$. This system has the n independent relations as below.

$$\begin{bmatrix} r & 0 & 0 & \cdots & 0 \\ -1 & r & 0 & \cdots & 0 \\ 0 & -1 & r & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots -1 & r \end{bmatrix} \begin{bmatrix} \rho_n \\ \rho_{n-1} \\ \vdots \\ \rho_1 \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \pmod{M = r^n}$$

The (nxn) matrix above gives the carry propagation rules, and is called the carry matrix. For a residue system with moduli $m_1, m_2, \ldots, m_n$, the relations are given by

$$\begin{bmatrix} m_1 & 0 & 0 & \cdots & 0 \\ 0 & m_2 & 0 & \cdots & 0 \\ 0 & 0 & m_3 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & m_n \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_n \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \pmod{M}.$$

Since there are no carries between digits, the (nxn) carry matrix above is appropriately diagonal. The triangular form of carry matrix is the crucial property of non-redundant systems. Triangularity means that the carries are propagated in one direction, and an ordering on the moduli is possible. Also, the termination of carries is guaranteed once they are propagated up to the highest ordered digit. In contrast, the carry matrices of redundant systems need not be triangular and thus a more elaborate arithmetic process is expected. Another complication associated with redundant number systems is the tranformation of equivalent representations to the preferred representation. This is the canonical reduction problem.

II. FINITE NUMBER SYSTEMS: LINEAR AND NON-LINEAR CATEGORIES

2.1 A Finite Number System

Here we define the concept of a number system which heretofore has been left to the intuition of the reader. A number system could well be considered as a method of representing or assigning names to the integers.

Let a system N, be a cartesian product of n digit sets.

$$N = D_1 \times D_2 \times \ldots \times D_n$$

where

$$D_i = \{0, 1, \ldots, m_i - 1\}$$

and is called the i-th digit set and m_i the i-th modulus. Then N will be called a number system if

1) N is closed under addition:

   $x, y \in N \Longleftrightarrow x + y \in N$ (closure law)

2) $\exists\ 0 \in N$ such that $x + 0 = 0 + x = x$ (identity).

It will be proved later that, in addition to the above two, the other axioms of an abelian group are obeyed by non-redundant linear number systems.

3) $\exists$ a mapping $w : N \to Z_M$ and w is a function of the n variables, such that $w(x+y) = w(x) + w(y) \pmod M$ for all $x, y \in N$.

This function, which we will call hereafter the weight function, is the most significant property of the number system and is the major factor determining the arithmetic and carry properties of N. We will observe further that the division of number systems into different categories is based on this function. The following definitions,* which are quite familiar, are included as a basis for further discussion on number systems.

Definition 1. A number system N (obeying the three axioms stated above) is complete $\Longleftrightarrow$ w is onto Z_M . This is to say that for all $a \epsilon Z_M$, $\exists\, x \;\epsilon N$ such that $w(x) = a$.

Also $M > \prod^{n} m_i \Longleftrightarrow$ N is incomplete.

Definition 2. N is a redundant system $\Longleftrightarrow x,y \epsilon N$ such that $x \neq y$ and $w(x) = w(y)$.

The above two definitions can be combined to obtain the lemma:

Lemma 1. A number system is complete and non-redundant $\Longleftrightarrow$ w is an isomorphism.

This lemma permits the seperation of number systems into redundant and non-redundant tyres.

2.2 Fundamental Definition of Linearity

Definition 3. N is said to be a linear system if and only if w is a linear function of n variables, the coefficients coming from Z_M .

Definition 4. A number system N is weighted $\Longleftrightarrow \exists\, \rho_i \epsilon Z_M$ for $i = 1, 2, \ldots, n$ such that for any $x = (x_1, \ldots, x_n) \epsilon N$ $w(x) = \left| \sum_{i=1}^{n} \rho_i x_i \right|_M$. The weight function for weighted systems is a linear homogeneous function.

Thus all the weighted systems are linear. However, not all linear systems are weighted. The two examples is Section 2.3 illustrate the above statement.

---

* Definitions 1, 2, and 4 are made by Garner in his earlier work.(6)

2.3 Non-weighted Codes

Before we go into the advantages of weighted and non-weighted systems, we shall examine the weight functions w of some non-weighted codes. Given below is a table of representation of the code known as excess three representing $Z_{10}(N_1 \to Z_{10})$, and the four-bit reflected binary code for $Z_{16}(N_2 \to Z_{16})$. It is well known that the excess three code has the advantage over the binary coded decimal system (which is linear homogeneous) in that the 9's complement is obtained by interchanging 0's and 1's. The advantage of the reflected binary code is that it is a unit distance code. Thus any single digit error causes a change of one in magnitude. We shall show that the weight function of the excess three code is linear and non-homogeneous, and that of the reflected binary code is non-linear.

Let N_1 and N_2 be two non-redundant number systems representing Z_{10} and Z_{16} respectively.

Let the weight functions $N_1 \to Z_{10}$ and $N_2 \to Z_{16}$ be defined as shown in Table I.

Both of these mappings are (1-1) and onto, and hence for all $X, Y \in N$, we have

$$X + Y = w^{-1}\,[w(Y) + w(X)] .$$

This shows closure under addition. The existence of an identity element is obvious.

We can easily find that the weight function for the excess three code is such that for

TABLE I

REPRESENTATIONS FOR CODES N_1 AND N_2

Excess Three Code N_1

Z_{10}	x_4	x_3	x_2	x_1
0	0	0	1	1
1	0	1	0	0
2	0	1	0	1
3	0	1	1	0
4	0	1	1	1
5	1	0	0	0
6	1	0	0	1
7	1	0	1	0
8	1	0	1	1
9	1	1	0	0

Four-bit Reflected Binary Code N_2

Z_{16}	x_4	x_3	x_2	x_1
0	0	0	0	0
1	0	0	0	1
2	0	0	1	1
3	0	0	1	0
4	0	1	1	0
5	0	1	1	1
6	0	1	0	1
7	0	1	0	0
8	1	1	0	0
9	1	1	0	1
10	1	1	1	1
11	1	1	1	0
12	1	0	1	0
13	1	0	1	1
14	1	0	0	1
15	1	0	0	0

$$X \;\; = (X_1, X_2, X_3, X_4)$$

$$w(X) = 2^3 X_1 + 2^2 X_2 + 2X_3 + X_4 - 3 \; ,$$

where the coefficients $2^3, 2^2, 2, 1, M - 3$ are in Z_M . Thus w is a non-homogeneous linear function on the variables $X_4, \ldots, X_1$.

However, for the reflected binary code, the function w is not so straightforward to obtain. From Table I we have that

$$\begin{aligned} w(0, 0, 0, 0) &= 0, \\ w(0, 0, 0, 1) &= 1 = \rho_1, \\ w(0, 0, 1, 0) &= 3 = \rho_2, \\ w(0, 1, 0, 0) &= 7 = \rho_3, \\ w(1, 0, 0, 0) &= 15 = \rho_4. \end{aligned}$$

For an n-bit code

$$w(0, 0, \ldots, 0, \underset{\substack{\uparrow \\ i^{th} \text{ place.}}}{1}, 0, \ldots 0) = 2^i - 1 = \rho_i$$

Next we denote the weight function for the n-bit reflected binary code as f_n . Taking note of the alternate negation in the weights of the digits in the positions where 1's are present, we can write as below:

In the case of a 1-bit code

$$f_1 = w(X_1) = \rho_1 X_1 = X_1 \; ,$$

and for a 2-bit code

$$\begin{aligned} f_2 = w(X_2, X_1) &= X_2\rho_2 + X_1\rho_1 - 2X_2X_1\rho_1 \\ &= X_2\rho_2 + (1 - 2X_2)f_1 \; . \end{aligned}$$

For a 3-bit code

$$\begin{aligned} f_3 &= w(X_3, X_2, X_1) \\ &= X_3\rho_3 + X_2\rho_2 + X_1\rho_1 - 2X_3X_2\rho_2 - 2X_2X_1\rho_1 - 2X_3X_1\rho_1 + 4X_3X_2X_1\rho_1 \\ &= X_3\rho_3 + (1 - 2X_3)(X_2\rho_2 + (1 - 2X_2)X_1\rho_1) \\ &= X_3\rho_3 + (1 - 2X_3)f_2 \ . \end{aligned}$$

Then

$$\begin{aligned} f_4 &= w(X_4, \ldots, X_1) \\ &= X_4\rho_4 + (1 - 2X_4)f_3 \end{aligned}$$

and

$$\begin{aligned} f_n &= w(X_n, \ldots, X_1) \\ &= X_n\rho_n + (1 - 2X_n)f_{n-1} \\ &= X_n\rho_n + (1 - 2X_n)X_{n-1}\rho_{n-1} + (1 - 2X_n)X_{n-2}\rho_{n-2} \\ &\qquad + \ldots\ldots\ldots \\ &\qquad + (1 - 2X_n)(1 - 2X_{n-1}) \ldots (1 - 2X_2)X_1\rho_1 \ , \end{aligned} \tag{2.1}$$

which is clearly a non-linear function of order n .

2.4 Digitwise Sum

In a non-redundant system (w is a (1-1) mapping), the addition in N is defined by w . This is because

$$\left.\begin{aligned} w(X + Y) &= |w(X) + w(Y)|_M \\ (X + Y) &= w^{-1} [|w(X) + w(Y)|_M] \end{aligned}\right\} \text{ for all } X, Y \in N$$

and w^{-1} is (1-1) and, Z_M is an additive abelian group, so is N .

An important property of all linear homogenous systems is that for all $X, Y \in N$

$$w(X+Y) = w(x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \tag{2.2}$$

This is trivial if $x_i + y_i < m_i$ for all $i = 1, \ldots, n$. In which case $X + Y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \in N$. But when for any i, $x_i + y_i \geq m_i$, then $(x_1 + y_1, \ldots, x_n + y_n) \notin N$. However, (2.2) still holds.

$$w(X) = |x_1\rho_1 + x_2\rho_2 + \ldots + x_n\rho_n|_M$$

$$w(Y) = |y_1\rho_1 + y_2\rho_2 + \ldots + y_n\rho_n|_M$$

$$w(X+Y) = w(X) + w(Y) = |(x_1 + y_1)\, \rho_1 + \ldots + (x_n + y_n)\rho_n|_M$$

$$= w(x_1 + y_1, \ldots, x_i + y_i, \ldots, x_n + y_n) .$$

Conversely let

$$w(x_1, x_2, \ldots, x_n) + w(y_1, y_2, \ldots, y_n)$$

$$= w(x_1 + y_1, \ldots, x_n + y_n).$$

Then replacing $y_i = 0$ for $i = 1, 2, \ldots, n$, we get

$$w(x_1, x_2, \ldots, x_n) + w(0, 0, \ldots, 0) = w(x_1 + 0, \ldots, x_n + 0)$$

$$= w(x_1, x_2, \ldots, x_n) .$$

Therefore,

$$w(0, 0, \ldots, 0) = 0 .$$

Also for any integer $a \in Z_M$

$$w(ax_1, ax_2, \ldots, ax_n) = \underbrace{w(x_1, x_2, \ldots, x_n) + \ldots + w(x_1, x_2, \ldots, x_n)}_{a \text{ times}}$$

$$= aw(x_1, x_2, \ldots, x_n) .$$

Let

$$w(0, 0, \ldots, x_i, \ldots, 0) = f_i(x_i) ,$$

then f_i is a homogeneous function on x_i .

Since $\quad w(0, \ldots, x_i + y_i, \ldots, 0) = f_i(x_i) + f_i(y_i)$,

f_i is a linear homogeneous function on x_i .

$$w(x_1, x_2, \ldots, x_n) = w(x_1, 0, \ldots, 0) + w(0, x_2, x_3, \ldots, x_n)$$
$$= w(x_1, 0, \ldots, 0) + w(0, x_2, 0, \ldots, 0) + w(0, 0, x_3, \ldots, x_n)$$
$$\vdots$$
$$= w(x_1, 0, \ldots, 0) + w(0, x_2, 0, \ldots, 0) + \ldots + x(0, \ldots, x_n)$$
$$= \sum_{i=1}^{n} f_i(x_i) ,$$

Where f_i is a linear homogeneous function of x_i , for

$$i = 1, 2, \ldots, n .$$

Therefore w is a linear homogeneous function on n variables

$$x_1, x_2, \ldots, x_n.$$

Thus we have proved the following:

Theorem 1.

N is a linear homogeneous system with a mapping $w: N \to Z_M$ if and only if w satisfies the digitwise sum rule given by (2.2).

In the excess three code N_1 , which is non-homogeneous

$$X = (x_4, x_3, x_2, x_1)$$
$$Y = (y_4, y_3, y_2, y_1)$$
$$w(X) = 8x_4 + 4x_3 + 2x_2 + x_1 - 3$$
$$w(Y) = 8y_4 + 4y_3 + 2y_2 + y_1 - 3$$
$$w(X+Y) = w(X) + w(Y) = 8(x_4+y_4) + 4(x_3+y_3) + 2(x_2+y_2) + (x_1+y_1) - 3 - 3$$
$$= w(x_4 + y_4, x_3 + y_3, x_2 + y_2, x_1 + y_1) - 3.$$

Thus

$$w(X+Y) \neq w(x_4 + y_4, x_3 + y_3, x_2 + y_2, x_1 + y_1) .$$

In the reflected two-bit binary code $(N_2 \to Z_4)$

$$X = (x_2, x_1)$$
$$Y = (y_2, y_1)$$
$$w(X) = 3x_2 + x_1 - 2x_2x_1$$
$$w(Y) = 3y_2 + y_1 - 2y_2y_1$$
$$w(X+Y) = w(X) + w(Y) = 3(x_2 + y_2) + (x_1+y_1) - 2(x_2x_1+y_2y_1)$$
$$\neq w(x_2 + y_2, x_1 + y_1) .$$

In linear homogeneous systems, we proved that digitwise addition can be carried out. If any digit sum is $\geq$ the corresponding modulus, then the result is not in the number system. Conventionally this is taken care of by carry generation and assimilation. For a general weighted number system, the carry generation and assimilation process is characterized later by using the theory of modules.

III. THEORY OF MODULES OVER INTEGERS

In this chapter the concepts of free module, submodule and quotient module are presented. The theorems of module theory relevant and necessary for our study of weighted number systems, are stated here. In Section 3.2, similarities are examined between the structures of a non-redundant weighted system N , and a quotient module ξ/S over integers. This leads to the concept that the quotient module ξ/S , where the submodule S of ξ is constructed from the carry generation rules of N , stands as an abstract model for N .

3.1 Algebraic Preliminaries

Let Z be the set of all integers. Algebraically, Z satisfies the axioms of a ring, integral domain and also an Euclidian domain.

Let ξ be n tuples of integers of the form $(x_1, x_2, \ldots, x_n)$, $x_i \epsilon Z$. Let addition in ξ be defined as follows:

$$x = (x_1, x_2, \ldots, x_n)$$

$$y = (y_1, y_2, \ldots, y_n)$$

$$x+y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) .$$

Let scalar multiplication of $x \epsilon \xi$ by any integer $a \epsilon Z$ be defined as

$$a(x_1, x_2, \ldots, x_n) = (ax_1, ax_2, \ldots, ax_n) .$$

If ξ satisfies the axioms of an abelian group (the axioms A1 to A5) with respect to addition, and the mapping of $Z \times \xi \rightarrow \xi$ called scalar multiplication obeys the axioms M1 to M4 and G1 , then ξ is called a module over Z , or a Z-module.

$$\left.\begin{array}{ll} \text{(A1)} & x + y \in \xi \\ \text{(A2)} & x + (y+w) = (x+y) + w \\ \text{(A3)} & x + y = y + x \\ \text{(A4)} & \exists\ 0 \in \xi \text{ such that } x + 0 = x \\ \text{(A5)} & \exists\ x' \in \xi \text{ such that } x + x' = 0 \end{array}\right\} \text{ for all } x, y, w \in \xi$$

$$\left.\begin{array}{ll} \text{(M1)} & a(bx) = ab(x) \\ \text{(M2)} & a(x + y) = ax + ay \\ \text{(M3)} & (a + b)x = ax + bx \\ \text{(M4)} & 1x = x \end{array}\right\} \text{ for all } x, y \in \xi \text{ and } a, b \in Z$$

(G1) There exists a set $\{e_1, e_2, \ldots, e_n\}$; $e_i \in \xi$ such that for any $w \in \xi$, there exist integers $w_1, w_2, \ldots, w_n$ such that $w = \sum_{i=1}^{n} w_i e_i$. The set $\{e_1, e_2, \ldots, e_n\}$ is called the generator set.

In connection with the above axioms A1 to A5 , M1 to M4 , and G1 , it must be pointed out that (1) a module need not necessarily be defined over integers. A general definition of a module can be over any ring with an identity. And (2) a module is a generalization of a vector space in that (i) a vector space is defined over a division ring and more often over a field, and (ii) a module may not have a basis. If a module has a basis, then it is called a free module. For a module to have a basis, the axiom G1 must be replaced by a strengthened form, G2 .

(G2) There exist a set $\{e_1, e_2, \ldots, e_n\}$ $e_i \in \xi$ such that for any $w \in \xi$ can be written in one and only one way in the form $w_1 e_1 + w_2 e_2 + \ldots + w_n e_n$ for some $w_i \in Z$.

It is now possible to observe that ξ as defined above is a free module having a basis $\{e_1, e_2, \ldots, e_n\}$, where $e_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ (1 in the i-th place) for all $i = 1, 2, \ldots, n$.

Definition 5. If C is a subgroup of ξ and for all $s \in C$, $a \in Z \Longleftrightarrow as \in C$, then C is said to be a submodule.

From the above definition, the totality $\{x\}$ of multiples ax of the fixed element x in ξ and for all $a \in Z$ is a submodule generated by x. Also, $C = \{c_1, c_2, \ldots, c_k\}$ is the submodule generated by the set $c_1, c_2, \ldots, c_k, c_i \in \xi$. For all $x \in C \Longrightarrow x = \sum a_i c_i$ for some $a_i \in Z$. In this connection, a well-known(4) theorem stated and proved in standard textbooks of modern algebra will be stated as follows.

Theorem 2.

If ξ is a free Z-module with a basis of n elements, then any submodule C of ξ is also free and has a basis of $m \leq n$ elements.

Definition 6. ξ is a free Z-module and if C is a submodule, then the quotient group ξ/C satisfies the axioms of a module over Z. Thus, ξ/C is a Z-module, called the quotient module or difference module.

The following notation will be used hereafter for the modules ξ, C and ξ/C:

(1) $$\xi = \underbrace{Z \times Z \times \cdots \times Z}_{n \text{ terms}}$$

and the basis $\{e_1, e_2, \ldots, e_n\}$ $e_i = (0, \ldots, 1, \ldots, 0)$ (1 in the i-th place).

(2) A submodule C with k generators $c_1, c_2, \ldots, c_k, c_i \in \xi$ is written as $C = \{c_1, c_2, \ldots, c_k\}$ where

$$c_i = \sum_{j=1}^{n} c_{ij} e_j \qquad i = 1, 2, \ldots, k \tag{3.1}$$

and written as $(c_{i1}, c_{i2}, \ldots, c_{in})$ with respect to the basis $(e_1, e_2, \ldots, e_n)$. Hence, the submodule C can be written as a $(k \times n)$ matrix

$$C = \{c_1, c_2, \ldots, c_k\} = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ & & \\ c_{k1} & \cdots & c_{kn} \end{bmatrix} \tag{3.2}$$

(3) ξ/C is denoted as

$$Z + Z + \ldots + Z \Big/ \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ & & \\ c_{k1} & \cdots & c_{kn} \end{bmatrix}$$

Any $(k \times n)$ matrix over integers represents a submodule generated by the k rows of that matrix, and each of the rows are an element of the module ξ' with respect to the basis $\{e_1, e_2, \ldots, e_n\}$. If $\{f_1, f_2, \ldots, f_n\}$ is another basis in ξ such that

$$e_i = \sum_{j=1}^{n} \mu_{ij} f_j \qquad \text{for } i=1,2,\ldots,n \tag{3.3}$$

the right multiplication of C by the $(n \times n)$ matrix μ gives C' with respect to the new basis $(f_1, f_2, \ldots, f_n)$ of ξ.

By considering the change of basis from $(f_1, f_2, \ldots, f_n)$ to $(e_1, e_2, \ldots, e_n)$, another (nxn) matrix μ' can be obtained such that $\mu\mu'$ = identity matrix. Thus, μ and μ' are both invertible and so must have determinant $= \pm 1$. Similarly, it can be shown that a change of basis of the submodule can be effected by a left multiplication of C by a (kxk) invertible matrix. We still have the same submodule but the basis for representation of the submodule is different. Also, the row representation is altered by a new basis of ξ . Hence we may state the following.

Theorem 3.

If C is a (kxn) matrix over integers representing a submodule of ξ , then $C' = uCv$ also represents the same submodule with respect to a different basis in ξ where u is a (kxk) and v an (nxn) invertible matrices over Z .

By definition we shall use the term equivalence for (kxn) matrices if there exists u , v as defined above such that $C' = uCv$. If C is an (nxn) matrix, then the absolute values of determinants C and C' are equal, since the invertible matrices have determinants equal to ± 1 . Also we need the following important theorem.

Theorem 4.

If C is a (kxn) matrix with elements in Z , there exists a matrix C' equivalent to C which has the diagonal form as below.

$$C' = \begin{bmatrix} a_1 & 0 & . & & & 0 \\ 0 & a_2 & . & & & 0 \\ & & & & & \\ 0 & 0 & . & a_r & 0 & 0 \\ 0 & 0 & . & 0 & 0 & 0 \\ & & & & & \\ 0 & 0 & . & & & 0 \end{bmatrix}$$

$$r \leq k \leq n$$

$a_i \neq 0$ and a_i divides a_j for $i < j$ and r is the row rank of submodule. (This is a theorem in Jacobson, Vol. 2, Chapter 3, p.79.)[4]

The row rank of a matrix coincides with the number of basis elements of the submodule represented by that matrix. If $r = k = n$, then $|\det C| = |\det C'| = \left|\prod_{i=1}^{n} a_i\right|$.

Let ξ be a Z-module with a basis $\{e_1, e_2, \ldots, e_n\}$, and C be a submodule with a basis of n elements $c_1, c_2, \ldots, c_n$ where

$$c_i = (c_{i1}, c_{i2}, \ldots, c_{in})$$

with respect to $(e_1, e_2, \ldots, e_n)$. That is

$$c_i = \sum_{j=1}^{n} c_{ij} e_j \qquad i = 1, 2, \ldots, n .$$

Then C is an $(n \times n)$ matrix. The diagonal equivalent matrix C' of the Theorem 4 will be of the form

$$C' = \begin{bmatrix} a_1 & 0 & . & . & 0 \\ 0 & a_2 & . & . & 0 \\ . & & & & \\ . & & & & \\ 0 & 0 & . & . & a_n \end{bmatrix}$$

Then ξ/C' in this form can be recognized as a group with cardinality equal to $|a_1 \; a_2 \; \ldots \; a_n|$. Since C is the same submodule as C' , except that the representation is with respect to a new basis, the cardinality of ξ/C is also equal to $|a_1 \; a_2 \; \ldots \; a_n| = |\det C|$. Hence we proved the following theorem.

Theorem 5.

Let ξ be a Z-module of basis $\{e_1, e_2, \ldots, e_n\}$ and C be a submodule with a basis $\{c_1, c_2, \ldots, c_n\}$ where

$$c_i = \sum_{j=1}^{n} c_{ij} e_j \qquad \text{for } i = 1, 2, \ldots, n$$

such that C can be represented as an (nxn) matrix. Then the cardinality of the ξ/C module is equal to the absolute value of the determinant C .

3.2 The Number System as a Quotient Module

Let N be a non-redundant weighted system with moduli $m_1, m_2, \ldots, m_n$ and the corresponding digit weights $\rho_1, \rho_2, \ldots, \rho_n$. Then

$$N = D_1 \times D_2 \times \ldots \times D_n$$

where

$$D_i = \{0, 1, 2, \ldots, m_{i-1}\}$$

and the cardinality of $N = \prod_{i=1}^{n} m_i$. Also $(x_1, x_2, \ldots, x_n)$ represents $|\sum_{i=1}^{n} x_i \rho_i|_M$ in Z_M . Let the digit weights be related by the following equations or congruences.

$$\left.\begin{aligned} m_1\rho_1 &\equiv \quad c_{12}\rho_2 + c_{13}\rho_3 + \cdots + c_{1n}\rho_n \\ m_2\rho_2 &\equiv c_{21}\rho_1 + \quad + c_{23}\rho_3 + \cdots + c_{2n}\rho_n \\ &\;\vdots \\ m_n\rho_n &\equiv c_{n1}\rho_1 + c_{n2}\rho_2 + c_{n3}\rho_3 + \cdots + c_{nn-1}\rho_{n-1} + \end{aligned}\right\} \pmod{M} \qquad (3.4)$$

The form of the above equations is justifiable from the following. Since the system N is closed under addition, the sum of any two elements of N , having a digitwise sum equal to $(0, 0, \ldots, m_i, \ldots, 0)$, must have an equivalent canonical form in N . Let this be $(c_{i1}, c_{i2}, \ldots, c_{in})$. Then $0 \leq c_{ij} < m_j$. Also if $c_{ii} \neq 0$, then $(0, 0, \ldots, m_i - c_{ii}, \ldots, 0)$ and $(c_{i1}, c_{i2}, \ldots, 0, \ldots, c_{in})$ are equal. Since N is non-redundant, this is contradictory. Therefore, $c_{ii} = 0$. Hence we can rewrite (3.4) in the form below.

$$\begin{bmatrix} m_1 & -c_{12} & \cdot & \cdot & -c_{1n} \\ -c_{21} & m_2 & \cdot & \cdot & -c_{2n} \\ \cdot & & & & \\ \cdot & & & & \\ -c_{n1} & -c_{n2} & \cdot & \cdot & m_n \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \\ \\ \rho_n \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \pmod{M} \quad \text{for } 0 \leq c_{ij} < m_j \qquad (3.5)$$

The (nxn) matrix in the above form will be called the <u>carry propagation matrix</u>, or simply, the <u>carry matrix</u> of the system N . It will be shown below that the carry matrix is identical with the submodule S , such that it is possible to define a structure of a quotient module ξ/S for N as follows.

Let the i-th row of the carry matrix be s_i, i.e.

$$s_i = (-c_{i1}, -c_{i2}, \ldots, -c_{i,i-1}, m_i, -c_{i,i+1}, \ldots, -c_{in})$$

for

$$i = 1, 2, \ldots, n.$$

Also, let $s_1, s_2, \ldots, s_n$ be considered as elements of ξ with respect to the basis $\{e_1, e_2, \ldots, e_n\}$. That is,

$$s_i = \sum_{i=1}^{n} -c_{ij}e_j$$

where

$$c_{ii} = -m_i .$$

Let the free submodule with basis $\{s_1, s_2, \ldots, s_n\}$ be S. We can then give the following definition.

Definition 7.* Let N be a weighted system with moduli $m_1, m_2, \ldots, m$ and corresponding digit weights $\rho_1, \rho_2, \ldots, \rho_n$ with the relation on the digit weights yielding n independent equations that can be expressed in the form (3.4). Let C be the $(n\text{x}n)$ matrix of (3.5), and S be the submodule of ξ generated by the rows of C, then N is said to have the structure of ξ/S.

By the above definition, if N has a structure of ξ/S, then S specifies completely the carry propagation in N. This idea will be further studied in the next section.

---

* The quotient module structure for number systems was first conceived by R. F. Arnold,(12) who gave a similar structure to linear number systems. The main difference in his work is that ξ is a finite module over Z_M instead of Z.

IV. NON-REDUNDANT WEIGHTED SYSTEMS

We have discussed in the preceding chapter a quotient module ξ/S structure for a non-redundant system N. While the cardinality of N is the range M and is equal to Πm_i, the cardinality of ξ/S is equal to the absolute value of the determinant of S. Besides the structural similarity between the system N and its model ξ/S, we will in this chapter, establish the conditions on the system, for the existence of an isomorphism between the two.

Two theorems on determinants are derived in sec. 4.1 in order to establish a property of the carry matrix of the non-redundant weighted systems. This property is that the determinant of such a matrix is less than or equal to the product of the main diagonal elements, and the equality holds if and only if the matrix has a triangular form.

4.1 Some Useful Theorems on Determinants

Let C be a (kxk) matrix of the form shown below:

$$\begin{bmatrix} m_1 & c_{12} & \cdot & \cdot & c_{1n} \\ c_{21} & m_2 & \cdot & \cdot & c_{2n} \\ \cdot & & & & \\ c_{n1} & \cdot & \cdot & \cdot & m_n \end{bmatrix}$$

$m_1, m_2, \ldots, m_n$ are used for the principal diagonal, so that they are easily distinguished from the rest. A diagonal permutation on C is a column and row permutation as defined below.

<u>Definition 8</u>. If i-th and j-th rows are exchanged, followed by an i-th and j-th column exchange, then the matrix is said to be <u>diagonally permuted</u>. Such row and column permutations are said to be <u>diagonal permutations</u>.

Diagonal permutations satisfy the following:

(1) The set of elements $m_1, m_2, \ldots, m_n$ of the matrix C remains on the principal diagonal after a diagonal permutation and also after any number of repeated diagonal permutations.

(2) The determinent of C is unaltered in sign and magnitude by diagonal permutation.

(3) Every diagonal permutation has an inverse diagonal permutation.

<u>Definition 9</u>. If a (kxk) matrix C can be made (lower or upper) triangular by a necessary number of repeated diagonal permutations, then C is said to be <u>triangularable</u>.

<u>Lemma 2</u>. If a (kxk) matrix C is triangularable, then there exists a $j \leq k$ such that $c_{ji} = 0$ for $j \neq i$.

<u>Proof</u>: The lemma in essence means that there must exist a row in C in which off-diagonal elements are zero.

$$\text{Now let } C = \begin{bmatrix} m_1 & c_{12} & \cdot & \cdot & c_{1k} \\ c_{21} & m_2 & \cdot & \cdot & c_{2k} \\ \cdot & & & & \\ \cdot & & & & \\ c_{k1} & c_{k2} & \cdot & \cdot & m_k \end{bmatrix}$$

Let C' be the (lower) triangular matrix obtained by diagonal permutation on C.

$$\text{Then } C' = \begin{bmatrix} m_1' & 0 & \cdot & \cdot & 0 \\ c_{21}' & m_2' & 0 & \cdot & 0 \\ \cdot & & & & \\ \cdot & & & & \\ c_{k1}' & \cdot & \cdot & \cdot & m_k' \end{bmatrix}$$

C' can also be diagonally permuted to obtain C (as the diagonal permutations have inverses). The first row of C' has at most one non-zero element. Column permutations of C' do not change the number of non-zero elements of any row. A row permutation involving the first row (which has at most one non-zero element) and j-th row would leave the j-th row with one non-zero element. Hence, there will always be a row having m_1' on the diagonal with that row satisfying the required condition. Since C is obtained by repeated diagonal permutation of C', C satisfies that condition; hence, the lemma is proved.

<u>From here on, all the matrices and determinants are over integers.</u>

<u>Theorem 6.</u>

Let C be a determinant as shown below:

$$C = \begin{vmatrix} m_1 & -c_{12} & \cdot & \cdot & -c_{1n} \\ -c_{21} & m_2 & \cdot & \cdot & -c_{2n} \\ \cdot & & & & \\ \cdot & & & & \\ -c_{n1} & \cdot & \cdot & \cdot & m_n \end{vmatrix}$$

where c_{ij} is any non-negative integer, and $m_i > 0$ for $i = 2, 3, \ldots, n$. Then

$$C \leq \prod_{i=1}^{n} m_i \tag{4.1}$$

Proof: For $n = 1, 2$ the theorem is true. Assume the theorem is true for $n = 1, 2, \ldots, k-1$.

Claim. Theorem is true for $n = k$.

$$C = \begin{vmatrix} m_1 & -c_{12} & \cdot & \cdot & -c_{1k} \\ -c_{21} & m_2 & \cdot & \cdot & -c_{2k} \\ \cdot & & & & \\ \cdot & & & & \\ -c_{k1} & -c_{k2} & \cdot & \cdot & m_k \end{vmatrix}$$

$$C = m_1 \Delta_{11} - \sum_{i=2}^{k} c_{i1}\Delta_{i1}(-1)^{i-1}$$

$$= m_1 \Delta_{11} + \sum_{i=2}^{k} (-1)^{-i} c_{i1}\Delta_{i1}. \tag{4.2}$$

Δ_{11} is a (k-1 by k-1) determinant satisfying the conditions of the theorem so by the induction hypothesis

$$\Delta_{11} \leq \prod_{i=2}^{k} m_i$$

$$m_1 \Delta_{11} \leq \prod_{i=1}^{k} m_i$$

If it is shown that $(-1)^i c_{i1}\Delta_{i1}$ is ≤ 0 for $i = 2, 3, \ldots, k$, then determinant $C = m_1\Delta_{11} + (-1)^i c_{i-1}\Delta_{i1} \leq m_1\Delta_{11} \leq m_1 \; m_2 \; \cdots \; m_k$ and thus proof will be complete. Therefore consider $(-1)^i c_{i1}\Delta_{i1}$ where

$$\Delta_{i1} = \begin{vmatrix} -c_{12} & -c_{13} & \cdot & \cdot & -c_{1i} & \cdot & -c_{1k} \\ m_2 & -c_{23} & \cdot & \cdot & -c_{2i} & \cdot & -c_{2k} \\ -c_{32} & m_3 & \cdot & \cdot & -c_{3i} & \cdot & -c_{3k} \\ \cdot & & & & & & \\ -c_{i-1,2} & \cdot & \cdot & m_{i-1} & -c_{i-1,i} & \cdot & -c_{i-1,k} \\ -c_{i+1,2} & \cdot & \cdot & \cdot & -c_{i+1,i} & m_{i+1} & -c_{i+1,k} \\ \cdot & & & & & & \\ \cdot & & & & & & \\ -c_{k,2} & -c_{k3} & \cdot & \cdot & -c_{ki} & \cdot & m_k \end{vmatrix}$$

This minor does not have m terms on the diagonal. By shifting the i-th column to the place of the first column, a new minor Δ'_{i1} is obtained such that $\Delta_{i1} = (-1)^{i-2}\Delta'_{i1}$. Also, the Δ'_{i1} has all off-diagonal elements negative and all terms except the first on the principal diagonal ≥ 0, thus satisfying the induction hypothesis. Therefore, $\Delta_{i1} = (-1)^{i-2}\Delta'_{i1}$ where

$$\Delta'_{i1} = \begin{vmatrix} -c_{1i} & -c_{12} & \cdot & \cdot & \cdot & -c_{1k} \\ -c_{2i} & m_2 & \cdot & \cdot & \cdot & -c_{2k} \\ \cdot & & & & & \\ -c_{i-1,i} & \cdot & \cdot & m_{i-1} & \cdot & -c_{i-1,k} \\ -c_{i+1,i} & \cdot & \cdot & \cdot & m_{i+1} & -c_{i+1,k} \\ \cdot & & & & & \\ -c_{ki} & -c_{k2} & \cdot & \cdot & \cdot & m_k \end{vmatrix}$$

From the induction hypothesis we have

$$\Delta'_{i1} \leq -c_{1i} \; m_2 \cdots m_{i-1} \; m_{i+1} \cdots m_k$$

≤ 0 as c_{1i}, m_j are all non-negative for $j = 2, 3, \ldots, n.$

Therefore, the summation term

$$c_{i1} \, (-1)^i \, (-1)^{i-2} \, \Delta_{i1} = c_{i1} \, (-1)^{2i-2} \Delta'_{i1},$$

and so has the same sign as Δ_{i1}.

Therefore, $(-1)^i c_{i1} \Delta_{i1} \leq 0$ for $i = 2, \ldots, n.$ Therefore determinant

$$C \leq \prod_{j=1}^{k} m_j \; .$$

Hence, the theorem is proved.

Lemma 3. Let there be two determinants c_{k-1}, D_k such that

$$c_{k-1} = \begin{vmatrix} m_1 & -c_{12} & \cdot & \cdot & -c_{1,k-1} \\ -c_{21} & m_2 & \cdot & \cdot & -c_{2,k-1} \\ \cdot & & & & \\ \cdot & & & & \\ -c_{k-1,1} & \cdot & \cdot & \cdot & m_{k-1} \end{vmatrix}$$

and

$$D_k = \begin{vmatrix} m_1 & 0 & 0 & \cdot & 0 \\ -d_{21} & m_2 & -d_{23} & \cdot & -d_{2k} \\ -d_{31} & -d_{32} & m_3 & \cdot & -d_{3k} \\ \cdot & & & & \\ -d_{k,1} & \cdot & \cdot & \cdot & m_k \end{vmatrix}$$

all

$$c_{ij} \geq 0 \ , \quad d_{ij} \geq 0$$

and

$$m_i > 0 \qquad i = 2, \ldots, k$$

m_1 any non-zero integer.

Then if

$$c_{k-1} = \prod_{i=1}^{k-1} m_i \Longrightarrow c_{k-1} \qquad \text{is triangularable,}$$

then

$$D_k = \prod_{i=1}^{k} m_i \Longrightarrow D_k \qquad \text{is triangularable.}$$

<u>Proof</u>: $D_k = m_1 \Delta_{11}$,

$$\text{where } \Delta_{11} = \begin{vmatrix} m_2 & -d_{23} & \cdot & -d_{2k} \\ -d_{32} & m_3 & \cdot & -d_{3k} \\ \cdot & & & \\ \cdot & & & \\ -d_{k,1} & \cdot & \cdot & m_k \end{vmatrix}$$

$$D_k = m_1 \Delta_{11} = m_1 \ m_2 \ \ldots \ m_k$$

$$\Delta_{11} = m_2 \ \ldots \ m_k \ .$$

Δ_{11} is a k-1 by k-1 determinant whose determinant is equal to the product of the principal diagonal elements. Thus Δ_{11} is triangularable. Diagonal permutation of D_k , so that Δ_{11} is triangular, would leave the first row of D_k unaltered, (since the zeros are permuted). Hence, we will have a D_k in triangular form. The proof is then complete.

Theorem 7.

Let

$$D_n = \begin{vmatrix} m_1 & -c_{12} & \cdot & -c_{1n} \\ -c_{21} & m_2 & \cdot & -c_{2n} \\ \cdot & & & \\ -c_{n1} & -c_{n2} & \cdot & m_n \end{vmatrix}$$

such that all $c_{ik} \geq 0$, and $m_i > 0$ for $i = 2, \ldots, k$, m_1 is any non-zero integer. Then

$$D_n = \prod_{i=1}^{n} m_i$$

if and only if D_n is triangularable.

Proof: Let D_n be triangularable. Then let D_n' be the triangular form of D_n . Therefore we have $D_n' = D_n$. Since D_n' is triangular and the diagonal is only permuted, we have $D_n = D_n' = m_1 \; m_2 \; \ldots \; m_n$. Therefore,

$$D_n \text{ is triangularable} \Longrightarrow D_n = m_1 \; m_2 \; \ldots \; m_n \; .$$

Yet to be proved is

$$D_n = m_1 \; m_2 \; \ldots \; m_n \Longrightarrow D_n \text{ is triangularable.}$$

Proof by Induction:

Induction step: For $n = 1$, it is trivial.

For $n = 2$, $\begin{vmatrix} m_1 & -c_{12} \\ -c_{21} & m_2 \end{vmatrix} = m_1 \; m_2$

$$\Longrightarrow c_{21} \; c_{12} = 0$$

$$\Longrightarrow c_{12} \text{ or } c_{21} = 0 \; .$$

Hence, D_2 is triangular.

<u>Induction Hypothesis</u>: The theorem is true for $n = 1, 2, \ldots, k-1$.

<u>Claim</u>: The theorem is true for $n=k$.

$$D_k = \begin{vmatrix} m_1 & -c_{12} & \cdot & \cdot & -c_{1k} \\ -c_{21} & m_2 & \cdot & \cdot & -c_{2k} \\ \cdot & & & & \\ -c_{k1} & \cdot & \cdot & \cdot & m_k \end{vmatrix}$$

$$D_k = m_1\Delta_{11} + \sum_{i=2}^{k} (-1)^i c_{i1}\Delta_{i1} = m_1 \; m_2 \; \cdots \; m_k \; .$$

$$\Delta_{11} = \begin{vmatrix} m_2 & -c_{23} & \cdot & -c_{2k} \\ -c_{32} & m_3 & \cdot & -c_{3k} \\ \cdot & & & \\ m_{k2} & \cdot & \cdot & m_k \end{vmatrix}$$

From Theorem 6 we have $\Delta_{11} \leq m_2 \; m_3 \; \ldots \; m_k$ and as all the terms in this summation are negative

$$m_1\Delta_{11} + \sum_{i=2}^{n} (-1)^i c_{i1}\Delta_{i1} = m_1 \; m_2 \; \cdots \; m_k$$

$$\Longrightarrow \Delta_{i1} = 0 \text{ for } i = 2, \ldots, k \; .$$

$$\text{and} \quad \Delta_{11} = m_2 \; \cdots \; m_k \; .$$

From the induction hypothesis, Δ_{11} is triangularable. Hence, let D_k be diagonally permuted so that Δ_{11} is lower triangular. Now reordering the subscripts (as m_2, m_k are all arbitrary) we have D_k as shown in Table II.

TABLE II

(kxk) DETERMINENT D_k , WITH TRIANGULAR (k-1 x k-1) MINOR

$$D_k = \begin{vmatrix} m_1 & -c_{12} & -c_{13} & \cdot & & -c_{1k-1} & c_{1k} \\ -c_{21} & m_2 & 0 & 0 & & 0 & 0 \\ -c_{31} & -c_{32} & m_3 & 0 & & 0 & 0 \\ \cdot & & & & & & \\ \cdot & & & & & & \\ -c_{k-1,1} & \cdot & \cdot & \cdot & & m_{k-1} & 0 \\ -c_{k1} & \cdot & \cdot & \cdot & & \cdot & m_k \end{vmatrix}$$

If there exists a row that has all zero terms, except the diagonal element, then we can bring that row to the top by diagonal permutation. The resulting determinant then satisfies the hypothesis of lemma 3, and so D_k is triangularable, and the theorem will be true. Therefore, we can assume that there does not exist in D_k a row having all zero off-diagonal elements.

From Table II,

$$D_k = m_1\Delta_{11} + \sum_{i=2}^{k} (-1)^i \; c_{1i}\Delta_{1i} = \prod_{i=1}^{k} m_i$$

since

$$\Delta_{11} = \prod_{i=2}^{k} m_i \; , \quad c_{1i}\Delta_{1i} = 0 \; , \qquad \text{for all } i = 2, \ldots, k \; .$$

<u>Case 1</u>: Let $c_{12} \neq 0$,

then $\Delta_{12} = 0 = -c_{21}\, m_3 \;\ldots\; m_n$.

Since $m_3 \; \ldots \; m_n$ are greater than zero, $c_{21} = 0$. This implies that the second row has all zero off-diagonal elements. This is a contradiction, therefore

$$c_{12} = 0 \, .$$

<u>Case 2</u> Let j be the smallest integer such that

$$c_{1j} \neq 0 \, .$$

Then

$$\Delta_{1j} = 0 \, .$$

Now by shifting the first column to the j-th position, we obtain Δ_{1j} in the form

$$\begin{array}{c|cccccc|ccc|}
m_1 & 0 & 0 & 0 & . & 0 & -c_{1j} & -c_{1j+1} & . & -c_{1k} \\
\hline
. & m_2 & 0 & 0 & . & . & -c_{21} & 0 & . & 0 \\
. & -c_{32} & m_3 & 0 & . & . & -c_{31} & 0 & . & 0 \\
. & -c_{42} & -c_{43} & m_4 & 0 & . & -c_{41} & 0 & . & 0 \\
. & . & & & & & & & & \\
. & -c_{j-1,2} & . & . & . & m_{j-1} & . & 0 & . & 0 \\
. & -c_{j,2} & . & . & . & . & -c_{j,1} & 0 & . & 0 \\
\hline
. & . & & & & & & m_{j+1} & 0 & 0 \\
 & & & & & & & & & \\
. & . & & & & & & & & 0 \\
. & . & & & & & & . & . & m_k
\end{array}$$

$$\Delta_{1j} = m_{j+1} \cdots m_k \, \Delta' = 0 \, , \quad \text{therefore} \quad \Delta' = 0 \, ,$$

where Δ' is top left $j-1$ by $j-1$ determinant shown within the lines.

$$\Delta' \leq m_2 m_3 m_4 \dots m_{j-1} (-c_{j,1}) .$$

Since $m_2 \dots m_{j-1}$ are all greater than zero, we have $c_{j1} = 0$. Δ' is $j-1$ by $j-1$ matrix whose determinant is equal to the product of diagonals with off-diagonal terms non-positive. It is triangularable, and so from lemma 2, there is a row in D_k with zero off-diagonal elements. This implies in Table II a row with all off-diagonal elements equal to zero. This is a contradiction. Hence, $c_{1j} = 0$. So for all $j=2, \dots, k$, $c_{1j} = 0$. Therefore, we have a triangular form. Hence, the theorem is proved.

4.2 Triangularity of the Carry Matrix of Non-redundant Weighted Number Systems

Let ξ be a free module over Z with a basis $\{e_1, e_2, \dots, e_n\}$, as before and φ be a mapping of ξ onto Z_M such that

$$\left.\begin{array}{l}\varphi(e_i) = \rho_i \\ \rho_i \in Z_M\end{array}\right\} \text{ for } i = 1, 2, \dots, n . \tag{4.3}$$

Now, for any $x, y \in \xi$, let

$$x = (x_1, x_2, \dots, x_n)$$

and

$$y = (y_1, y_2, \dots, y_n)$$

with respect to the basis $\{e_1, e_2, \dots, e_n\}$. Then

$$\varphi(x) = \left| \sum_{i=1}^{n} \rho_i x_i \right|_M ,$$

since
$$\varphi(x+y) = \varphi(x_1+y_1,\ x_2+y_2,\ \ldots,\ x_n+y_n)\ ,$$

$$\varphi(x+y) = \left|\sum_{i=1}^{n} \rho_i(x_i+y_i)\right|_M$$
$$= \left|\sum_{i=1}^{n} \rho_i x_i\right|_M + \left|\sum_{i=1}^{n} \rho_i y_i\right|_M$$
$$= \varphi(x) + \varphi(y)\ .$$

Therefore φ is a homomorphism of ξ onto Z_M. Now consider a non-redundant number system N, with moduli $m_1, m_2, \ldots, m_n$, and digit-weights $\rho_1, \rho_2, \ldots \rho_n$. Then the carry relations of (3.4) are

$$\left.\begin{aligned} m_i\rho_i &= \sum_{\substack{j=1\\ j\neq i}}^{n} c_{ij}\rho_j \\ \text{and}\quad 0 &\leq c_{ij} < m_j \end{aligned}\right\} \text{for } i = 1, 2, \ldots, n\ .$$

The carry relation of (3.4) are equivalent to

$$\varphi(c_{i1},\ c_{i2},\ \ldots,\ c_{i,i-1},\ -m_i,\ c_{i,i+1},\ \ldots,\ c_{in}) = 0$$
$$\text{for } i = 1, 2, \ldots, n\ .$$

ρ_1 has a representation $(1, 0, \ldots, 0)\ \epsilon N$ and $M-\rho_1$ must have some representation of the form $(c_{11}, c_{12}, \ldots, c_{1n})\ \epsilon N$, so that $0 \leq c_{1j} \leq m_j-1$. Since the ring sum of ρ_1 and $M-\rho_1$ is zero the digit-wise sum of their representations $(c_{11}+1, c_{12}, \ldots, c_{1n})$ is in N, if $c_{11} \leq m_j-2$. This is contradictory. Therefore,

$$c_{11} = m_j-1$$

and

$$\varphi(m_1,\ c_{12},\ \ldots,\ c_{1n}) = 0 \qquad (4.4)\ *$$

---

* This form with all non-negative entries for s_i is important for this proof and was contributed by H. L. Garner.

Therefore the carry relation between the digit weights can be written in the modified form as below in (4.5).

$$\begin{bmatrix} m_1 & c_{12} & c_{13} & \cdot & \cdot & c_{1n} \\ -c_{21} & m_2 & -c_{23} & \cdot & \cdot & -c_{2n} \\ \cdot & & & & & \\ \cdot & & & & & \\ -c_{n1} & -c_{n2} & \cdot & \cdot & \cdot & m_n \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \cdot \\ \rho_n \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ 0 \end{bmatrix} \pmod{M} \qquad (4.5)$$

$$\text{where} \quad 0 \leq c_{ij} < m_j$$

Let the new carry matrix of (4.5) be S , and let the i-th row be denoted as $s_i \epsilon \xi$, and $\varphi(s_i) = 0$ for $i = 1, 2, \ldots, n$.

The totality $\{\alpha s_i\}$ for $\alpha \epsilon Z$ is a submodule of ξ and denoted as $\{s_i\}$. The generator s_i is minimal in the sense that no smaller integer than m_i can generate carries from the i-th digit, and the submodule $\{s_i\}$ is maximal, for all $i = 1, 2, \ldots, n$. Furthermore, if the determinant of S is non-zero, then the n generators $\{s_1, s_2, \ldots, s_n\}$ are independent, and they generate K , the kernel of φ .

Consider now a matrix S' obtained from S, by multiplying the first row of S by -1 . Then

$$\det. S' = - \det. S .$$

S' has now all off-diagonal elements non-positive, and except for the first one all diagonal elements are positive. Therefore from Theorem 6,

$$\det. S' \leq -\Pi\, m_i \qquad (4.6)$$

So det.S' and det.S are both non-zero and this establishes the independence of the generator set $\{s_1, s_2, \ldots, s_n\}$.

Thus S is identical to K .

Since φ is a group homomorphism of ξ onto Z_M , ξ/K is isomosphic to Z_M , and from Theorem 5 on modules

$$|\text{det. } K| = M = |\text{det. } S| \tag{4.7}$$

This is consistent with (4.6) only if det. S = M. Therefore, det. S' = -M .

Since S' satisfies the hypothesis of Theorem 7, S' must have a triangular form. So S also must have a triangular form. By diagonally permuting and reordering the subscripts we can obtain a triangular form of the carry matrix S . Thus we have proved the following important theorem.

Theorem 8.

The carry matrix of a non-redundant weighted number system has a triangular form.

Since S and K are identical, ξ/S is isomosphic to Z_M . Since the non-redundant system N is also isomosphic to Z_M , we have that the system N is isomosphic to its mathematical model ξ/S .

4.3 Examples of Quotient Module Structure

Example 1

A conventional non-redundant n-digit decimal system, having a range $M=10^n$, will have a carry matrix

$$S = \begin{bmatrix} 10 & 0 & . & . & 0 \\ -1 & 10 & . & . & 0 \\ . & & & & \\ 0 & . & . & -1 & 10 \end{bmatrix}$$

Since the digit weights are $\rho_1 = 10^{n-1}$, $\rho_2 = 10^{n-2}$, ..., $\rho_{n-1} = 10$, $\rho_n = 1$; the digit weight relations satisfy the condition below.

$$\begin{bmatrix} 10 & 0 & 0 & . & 0 \\ -1 & 10 & 0 & . & 0 \\ 0 & -1 & 10 & . & 0 \\ . & & & & . \\ . & & & & . \\ 0 & 0 & 0 & -1 & 10 \end{bmatrix} \begin{bmatrix} 10^{n-1} \\ 10^{n-2} \\ . \\ . \\ 10 \\ 1 \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix} \pmod{10^n}$$

Thus the n digit decimal system can be given a structure of a quotient module ξ/S .

<u>Example 2</u>

A consistently based system of k digits with all moduli $m_i = r$, having a range $M = r^k$, can be given a structure of

$$\xi/S = Z + Z + \ldots + Z \Big/ \begin{bmatrix} r & 0 & 0 & . & . & 0 \\ -1 & r & 0 & . & . & 0 \\ 0 & -1 & r & . & . & \\ . & & & & & \\ . & & & & & \\ 0 & 0 & . & . & -1 & r \end{bmatrix}$$

Example 3

A residue system with moduli 2, 3, 5, 7 (which are pairwise relatively prime) can represent integers modulo 210. The carry matrix

$$S = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix}$$

is diagonal, indicating that there are no carries generated by this system.

The weights are $\rho_1 = 105$, $\rho_2 = 70$, $\rho_3 = 126$, and $\rho_4 = 120$ such that the condition

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix} \begin{bmatrix} 105 \\ 70 \\ 126 \\ 120 \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \pmod{210}$$

is satisfied giving a structure of $Z + Z + Z + Z/S$ to the system.

Example 4

A redundant representation of integers modulo 7, using three binary digits, can be constructed as follows:

$$N = D_1 \times D_2 \times D_3 \qquad D_i = \{0,1\} \quad \text{for } i = 1, 2, 3.$$

Let the digit weights be $\rho_3 = 1$, $\rho_2 = 2$, $\rho_1 = 4$ as in conventional binary. Then the digit relations give a carry matrix

$$S = \begin{bmatrix} 2 & 0 & -1 \\ -1 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix}$$

such that condition (4.8) is satisfied. That is

$$\begin{bmatrix} 2 & 0 & -1 \\ -1 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \pmod 7 \qquad (4.8)$$

N is a redundant system, since 0 has two representations (0, 0, 0) and (1, 1, 1). Two properties should be noted. (1) The determinant of C = 7, and (2) S is not triangular.

V. GENERAL WEIGHTED SYSTEM STRUCTURE AND CANONICAL TRANSFORMATIONS

5.1 Brief Introduction

The discussion here is generalized to include the redundant number systems. Let N be a system with moduli $m_1, m_2, \ldots, m_n$ having a range $M \leq \Pi\, m_i$. (Note: $M < \Pi\, m_i$ makes the system redundant.) Since N is closed under addition, the sum of two elements in N , having a digitwise sum equal to $(0, 0, \ldots, m_i, \ldots, 0)$, is in N and is of the form $(c_{i1}, c_{i2}, \ldots, c_{ii}, \ldots, c_{in})$ for $0 \leq c_{ij} < m_j$ and for all $i = 1, 2, \ldots, n$. In the non-redundant case we proved in Section 3.2, that c_{ii} should be zero. But here c_{ii} need not be zero. Thus if $m_i - c_{ii} = c_i$ for $i = 1, 2, \ldots, n$, the relations between the digit-weights $\rho_1, \rho_2, \ldots, \rho_n$ and the range M of the system can be expressed as below in (5.1):

$$\begin{bmatrix} c_1 & -c_{12} & \cdot & \cdot & -c_{1n} \\ -c_{21} & c_2 & \cdot & \cdot & -c_{2n} \\ \cdot & & & & \\ \cdot & & & & \\ -c_{n1} & -c_{n2} & \cdot & \cdot & c_n \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdot \\ \\ \rho_n \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ \\ \\ 0 \end{bmatrix} \pmod{M} \quad (5.1)$$

where $0 \leq c_{ij} < m_j$, for $i, j = 1, 2, \ldots, n$.

The (nxn) matrix of (5.1) can be called the carry matrix as before and the structural similarity between N and ξ/S , (where the submodule S is constructed as earlier), can be established. The main difference from the non-redundant case is that S is different from K , the kernel of the mapping $\varphi : \xi \to Z_M$. The importance of K and the basis elements of K for understanding the arithmetic process in redundant weighted systems will be seen later when the canonical transformation and canonical reduction methods are dealt with. However, to make clear the distinction of the subgroups S and K in redundant systems, the following two examples are provided.

Example 5

Let N be a residue system with moduli 6 and 15. The cardinality of N is 90. Since the moduli are not relative prime, they can represent integers Z_M , where $M \leq \langle 6, 15 \rangle = 30$. Let M be 30. Since there are no carries generated in a residue system, the carry matrix will be $\begin{bmatrix} 6 & 0 \\ 0 & 15 \end{bmatrix}$. Assuming that $(1,1) \in N$ represents 1, digit weights can be $\rho_1 = 5$, $\rho_2 = 26$, satisfying

$$\begin{bmatrix} 6 & 0 \\ 0 & 15 \end{bmatrix} \begin{bmatrix} 5 \\ 26 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \pmod{30} ,$$

$$\rho_1 + \rho_2 = 5 + 26 \equiv 1 \pmod{30} .$$

ξ in this example is pairs of integers, and the mapping $\varphi : \xi \to Z_{30}$ is such that

$$\varphi(x_1, x_2) = |5x_1 + 26x_2|_{30} .$$

Trivially,

$$\varphi(6,0) = \varphi(0,15) = 0 .$$

However, (6,0) and (0,15) cannot be generators of K since $\varphi(-2,5) = -2.5 + 5.26 \equiv 0 \bmod 30$, and (-2,5) is not in the submodule generated by (6,0) and (0,15). On the other hand, (2,-5) and (0,15) generate K. Equally well (6,0) and (2,-5) generate K.

Example 6

Let N be a system with moduli $m_1 = m_2 = m_3 = m_4 = 2$ representing Z_5 and let $\rho_4 = 1$, $\rho_3 = 2$, $\rho_2 = 2^2 = 4$, $\rho_1 = 2^3 = 3 \pmod 5$. Carry matrix C can be written as

$$C = \begin{bmatrix} 2 & 0 & 0 & -1 \\ -1 & 2 & 0 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

Also we have the relations between the digit weights as follows:

$$\begin{bmatrix} 2 & 0 & 0 & -1 \\ -1 & 2 & 0 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 2 \\ 1 \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \pmod 5 .$$

It can easily be seen that C is not equal to K = kernel φ where $\varphi : \xi \to Z_5$ such that

$$\varphi(x_1,x_2,x_3,x_4) = \left|3x_1 + 4x_2 + 2x_3 + x_4\right|_5$$

for all $x_i \in Z$. This is because $(0,1,0,1) = |0 + 4 + 0 + 1|_5 = 0$ and (0,1,0,1) is not in the space generated by the row elements of C.

However, it can be observed that the row elements of the matrix

$$\begin{bmatrix} -1 & 2 & 0 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 0 & -1 & 2 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

generate K . K certainly contains C , since (2,0,0,-1) can be obtained readily from the rows of the K matrix. It is also interesting to observe that determinant C = 15 and determinant K = 5 , showing that ξ/C has cardinality 15 and ξ/K is isomorphic to Z_5. This example will be further investigated in later sections where the arithmetic process in redundant systems is explained.

5.2 Condition on the Determinant of the Carry Matrix

In weighted systems the digit weight relations expressed in the form of (5.1) govern the arithmetic process. Some significant results can be obtained from the following theorem relating to the condition (5.1).

Theorem 9

The n independent linear congruences expressed below as

$$\begin{bmatrix} c_{11} & c_{12} & \cdot & c_{1n} \\ c_{21} & c_{22} & \cdot & c_{2n} \\ \cdot & & & \\ c_{n1} & c_{n2} & & c_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \\ x_n \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \\ 0 \end{bmatrix} \pmod{M} \qquad (5.2)$$

have solutions $x_i = \rho_i$ where $(\rho_1, \rho_2, \ldots, \rho_n, M) = 1$ if only M divides the determinant of the (nxn) matrix above.

A conventional proof of the above theorem can be found in the Appendix. A module theoretic proof will be provided here, to exhibit the usefulness of module theory as such to weighted systems.

Proof: Consider a weighted system N with n moduli and corresponding digit weights satisfy the linear congruences as in the theorem

$$\left.\begin{array}{l} c_{11}\rho_1 + c_{12}\rho_2 + \;\ldots\ldots + c_{1n}\rho_n \equiv 0 \\ c_{21}\rho_1 + c_{22}\rho_2 + \;\ldots\ldots + c_{2n}\rho_n \equiv 0 \\ \\ c_{n1}\rho_1 + c_{n2}\rho_2 + \;\ldots\ldots + c_{nn}\rho_n \equiv 0 \end{array}\right\} \pmod{M} \qquad (5.3)$$

Then from the definition made earlier, M can be a given structure of a quotient Z-module ξ/S where S is the submodule represented by the (nxn) matrix of (5.2).

Let $\varphi : \xi \rightarrow Z_M$ be defined as in (4.3) . It is proved in Section 4.2, that φ is a homomosphism and ξ/K is isomorphic to Z_M , so has cardinality equal to M .

Since S is a subroup of K , and ξ/K is a subgroup of ξ/S , the cardinality of the set ξ/K divides the cardinality of ξ/S .

Therefore it is evident that M divides determinant S .

5.3 Canonical Forms and Transformations

Let N be a redundant weighted system defined as in Section 5.1. The relations between the digit weights of N can be expressed as (5.1).

Also let $\varphi : \xi \rightarrow Z_M$ be defined as before in Section 4.2. Since $N \subset \xi$, φ is a mapping of N onto Z_M. Define an equivalence relation $\sim$ in N such that

$$\forall\, x,y \in N,\; x \sim y \iff \varphi(x) = \varphi(y) \,. \tag{5.4}$$

This amounts to saying that all such elements in N representing any particular element in Z_M are considered equivalent. Since the map φ of N is onto Z_M, (in order that N be a complete system) there are exactly M equivalence classes in N. Whenever these equivalence classes contain a large number of elements, there may be a few elements with some advantages. Some elements may have fewer 1's (when expressed in binary coded form) than the other members of the class. Or in certain other examples, some other interesting properties could be sought for. In any case, such a desired form for elements in N will be called the <u>canonical form</u> or <u>canonical property</u>. The totality of elements in N having a canonical property is said to be a <u>canonical subset</u> C. A necessary condition on the canonical form is that there exist at least one element in canonical form in each equivalence class. However, there exists some difficulty with this setup. The arithmetic sum of two elements in C, obtained by means of the carry transformation associated with the rows of S may not satisfy the closure law. That is, there may be elements $x,y \in C$ such that $x + y \in N$ but $x + y \notin C$. The additional transformations necessary for putting the result in C are called <u>canonical transformations</u>. Recovery of any element in an equivalence class to the canonical form is called <u>canonical reduction</u>. If T is the set of all canonical transformations, then $t \in T$ must satisfy the following criterion.

$$\left.\begin{array}{ll} 1) & (\forall)x\in N \quad t(x)\sim x \quad \text{or} \quad t(x) = x \quad (\text{mod } M) \\ 2) & (\exists)y\in N,\ y\in C,\ t(y)\in C \\ 3) & (\forall)Z\in C,\ t(z)\in C\ . \end{array}\right\} (5.5)$$

Carry transformations satisfy the above criterion, but they are not sufficient for taking any digitwise sum into the canonical form.

The major problem of canonical reduction is finding the exact compound transformations which take every possible digitwise sum into canonical form. This can be obtained for any particular system by investigation of the transformations derived from the basis or generator elements of K instead of S.

Since $K \supseteq S$ and determinant of $K = M$, by theorem, the basis transformation of K must be sufficient for reducing any element of ξ into the required form.

In particular, if the basis of K is triangular for the system N, then the transformations associated with the row elements of K, will simplify the canonical reduction. The exact nature of the transformations can be given and the time taken for the implementation of the canonical reduction can be estimated. This type of system could be conceived for practical use.

The following simple examples demonstrate the canonical reduction problem and in each case the necessary canonical transformation is suggested.

5.4 Some Examples of Redundant Systems and Their Canonical Forms

Example 7

Example 6 will be reviewed here. N is a system with moduli

$m_1 = m_2 = m_3 = m_4 = 2$. N is a set of 16 elements, starting from (0,0,0,0), (0,0,0,1), ..., up to (1,1,1,1), representing Z_5 . The digit weights of the system are given as

$$\rho_4 = 1$$
$$\rho_3 = 2$$
$$\rho_2 = 4$$
$$\rho_1 = 8 \equiv 3 \pmod 5 .$$

Since $\rho_4 \equiv 2\rho_1 \pmod 5$, an end around carry exists in the system.

The set N divides into 5 equivalence classes, representing integers 0,1,2,3, and 4 as below.

0	1	2	3	4
[0 0 0 0]	[0 0 0 1]	[0 0 1 0]	0 0 1 1	[0 1 0 0]
0 1 0 1	0 1 1 0	0 1 1 1	[1 0 0 0]	1 0 0 1
1 0 1 0	1 0 1 1	1 1 0 0	1 1 0 1	1 1 1 0
1 1 1 1				

If we define that the canonical form has not more than a single 1 among the four bits, then there are exactly five elements in the canonical subset C , one in each equivalence class, shown within the enclosures in the table. The muliplication of any two elements in C is carried out by a suitable shift and the result is also in C . This multiplication is very simple and correspondingly faster compared with a conventional binary notation. The addition of two elements in C may result in two 1's, then the result is not in C . A simple canonical transformation will then be necessary. If addition of two non-zero elements in C does not generate a carry, a canonical transformation is necessary. If they produce a carry, the result is already in canonical form.

One canonical transformation t_1 will replace two nonadjacent 1's by 0's. It is so because

$$\varphi(1,0,1,0) = 0 = \varphi(0,1,0,1) .$$

Another transformation t_2 will transform two adjacent 1's into zeros followed by a 1 in the next place to the right.

$$t_2 : \quad \begin{array}{l} 0\,1\,1\,0 \rightarrow 0\,0\,0\,1 \\ 1\,1\,0\,0 \rightarrow 0\,0\,1\,0 \\ 1\,0\,0\,1 \rightarrow 0\,1\,0\,0 \\ 0\,0\,1\,1 \rightarrow 1\,0\,0\,0 \end{array}$$

These transformations can be constructed without much difficulty, but in view of the simple example, the desirability of the transformation and therefore the usefulness of the code is questionable. But multiplication is far simpler than in any other known weighted code representation for integers modulo 5.

Example 8

Let N be a residue system with moduli 6,15 and 21, and (1,1,1) be a representation for 1 in the system. N can represent integers modulo M where M = Least Common Multiple of the moduli = $< 6,15,21 > = 210$. It will be shown in Section 6.2 that the following two conditions on digit weights ρ_1, ρ_2, ρ_3 have to be satisfied:

1)
$$\begin{bmatrix} 6 & 0 & 0 \\ 0 & 15 & 0 \\ 0 & 0 & 21 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mod 210 . \qquad (5.6)$$

2)
$$\rho_1 + \rho_2 + \rho_3 \equiv 1 \pmod{210} .$$

It will be proved later in Section 6.3, that there are $\frac{15 \cdot 6 \cdot 21}{210} = 9$ different sets of digit weights satisfying the above conditions. One of the nine sets of digit weights is (35,56,120), as they satisfy (5.6):

$$6 \times 35 \equiv 15 \times 56 \equiv 21 \times 120 \equiv 0 \pmod{210}$$
$$35 + 56 + 120 = 211 \equiv 1 \pmod{210} .$$

N is a set of $6 \times 15 \times 21 = 1890$ elements representing Z_{210}. The subset $H \subset N$, containing all elements generated by (1,1,1), constitute the non-redundant system representing Z_{210}. The system H has interesting error detecting properties, due to the fact that the moduli are not relatively prime. 3 divides all the moduli 6, 15, and 21.

If $Y \epsilon Z_{210}$, then Y has a representation in H as (y_1, y_2, y_3) where

$$y_i \equiv Y \pmod{m_i} \quad i = 1,2,3.$$
$$= Y - a_i m_i \qquad \text{for some integers } a_i$$

$$y_i - y_j = y - a_i m_i - (y - a_j m_j)$$
$$= a_j m_j - a_i m_i$$

Since 3 divides m_j and m_i, 3 divides the $a_j m_j - a_i m_i$ so also $y_i - y_j$.

If the residues of H are coded in binary form, a single error in the bits would result in a change of 1, 2, or 4, etc. Since 3 is not a factor of the error, it can be detected. Thus, if H is considered the canonical subset of N , any single error would take it out of the canonical form. Error correction could be used to put it back in its canonical form. This particular code is obviously unsuitable for error correction, since the error does not keep it in the equivalence class. On the other hand, if C is a canonical subset, containing all the elements of the form (x_1, x_2, x_3) where $0 \leq x_i \leq 7$, then any element in its canonical form needs only 3 bits for each digit.

Carry transformations based on the rows of

$$S = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 15 & 0 \\ 0 & 0 & 21 \end{bmatrix}$$

are not sufficient for canonical reduction, since there is no way to continue if the sum in the second or third digit exceeds 7. This is where the canonical transformation is required.

If $T = \begin{bmatrix} 6 & 0 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 7 \end{bmatrix}$ is the associated matrix of submodule T of ξ, then

$$\begin{bmatrix} 6 & 0 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 7 \end{bmatrix} \begin{bmatrix} 35 \\ 56 \\ 120 \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \mod 210 .$$

Thus, if

$$\varphi : \xi \to Z_{210}$$

such that

$$\varphi(x_1,x_2,x_3) = \left|35x_1 + 56x_2 + 120x_3\right|_{210}$$

then $\varphi : T \to 0$ and $T \supseteq S$.

It can be seen that the row elements of T provide the transformation that is sufficient for canonical reduction. However, one of the canonical transformations involves a carry propagation of 2, into the first digit, whenever the sum in the second digit exceeds 4. These canonical transformations can also be viewed as a homomorphism of the redundant sustem N onto a non-redundant weighted system C of moduli 6, 5, and 7. Thus, the cardinalities of N, H and D are:

$$(N) = 1890$$

$$(H) = 210$$

$$\text{and} \quad (C) = 210 \, .$$

Thus, if H is considered as a system representing Z_{210} , single error detection in H is possible, A mapping of N to C can be carried out by canonical reduction based on the row elements of T .

Example 9

Consider a system N with moduli $m_1 = m_2 = \ldots = m_k = r+1$ and the digit weights $\rho_k = 1$, $\rho_{k-1} = r$, ..., $\rho_2 = r^{k-2}$, $\rho_1 = r^{k-1}$ representing integers modulo r^k . This system has the same set of digit weights as a consistently based system with $m_i = r$. The digit weight relations satisfy the condition (4.5).

$$\begin{bmatrix} r & 0 & 0 & \cdot & \cdot & 0 \\ -1 & r & 0 & \cdot & \cdot & 0 \\ \cdot & & & & & \\ \cdot & & & & & \\ 0 & 0 & 0 & \cdot & \cdot & r \end{bmatrix} \begin{bmatrix} r^{k-1} \\ r^{k-2} \\ \cdot \\ r \\ 1 \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \\ 0 \end{bmatrix} \pmod{r^k} .$$

If the submodule generated by the rows of the above square matrix is S , then N has a structure of ξ/S . N is a redundant system with cardinality = $(r + 1)^k > r^k$.

Addition of two numbers in N is possible by carry generation based on the structure of ξ/S . The increase by 1 of the moduli, can be used to restrict the carry propagation to at most one level. On this basis it can be said that the addition is totally parallel. Totally parallel addition is defined to be the case when a digitwise sum produces a carry and partial sum, the carries getting absorbed without producing further carries. Based on this idea there are some interesting redundant systems such as signed digit representation(3) in which any digit can be negative or positive, and carries or borrows can be propagated to higher ordered digits.

Of particular interest is symmetric signed digit representation(3) in which each digit position can take a value from $-k$ to $+k$ where $2k + 1 \leq r + 2$. Even though the addition is complicated by canonical reduction, the symmetric signed digit representation was shown to have many computational advantages besides totally parallel addition.

Then preceding examples are meant to explain the arithmetic process in redundant systems and codes. The logical advantages of these coding methods depend largely on the canonical reduction methods.

Although the carry assimilation and the canonical reduction can be combined and the arithmetic operation can be obtained in one step, it will be desirable sometimes to treat them separately. Since the coding simplifies the carry process, which is the first stage of the operation, it can be performed and the canonical reduction can be left for some other convenient time. It might be possible to do canonical reduction in parallel with other operations. In this way some time sharing techniques can be used. Also redundant codes for residue number systems can be shown to be advantageous. The other aspect of redundant systems is error checking in arithmetic operations. This problem has been studied and several coding methods have been suggested for consistently based number systems by other researchers.(8,9) Methods of error checking in residue arithmetic are covered in Chapter VII.

VI. REDUNDANCY IN RESIDUE NUMBER SYSTEMS

6.1 Introduction and Results

This chapter investigates the conditions for a finite, redundant residue system using the moduli, $m_1, m_2, \ldots, m_n$ to represent integers modulo M. It will be proved that for a general residue system (without any restriction on moduli) it is necessary and sufficient that M be a divisor of the $< m_1, m_2, \ldots, m_n >$ in order that the system be redundant and weighted. M need not be a divisor of Πm_i for redundant, non-residue systems (refer to example 4 on page 41). It will also be proved that for a residue system, if $M = < m_1, m_2, \ldots, m_n >$, there exist exactly d sets of digit weights for the system where d is called the factor of redundancy and is given as

$$d = \frac{m_1 m_2 \cdots m_n}{< m_1, m_2, \ldots, m_n >} = \frac{\prod_{i=1}^{n} m_i}{M} \qquad (6.1)$$

These results are useful in the discussion of the methods of error-checking in the arithmetic of residue systems as described in the next chapter.

6.2 Necessary and Sufficient Conditions on the Digit Weights of a Residue System

Lemma 5. $\rho_1, \rho_2, \ldots, \rho_n$ is a set of digit weights for a general residue system N with moduli $m_1, m_2, \ldots, m_n$, and range M, if and only if

$$\left.\begin{array}{ll} i = 1, 2, \ldots, n & m_i\rho_i \equiv 0 \pmod{M} \\ & (\rho_1, \rho_2, \ldots, \rho_n, M) = 1 \end{array}\right\} \qquad (6.2)$$

The reader should note the distinction between the greatest common divisor $(x_1, x_2, \ldots, x_n)$ and an element of N or ξ indicated as $(c_1, c_2, \ldots, c_n) \in N$ or ξ.

Proof: N is a residue system with moduli $m_1, m_2, \ldots, m_n$ and the corresponding weights $\rho_1, \rho_2, \ldots, \rho_n$ representing Z_M. Since there are no carries in the system, $m_i\rho_i \equiv 0 \pmod{M}$ for $i = 1,2,\ldots,n$.

If $(a_1, a_2, \ldots, a_n)$ represents 1 in the system, then

$$\left|\sum a_i\rho_i\right|_M = 1 \quad .$$

This implies $(\rho_1, \rho_2, \ldots, \rho_n, M) = 1$. Conversely, if a residue system N with weights $(\rho_1, \rho_2, \ldots, \rho_n)$ satisfies (6.2), then

$$\exists\ a_1, a_2, \ldots, a_n, a_{n+1}$$

such that

$$\sum_{i=1}^{n} a_i\rho_i + a_{n+1} M = 1$$

Therefore,

$$\left|\sum_{i=1}^{n} a_i\rho_i\right|_M = 1$$

So $(a_1, a_2, \ldots, a_n)$ represents 1 in the system. For an $x \in Z_M$ there exists $(x_1, x_2, \ldots, x_n) \in N$ such that

$$x_i = \left| X a_i \right|_{m_i}$$

and

$$\left| \Sigma x_i \rho_i \right|_M = x$$

Also, x, $y \epsilon Z_M$, $x \neq y$. Then

$$x_i = \left| x a_i \right|_{m_i}, \quad x = \left| \Sigma x_i \rho_i \right|_M$$

$$y_i = \left| y a_i \right|_{m_i}, \quad y = \left| \Sigma y_i \rho_i \right|_M$$

So $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ have to be distinct. Thus, all integers in Z_M have a representation in N and that completes the proof of the lemma.

The residue system in which (1, 1,..., 1) represents $1 \epsilon Z_M$ is sometimes called the system of residue classes. In such systems

$$\sum_{i=1}^{n} \rho_i \equiv 1 \pmod{M}$$

and the necessary and sufficient conditions of (6.1) are modified as

$$\left.\begin{array}{ll} \rho_i m_i \equiv 0 \pmod{M} & \text{for } i = 1, 2, \ldots, n \\ \sum_{i=1}^{n} \rho_i \equiv 1 \pmod{M} & \end{array}\right\} \quad (6.3)$$

6.3 Number of Acceptable Sets of Digit Weights of a Residue System with Moduli That are Not Relatively Prime

The following theorems are for residue systems which have a representation (1, 1, ..., 1) for 1, and so (6.3) applies. However, they can all be proved for the general case by using (6.2).

Theorem 10.

A necessary and sufficient condition that a congruence

$$\sum_{i=1}^{n} k_i \frac{M}{m_i} \equiv 1 \pmod{M}$$

where

$$M = \langle m_1, m_2, \ldots, m_n \rangle$$

is solvable for $k_1, \ldots, k_n$ is that

$$(M/m_1, M/m_2, \ldots, M/m_n, M) = 1 \qquad (6.4)$$

If d is defined as in (6.1), then there are exactly d sets of solutions to $k_1, k_2, \ldots, k_n$ such that $0 \leq k_i \leq m_i - 1$.

Proof: Let $(m_1, m_2) = d_{12}, \quad M_{12} = \dfrac{m_1 m_2}{d_{12}}$

$$(M_{12}, m_3) = d_{13}, \quad M_{13} = \frac{m_1 m_2 m_3}{d_{12} d_{13}}$$

$$(M_{13}, m_4) = d_{14}, \quad m_{14} = \frac{m_1 m_2 m_3}{d_{12} d_{13} d_{14}}$$

$$(M_{1,n-1}, m_n) = d_{1n}, \quad M = M_{1n} = \frac{m_1 m_2 \ldots m_n}{d_{12} d_{13} \ldots d_{1n}}$$

d in the theorem is now obtained by

$$d = \frac{m_1 m_2 \cdots m_n}{M} = d_{12} d_{13} \cdots d_{1n} .$$

The first part of the theorem stating the necessary and sufficient condition for solvability is well established and proved in most of the textbooks on number theory.[5] However, we will show that the condition (6.4) is true.

Let

$$(M/m_1, M/m_2, \ldots, M/m_n, M) = d$$

then

$$(M/(m_1 d), M/(m_2 d), \ldots, M/(m_n d), M/d) = 1$$

$$(\tfrac{M}{d}/m_1, \tfrac{M}{d}/m_2, \ldots, \tfrac{M}{d}/m_n, M/d) = 1$$

So

$$m_i \mid M/d \quad \text{for } i = 1, 2, \ldots n .$$

Therefore M/d is a common multiple of $m_1, m_2, \ldots, m_n$.
The least common multiple, M divides all common multiples of $m_1, m_2, \ldots, m_n$.
Therefore M divides M/d, and so $d = 1$.
Thus we have proved the condition (6.4). We have the congruence

$$k_1 \frac{M}{m_1} + k_2 \frac{M}{m_2} + \ldots + k_n \frac{M}{m_n} \equiv 1 \pmod{M}$$

From the definition of $d_{12}, \ldots, d_{1n}$ we have

$$\left(\frac{M}{m_1}, \frac{M}{m_2}, \ldots, \frac{M}{m_{n-1}}, M\right)$$

$$= \left(\frac{m_2 \cdot\cdot m_n}{d_{12}\cdot\cdot d_{1n}}, \frac{m_1 m_3 \cdots m_n}{d_{12}d_{13}\cdot\cdot d_{1n}}, \ldots, \frac{m_1 m_2 \cdots m_{n-2}m_n}{d_{12}d_{13} \cdot\cdot d_{1n}}\right)$$

Using the formulas (1), (2) and (3) given below

(1) If $(m_1, m_2) = d_{12}$

then $\left(\frac{m_2}{d_{12}}, \frac{m_1}{d_{12}}\right) = 1$

(2) $\left(\frac{a}{t}, \frac{b}{t}\right) = \frac{(a,b)}{t}$; $(ta, tb) = t(a,b)$

(3) $(x_1, x_2, \ldots x_n) = (\ldots(((x_1, x_2), x_3), x_4), \ldots, x_n),$

we have

$$\left(\frac{m_2 \cdots m_n}{d_{12}\cdots d_{1n}}, \frac{m_1 m_3 \cdots m_n}{d_{12} \cdots d_{1n}}\right) = \frac{m_3 \cdots\cdot m_n}{d_{13} \cdots d_{1n}}$$

$$\left(\frac{m_3 \cdots m_n}{d_{13}\cdots d_{1n}}, \frac{m_1 m_2 m_4 \cdots m_n}{d_{12}d_{13} \cdots d_{1n}}\right) = \frac{m_4 \cdots m_n}{d_{14}\cdots d_{1n}}$$

because

$$\left(\frac{M_{12}}{d_{13}}, \frac{m_3}{d_{13}}\right) = 1; \qquad (M_{12}, m_3) = d_{13} ; \quad \frac{m_1 m_2}{d_{12}} = M_{12}$$

Continuing the process, we get

$$\left(\frac{m_{n-1}\, m_n}{d_{1,n-1} d_{1n}} , \frac{m_1 m_2 \cdots m_{n-2} m_n}{d_{12} d_{13} \cdots d_{1n}}\right) = \frac{m_n}{d_{1n}} ,$$

because

$$\left(\frac{m_{n-1}}{d_{1,n-1}}, \frac{M_{1,n-1}}{d_{1,n-1}}\right) = 1$$

Therefore $\left(\frac{M}{m_1} , \frac{M}{m_2} \cdots \frac{M}{m_{n-1}} , M\right) = \frac{m_n}{d_{1n}}$

$$k_1 \frac{M}{m_1} + k_2 \frac{M}{m_2} +, \ldots, + k_{n-1} \frac{M}{m_{n-1}} + k_n \frac{M}{m_n} \equiv 1 \pmod M$$

From the above two equations and from (6.4) we have

$$\left(\frac{M}{m_n} , \frac{m_n}{d_{1n}}\right) = 1$$

and

$$k_n \frac{M}{m_n} \equiv 1 \pmod{m_n/d_{1n}}$$

k_n has exactly one solution mod m_n/d_{1n}; however, it has d_{1n} solutions mod m_n or $0 \leq k_n \leq m_{n-1}$. Now substituting for k_n one of the $d_{1,n}$ possible values we obtain a congruence in $n-1$ variables

$$k_1 \frac{M}{m_1} + k_2 \frac{M}{m_2} + k_{n-1} \frac{M}{m_{n-1}} \equiv \left(1 - k_n \frac{M}{m_n}\right) \pmod M .$$

This equation is divisible on both sides by

$$\frac{m_n}{d_{1n}}$$

Thus we have

$$k_1 \frac{M_{1,n-1}}{m_1} + k_2 \frac{M_{1,n-1}}{m_2} \ \ldots + k_{n-1} \frac{M_{1,n-1}}{m_{n-1}} = C_{n-1} \pmod{M_{1,n-1}} .$$

Repeating the same step

$$\left(\frac{M_{1,n-1}}{m_1}, \frac{M_{1,n-1}}{m_2} \ldots \frac{M_{1,n-1}}{m_{n-2}}, M_{1,n-1} \right) = \frac{m_{n-1}}{d_{1,n-1}}$$

we can show that k_{n-1} has exactly $d_{1,n-1}$ solutions modulo m_{n-1} and k_{n-2} has $d_{1,n-2}$ and so on. This proves that we have a total of

$$d_{1n} \cdot d_{1,n-1} \, d_{1n-2} \cdots d_{12} = d$$

solutions for $k_1, k_2, \ldots, k_n$, such that $0 \le k_i \le m_i - 1$. Hence the theorem is proved.

From the above theorem the congruence

$$\sum_{i=1}^{n} k_i \frac{M}{m_i} \equiv 1 \pmod{M}$$

has d sets of solutions for $k_1, \ldots, k_n$, such that $0 \le k_i < m_i$. Now applying (6.3) on the digit weights of a residue system N with the operating moduli $m_1, m_2 \ldots m_n$ we have $\rho_i = k_i \, M/m_i$

$$\sum_{1}^{n} \rho_i = \sum_{1}^{n} k_i \, M/m_i \equiv 1 \pmod{M}$$

which has d sets of solutions for $k_1, \ldots, k_n$, $0 \leq k_i \leq m_i$. Thus we have proved the theorem stated below.

Theorem 11.

For a residue system N with moduli $m_1, m_2, \ldots m_n$ representing integers modulo M where $M = < m_1, m_2, \ldots, m_n >$, there are exactly

$$d = \frac{m_1 m_2 \cdots m_n}{M}$$

sets of digit weights.

Example 10. Consider a residue system N with moduli 6, 10, 21 representing Z_{210}. Then

$$d = \frac{6 \cdot 10 \cdot 21}{210} = 6 \; ; \quad m_1 = 6,\ m_2 = 10,\ m_3 = 21.$$

The six sets of digit weights are given below.

ρ_1	ρ_2	ρ_3
0	21	190
35	126	50
175	126	120
70	21	120
105	126	190
140	21	50

$$\rho_1 + \rho_2 + \rho_3 \equiv 1 \pmod{210}$$

$$m_1\rho_1 \equiv m_2\rho_2 \equiv m_3\rho_3 \equiv 0 \pmod{210}$$

6.4 Condition on the Range M of a Residue System

Theorem 12.

$m_1, m_2, \ldots, m_n$ are the moduli of a residue system N. N can represent integers modulo M, if and only if M is a divisor of $< m_1, m_2, \ldots, m_n >$.

Proof: If N is a residue system, then $\exists \rho_1, \rho_2, \ldots, \rho_n \in Z_M$ satisfying the condition (6.3). For any i, if $(m_i, M) = 1$

$$\rho_i m_i \equiv 0 \bmod M$$
$$= k_i M$$
$$M | \rho_i m_i \qquad (M, m_i) = 1$$

Therefore $M | \rho_i$

Therefore $\rho_i = c_i M$ for some integer c_i

$$\equiv 0 \pmod M$$

This means that for any modulo that is relatively prime to M its digit weight is zero. Such digits exist in the system as purely redundant digits. Assume there are r such bits where $0 \leq r < n$. If $r = n$ then $\rho_i = 0$ for $i = 1, 2, \ldots n$ and $\sum \rho_i \equiv 0 \pmod M$ which contradicts condition (6.1).

Now reordering the moduli so that the last r moduli

$$m_{n-r+1}, m_{n-r+2}, \ldots, m_n$$

are the ones that are relatively prime to M, let

$$(M, m_i) = d_i \qquad \text{for} \quad i = 1, 2, \ldots, n-r$$

and

$$\frac{M}{d_i} = M_i'$$

$$d_i > 1 .$$

Then applying the first part of (6.3), we have

$$\rho_i m_i \equiv 0 \pmod M$$
$$= k_i M .$$

Dividing by d_i

$$\rho_i \frac{m_i}{d_i} = k_i \frac{M}{d_i}$$

Let $$\frac{m_i}{d_i} = m_i' \ ; \qquad \frac{M}{d_i} = M_i'$$

such that

$$(m_i', M_i') = 1 \ .$$

$$\rho_i = \frac{k_i}{m_i'} M_i = c_i \frac{M}{d_i} \qquad \text{for some integer } c_i$$

for $i = 1, 2, \ldots n-r$.

Applying the second part of (6.3), we have

$$\sum_1^{n-r} c_i \frac{M}{d_i} - CM = 1 \ .$$

This equation has solutions for ρ_i if and only if

$$\left(\frac{M}{d_1}, \frac{M}{d_2}, \ldots \frac{M}{d_{n-r}}, M\right) = 1 \ .$$

This is possible only if

$$M = \langle d_1, d_2, \ldots d_{n-r} \rangle$$

$$= \left\langle \frac{m_1}{m_1'}, \frac{m_2}{m_2'}, \ldots, \frac{m_{n-r}}{m_{n-r}'} \right\rangle$$

which implies that

$$M \,\Big|\, \langle m_1, m_2, \ldots, m_{n-r} \rangle$$

and so divides

$$< m_1, m_2, \ldots, m_n >$$

Now, if

$$M \text{ divides } < m_1, m_2, \ldots, m_n >$$

it is necessary to prove the following.

Claim. N represents Z_M. If we can show that $\exists\, \rho_1, \rho_2, \ldots, \rho_n$ satisfying (6.3), then we will have completed the proof.

We know that N with moduli $m_1, m_2, \ldots, m_n$ represents Z_t where $t = < m_1, m_2, \ldots, m_n >$. Let $\rho_1, \rho_2, \ldots, \rho_n \in Z_t$ be the digit weights of the weight functions $W:N \to Z_t$. Since M divides t we have $\rho_i m_i \equiv 0 \pmod t$

$$\rho_i m_i \equiv 0 \pmod t \Longrightarrow \rho_i m_i \equiv 0 \pmod M$$

$$\Sigma\, \rho_i \equiv 1 \pmod t \Longrightarrow \Sigma\, \rho_i \equiv 1 \pmod M$$

so $|\rho_1|_M, \ldots, |\rho_n|_M$ are the digit weights for the system $N \to Z_M$. so the theorem is proved.

Let $m_1, m_2, \ldots, m_n$ be the moduli of a residue system N having a range $M = < m_1, m_2, \ldots, m_n >$ and let $\rho_1, \rho_2, \ldots, \rho_n$ be one of the

$$d = \frac{\Pi\, m_i}{M}$$

sets of acceptable weights. As explained in Chapter IV, the system N can be given a quotient structure ξ/S, where S is the row space of the diagonal matrix.

$$\begin{bmatrix} m_1 & & & \bigcirc \\ & m_2 & & \\ & & \ddots & \\ \bigcirc & & & m_n \end{bmatrix}$$

If $m'_1, m'_2, \ldots, m'_n$ are the smallest integers satisfying $m'_i \rho_i \equiv 0 \pmod{M}$, then S' is the row space of the matrix

$$\begin{bmatrix} m'_1 & & & \bigcirc \\ & m'_2 & & \\ & & \ddots & \\ \bigcirc & & & m'_n \end{bmatrix}$$

which contains S. Further, if $m'_1, m'_2, \ldots, m'_n$ are all pairwise relatively prime, then $\Pi\, m'_i = M$ and S' will be identical to the kernel of φ where $\varphi\colon \xi \to Z_M$ such that

$$\varphi(e_i) = \rho_i \quad \text{for } i = 1, 2, \ldots, n.$$

Then the residue system N' with moduli $m'_1, m'_2, \ldots, m'_n$ and weights $\rho_1, \rho_2, \ldots, \rho_n$ having a range M, is nonredundant. The redundant system N can be considered an extension or coded form of N' and the factors

$$\frac{m_i}{m'_i}$$

in that case will be called the coding factors. These factors can be used for error checking purposes as will be described later in the next chapter. Also, simplification of any residue number in N can be done by means of the transformations based on S'.

However, the condition that there exist a set of weights $\rho_1, \rho_2, \ldots, \rho_n$ of N for which the corresponding $m'_1, m'_2, \ldots, m'_n$ are all pairwise relatively prime is yet to be established. This can be done by choosing a set $m'_1, m'_2, \ldots, m'_n$ satisfying (6.5).

$$m'_1 = m_1$$

$$m'_j = \frac{< m_1, m_2, \ldots, m_{j-1}, m_j >}{\prod_{i=1}^{j-1} m'_i} \quad \text{for } j = 2, \ldots, n \tag{6.5}$$

This way we can obtain $m'_1, m'_2, \ldots, m'_n$ that are pairwise relatively prime and

$$\prod_{i=1}^{n} m'_i = < m_1, m_2, \ldots, m_n > .$$

The residue system N' with moduli $m'_1, m'_2, \ldots, m'_n$ having a range equal to $\Pi\, m'_i = M$, will have weights $\rho'_1, \rho'_2, \ldots, \rho'_n$ that satisfy

$$m'_i\, \rho'_i \equiv 0 \pmod M$$

$$\sum \rho'_i \equiv 1 \pmod M$$

Since m'_i divides m_i, $m_i \rho'_i \equiv 0 \pmod M$, and therefore $\rho'_1, \rho'_2, \ldots, \rho'_n$ is an acceptable set of weights of N. This establishes the desired condition. Since the ordering of the moduli is arbitrary there can be more than one such set of moduli $m'_1, m'_2, \ldots, m'_n$ and also of the corresponding weights. Table III shows that the residue system with moduli 6, 10 and 21, has four such sets of weights, leading to m_1', m_2', m_3' that are relatively prime,

$$d = \frac{\Pi m_i}{M} = \frac{6 \times 10 \times 21}{210} = 6$$

for that system. The six sets of weights and the corresponding values of m_1', m_2', m_3' are listed in the table below.

TABLE III

DIGIT WEIGHTS AND CORRESPONDING VALUES OF m_1', m_2', m_3'

	ρ_1	ρ_2	ρ_3	m_1'	m_2'	m_3'
1	0	21	190	1	10	21
2	35	126	50	6	5	21
3	175	126	120	6	5	7
4	70	21	120	3	10	7
5	105	126	190	2	5	21
6	140	21	50	3	10	21

If m_i' is the smallest positive integer such that $|m_i'\rho_i|_M = 0$ then the residue system with moduli m_1', m_2', m_3' can also have weights ρ_1, ρ_2, ρ_3 and represent Z_M. Such a system also has diagonal carry matrix S.

$$\begin{bmatrix} m_1' & 0 & 0 \\ 0 & m_2' & 0 \\ 0 & 0 & m_3' \end{bmatrix}$$

In particular, if m_1', m_2', m_3' are pairwise relatively prime, then S is identical to K of the system. Otherwise, we will have K that is not diagonal but is triangular.

For the second set of weights from the table $m_1' = 6$, $m_2' = 5$, $m_3' = 21$, and the carry matrix S is given as

$$S = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 21 \end{bmatrix} \quad \text{and det. } S = 360 \text{ .}$$

The null space K has a matrix

$$K = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 5 & 0 \\ -2 & 0 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -7 \\ 0 & 5 & 0 \\ 0 & 0 & 21 \end{bmatrix} \quad \text{and det. } K = 210 \quad ,$$

K is not diagonal but triangular, and the system has carries between the digits which are very much undesirable. This is the case also with the sixth set of weights which have

$$m_1' = 3, \quad m_2' = 10 \quad \text{and} \quad m_3' = 21.$$

But in the case of the weight sets 1, 3, 4 and 5 (from the table) m_1', m_2', and m_3' are pairwise relatively prime and there are no carries at all.

Hence, a non-redundant residue system with 6, 5, 7 as moduli having respective weights 175, 126, 120 can be represented by the redundant system with moduli 6, 10 and 21 with the same weights.

This is why the selection of the set of weights is important from the arithmetic point of view and also decides the type of redundancy.

If the range M is a proper divisor of $< m_1, m_2, \ldots, m_n >$ the situation is not very much different. It can be shown by similar reasoning that there exists a set of weights $\rho_1, \rho_2, \ldots, \rho_n$ for the system which have corresponding integers $\ell_1, \ell_2, \ldots, \ell_n$ satisfying

$$\left.\begin{array}{ll} \text{(i)} & \ell_i \rho_i \equiv 0 \pmod{M} \\ \text{(ii)} & \ell_i \mid m_i \end{array}\right\} \text{ for } i = 1, 2, \ldots, n$$

$$\text{(iii)} \quad \prod_{i=1}^{n} \ell_i = M \, .$$

VII. ERROR CHECKING IN RESIDUE ARITHMETIC

7.1 Introduction

Residue number systems, their properties[1,3,10] and computational methods have been investigated by several researchers.[5,10,11] Some of the persistent problems in residue systems such as magnitude and sign determination, overflow detection and division methods have been studied by them. The reliability aspect of residue arithmetic, and error checking in residue number systems using pairwise relatively prime moduli are discussed by Garner.[7] Some of the coding techniques[8,9,13] of error checking in conventional number systems can be applied with advantage to residue systems. Different methods of residue coding, error checking and their relative advantages are discussed in this chapter.

7.2 Residue Representation

Error checking can be based on the notion of weights* of the coding elements and the distance between two code elements. The representation of the residues will be an important consideration in defining the weight of a code element. If each modulus is considered a single digit, (irrespective of the type of representation of the residues), then the weight of a code element is equal to the number of non-zero residues. In particular, if a residue system in which $(1, 1, \ldots, 1)$ represents the integer 1, the weight of $(1, 1, \ldots, 1)$ is equal to n and so also the distance between

* The weight of a residue number referred here in this chapter is related to the concept of Hamming distance and so must be distinguished from the digit weight used before.

any two adjacent numbers. The distance in this case, between two code elements is the weight of the arithmetic difference between them. This coincides with Hamming distance since there are no carries between the moduli. On the other hand, if the residues are binary coded, it is possible to look at each of the residues, and weight can be attached to them separately. The weight of any particular residue is equal to the number of 1's in it. The distance between any two residues (of modulus m_i) will then be considered as the weight of their difference.

In residue systems, even though there are no carries between moduli, there is difficulty in obtaining the least positive residue (with respect to m_i) anytime the arithmetic result exceeds $m_i - 1$. It will be shown later if the modulus is $2^s - 1$ for any positive integer s , binary coding of the residues will be advantageous.

7.3 Pairwise Relatively Prime Moduli

Consider a residue system N with moduli $m_1, m_2, \ldots, m_{n+k}$ (relatively prime, pairwise) representing integers modulo

$$M = \prod_{i=1}^{n} m_i \qquad (7.1)$$

Let

$$M' = \prod_{i=1}^{n+k} m_i \qquad (7.2)$$

The k moduli $m_{n+1}, \ldots, m_{n+k}$ are used for redundancy. And we shall consider the error checking possibilities of this redundant system with M' elements representing Z_M . Let us examine this as a system representing $Z_{M'}$, non-redundantly and pick the subset representing $\{0, 1, \ldots, M-1\}$ of $Z_{M'}$ as the correct elements of the system and the rest erroneous elements. Then if the system has a minimum weight $t+1$ (or t error detecting capabilities), any element of weight less than t must have a magnitude greater than M.

Also let $\rho_1, \rho_2, \ldots, \rho_{n+k}$ be the weights of each of the moduli. That is $(0, ., \underset{\substack{\uparrow \\ (i\text{-th place})}}{1}, 0, \ldots, 0)$ represents ρ_i. Such that $\rho_i \epsilon Z_{M'}$, for $i=1, 2, \ldots, n+k$. For a non-redundant system, we have that

$$\rho_i = k_i \frac{M'}{m_i} \qquad \text{for all } i=1, 2, \ldots, n+k$$

where k_i is any integer such that $(k_i, m_i) = 1$. If $x = (x_1, x_2, \ldots, x_{n+k}) \epsilon N$ has a weight equal to t, then exactly t of the x's are nonzero. The magnitude of x can then be written as

$$|x|_{M'} = \left| \sum_{j=1}^{t} \delta_j \rho_j' \right|_{M'} \qquad 1 \leq \delta_j \leq m_j' - 1$$

where $\rho_1', \rho_2', \ldots, \rho_t'$ are any combination of t weights out of $n+k$ weights $\rho_1, \rho_2, \ldots, \rho_{n+k}$. Then the condition

$$|x|_{M'} \geq M$$

is sufficient for t error detection, or $\left[\frac{t}{2}\right]$ error correction.

Theorem 13.*

A residue system with $n+k$ moduli $m_1, m_2, \ldots, m_{n+k}$ (all pairwise relatively prime) representing integers modulo

* For the case $k=1, 2,$ the theorem has been proved by H. L. Garner, in some of his unpublished work.

$$M = \prod_{i=1}^{n} m_i$$

has a minimum distance k+1 if and only if $m_{n+j} > m_i$ for

$$j=1, 2, \ldots, d \text{ and } i=1, 2, \ldots, n . \qquad (7.3)$$

Proof: If $m_1, m_2, \ldots, m_{n+k}$ satisfy (7.3), then any element with weight t , $1 \leq t \leq k$ has magnitude X

$$|X|_{M'} = \left| \sum_{j=1}^{t} \delta_j \rho'_j \right|_{M'}$$

$$|X|_{M'} = \left| \sum_{j=1}^{t} \delta_j \frac{M'}{m'_j} \right|_{M'}$$

where $1 \leq \delta_j \leq m'_j - 1$ and $m'_1, m'_2, \ldots, m'_t$ are any t moduli out of the n+k .

$$|X|_{M'} = \frac{C\, M'}{\prod_{j=1}^{t} m'_j} \qquad \text{for } C \quad 1 \leq C \leq \prod_{i=1}^{t} m'_j - 1$$

$$|X|_{M'} \geq \frac{M'}{\prod_{j=1}^{t} m'_j} \geq \frac{M'}{\prod_{j=1}^{t} m_{n+j}} ,$$

since

$$m_{n+j} > m_i \qquad \text{for all } i=1, 2, \ldots, n .$$

Also

$$\frac{M'}{\prod_{j=1}^{t} m_{n+j}} \geq \frac{M'}{\prod_{j=1}^{k} m_{n+j}} = M \qquad \text{since } k > t.$$

Thus, any element of weight t, $1 \leq t \leq k$ has magnitude greater than M and so is not in the code. Thus every non-zero code element has weight greater than k, Conversely, if every non-zero code element has weight greater than k, then

$$\left| \prod_{j=1}^{t} \delta_j \frac{M'}{m_j'} \right|_{M'} \geq M \qquad \text{for all} \quad 1 \leq t \leq k$$

$$\frac{C\ M'}{\prod_{j=1}^{k} m_j'} \geq M \qquad \text{where} \quad 1 \leq C \leq \prod_{j=1}^{k} m_j' - 1$$

Therefore

$$\frac{M'}{\prod_{j=1}^{k} m_j'} \geq M = \frac{M'}{\prod_{j=1}^{k} m_{n+j}}$$

This implies

$$\prod_{j=1}^{k} m_j' \leq \prod_{j=1}^{k} m_{n+j}$$

This is satisfied for all $t \leq k$ and any combination

$$m_1', m_2', \ldots, m_t'$$

only if

$$m_{n+j} > m_i \qquad \text{for} \quad j=1, 2, \ldots, k$$
$$i=1, 2, \ldots, n$$

Thus, the theorem is proved.

If the magnitude of any arithmetic result is checked and is found to be greater than $M-1$ then the arithmetic operation is in error. Error correction is based on the principle of table look up or by trial and error check by procedure. Either of them is inconvenient as they involve repeated magnitude determination which is not simple in residue systems. Thus, the excess moduli coding is good at best for single error detection. Single error detection can be obtained with one extra modulus m_{n+1}, such that $m_{n+1} > m_i$ for $i=1, 2, \ldots, n$.

7.4 Moduli that are not Pairwise Relatively Prime

Error checking in residue systems with relatively non-prime moduli is based on the following

Theorem 14.

If $m_1, m_2, \ldots, m_n$ are moduli of a residue system representing Z_M where $M = < m_1, m_2, \ldots, m_n >$, $(m_i, m_j) = d$ for some $i, j, i \neq j$ and $(1, 1, \ldots, 1) \epsilon N$ represents $1 \epsilon Z_M$, then d divides $x_i - x_j$, where x_i and x_j are residues with respect to the moduli m_i and m_j respectively.

Proof: It was shown in Chapter VI, that if $(1, 1, \ldots, 1)$ is a representation for 1, then any $x \epsilon Z_M$ has a representation

$$(x_1, x_2, \ldots, x_n) \epsilon N$$

such that

$$x_i = |x|_{m_i}$$

For $i \neq j$

$$\left.\begin{aligned} x_i &= x - k_i m_i \\ x_j &= x - k_j m_j \end{aligned}\right\} \text{for some integers } k_i \text{ and } k_j$$

$$\begin{aligned} x_i - x_j &= x - k_i m_i - (x - k_j m_j) \\ &= k_j m_j - k_i m_i \end{aligned}$$

Since $(m_i, m_j) = d$, d divides $k_j m_j - k_i m_i$ and so also $x_i - x_j$. Thus the theorem is proved.

The following example illustrates the error detecting properties of a residue system using $2^s - 1$ type of moduli, which are not pairwise relatively prime. Furthermore, this system uses binary coding of the residues.

Example 11

Let N be a residue system with moduli $m_1 = 63$, $m_2 = 255$, $m_3 = 511$, $m_4 = 1023$. The prime factorization of the moduli will yield the following

$$m_1 = 63 = 3 \times 3 \times 7 = 2^6 - 1$$

$$m_2 = 255 = 3 \times 5 \times 17 = 2^8 - 1$$

$$m_3 = 511 = 7 \times 73 = 2^9 - 1$$

$$m_4 = 1023 = 3 \times 11 \times 31 = 2^{10} - 1.$$

Since the moduli are not pairwise relatively prime, they can represent Z_M when $M = < m_1, m_2, m_3, m_4 > = 63 \times 85 \times 73 \times 341$. If $(x_1, x_2, x_3, x_4) \epsilon N$, then 7 must divide $x_1 - x_3$ and 3 must divide $x_2 - x_4$. If the error is divisible by these factors then it can not be detected. On the otherhand if the residues are coded in the binary form, a single error in one of the residues causes a change of $\pm(bm_i - 2^k)$ where b=o or 1 , and k is any non-negative integer such that $2^k < m_i$. Some simpler methods of obtaining residues modulo 3, 7 or $2^k - 1$ (for some integer k) are covered in the next section.

7.5 Binary Coded Residue Systems

Since the residues are coded binary, conventional binary arithmetic units can be used with certain modifications. Whenever the result of an arithmetic operation produces a residue value greater than $m_i -1$, the least positive residue $(\text{mod } m_i)$ has to be recovered. This can be done by a division of the result by m_i and the remainder be taken as the residue. That is a very cumbersome method. On the other hand, residue addition and recovery of the least positive residue can be done by generation of suitable number of end-around carries. The idea of end-around carries is based on the following. If a certain modulus m_i is a power of 2, say $m_i = 2^k$ then arithmetic modulo m_i is done by a k bit binary unit with overflows ignored. Otherwise, m_i is such that $2^k > m_i > 2^{k-1}$, for some positive integer k let $2^k - m_i$ be equal to C. Then $C < 2^{k-1}$, and $2^k \equiv C \pmod{m_i}$. Thus an overflow from the k-th stage, (equivalent to 2^k) can be taken care of by addition of C. If C in its binary form, has only a few 1's then these can be absorbed as end-around carries. The result obtained by this technique is $< 2^k$, and is the correct value if it is $< m_i$, otherwise m_i should be subtracted. This method can be employed for multiplication using end-around carries absorption, comparison with m_i being left to the end.

If $m_i = 2^k-1$, there is only one end-around carry. Multiplication by 2 or by any power of 2 is obtained by a suitable number of cyclic left shifts. Also, since $m_i = 2^k-1$ is expressed as 11...1 in binary, the complement of any residue is obtained by switching zeros into ones and vice versa. If a set of moduli $m_1, m_2, \ldots, m_n$ is chosen, in which $m_i = 2^{s_i}-1$, $(i=1, 2, \ldots, n;)$ then the moduli may not be relatively prime.

The example 11 in the previous section uses moduli of the type $2^{s_i}-1$ and the redundancy factors are 3 and 7. If the residues are binary coded then error detection proceeds as follows.

Single errors in binary residue arithmetic would result in an error of $\pm(bm_i - 2^k)$ where $b = 0$ or 1 and k any non-negative integer such that $2^k < m_i$. Since 3 divides m_i but does not divide 2^k the error is detectable. The same thing can be said about the moduli that have 7 as a common factor. Thus, single errors can be detected by verifying whether

$$3 \text{ divides } x_2 - x_4 \text{ and } 7 \text{ divides } x_1 - x_3 .$$

To check whether 3 divides any binary number several methods exist. One method is to delete all sets of two adjacent 1's or 0's and group the rest of the number and do it over again until no two adjacent zeros or ones exist. The residue modulo 3 will be equal to the number of 1's if they are in odd places, or will be equal to -(the number of one's) if they are in even places.

Another method is to use a modulo three adder-subtractor to add the odd digit 1's and subtract the even digit 1's.

Residue modulo 2^k-1 of a binary number n digits long can be obtained by treating the number in groups of K digit long and adding the $[\frac{n}{k}] + 1$, k bit numbers with an end-around carry. This is possible because

$$2^k \equiv 1 \bmod (2^k - 1)$$

$$2^{ak} \equiv 1 \bmod (2^k - 1)$$

Also k adjacent 1's or 0's forming part of the number can be deleted, the remaining ones joined.

An example of a binary number modulo 7 is given below.

Example 12

To obtain $X = |010\ 111\ 001\ 011\ 011|_7$

Deleting three adjacent 1's we have

010 001 011 011 .

Again deleting the three adjacent zeros formed, we have

011 011 011 .

Now dividing into three bit numbers, and adding them, we have

$$\begin{array}{r} 011 \\ \underline{011} \\ 110 \\ \underline{011} \\ 1001 \end{array}$$

An end-around carry has to be generated. Therefore the result is

$$\begin{array}{r} 001 \\ \underline{1} \\ 010 \end{array}$$

Therefore, $X = 2$.

A residue system with moduli $3m_1, 3m_2, \ldots, 3m_n$, where $m_1, m_2, \ldots, m_n$ are pairwise relatively prime, permits single error detection in the residues. In fact, simultaneous detection of single errors in t of the moduli is possible where t is any integer such that $t < \frac{n-1}{2}$. Also, the exact moduli in which the errors occurred can be located. If $(x_1, x_2, \ldots, x_n)$ is the arithmetic result from Theorem 7, we have that all the residues $|x_1|_3, |x_2|_3, \ldots, |x_n|_3$ should be equal. If these are single errors in any of the residues, they can be detected. If there are t moduli in which single errors have occurred, corresponding to these moduli, the residues modulo 3 would be different. Since some of the moduli of the $2^s - 1$ type have 3 as a factor, this method can be considered advantageous. Redundancy of the system can be expressed as

$$\text{Information per bit} = \frac{\log_2 \left[3 \prod_{i=1}^{n} m_i \right]}{\log_2 \left[3^n \prod_{i=1}^{n} m_i \right]},$$

which will be greater than the single error detecting system using relatively prime moduli, if any of the moduli is greater than 3^{n-1}.

7.6 An + B Type Coding of Residues

Error checking of the residues is possible by coding each of the residues separately. This is based on the principle of $An + B$ type codes.[9] If $Am_i + 2B = 2^{s_i} - 1$ for any particular modulus m_i, and for positive integers A, B and s_i, then an s_i bit binary representation of the residues is possible. Since k and its complement $m_i - k$ are coded as $A k + B$ and $A(m_i - k) + B$ respectively, their sum

$$A k + B + A(m_i - k) + B = Am_i + 2B = 2^{s_i} - 1,$$

which is expressed as 11...1 in the binary form. Complementation can be done by switching 0's and 1's. However for $B \neq 0$, the addition and subtraction of two code elements should be accompanied by proper correction.

This is not suitable for multiplication of two residues in the coded form. However, for codes $Am_i = 2^{s_i} - 1$ (that is for $B = 0$) no correction will be necessary for addition or subtraction. If k_1 and k_2 are two residues coded as $A\,k_1$ and $A\,k_2$, their multiplication will have to be $|A\,k_1\,k_2|_{Am_i}$. This can be obtained by multiplying $A\,k_1$ by k_2 or Ak_2 by k_1. That is, one of the operands have to be decoded before multiplication. Also, the minimum distance or weight of the code elements depends on the selection of A. For single error detection A can be any odd integer ≥ 3. For single error correction the minimum distance has to be ≥ 3. For each odd integer A there exists integer r_{max}, such that $A\,r_{max}$ is of the form $2^t \pm 1$ for smallest integer t. Then $m_i \leq r_{max}$, since Am_i is required to be of the form $2^{s_i} \pm 1$, $m_i = r_{max}$. Some values of A and m_i are given in the table below.

A	m_i	$Am_i = 2^{s_i} \pm 1$
19	27	$2^9 + 1$
21	3	$2^6 - 1$
23	89	$2^{11} - 1$
29	565	$2^{14} + 1$
37	3085	$2^{18} + 1$
39	105	$2^{12} - 1$
91	45	$2^{12} - 1$
99	331	$2^{15} + 1$
105	39	$2^{12} - 1$

If $Am_i = 2^{s_i} + 1$ type, the arithmetic is not so straightforward as in $2^{s_i} - 1$ type. An end-around borrow will have to be propagated in $2^{s_i} + 1$ type. Also, complementing a code element can be done by switching 0's and 1's followed by an addition of 2.

7.7 Suitable Moduli for Residue Computation

The single error detecting system using moduli 63, 255, 1023, and 511 has several advantages as shown before. The redundancy per bit in the system is:

$$\frac{\log_2 63}{\log_2[63 \times 255 \times 1023 \times 511]}$$

$$\approx \frac{6}{6+8+9+11} = .176 = 17.6 \text{ percent,}$$

and the information per bit = 1 - .176 = .824 (approximately).

An n-single error correcting system using moduli 89, 117, 565, and 331 which are all pairwise relatively prime has a range $M = \Pi m_i$, of the order of 2^{31}. The corresponding coding factors A_i and their products are as below.

i	A_i	m_i	$A_i m_i$
1	23	89	$2^{11} - 1$
2	35	117	$2^{12} - 1$
3	29	565	$2^{14} + 1$
4	99	331	$2^{15} + 1$

Any single error in each of the residues of the arithmetic result (x_1, x_2, x_3, x_4) can be corrected by obtaining $|x_i|_{A_i}$. If the result is correct,

$$|x_i|_{A_i}$$

would be all 0 . Since any single error causes a change of $\pm 2^k$, $k \leq s_i$ for the $2^{s_i} + 1$ type, $k < s_i$ for the $2^{s_i} - 1$ type of moduli, and their residues modulo A_i are all distinct; the single error can be corrected. This system enables n-single error correction (or a single error correction in each of the n moduli). As expected, the redundancy per bit given by r

$$r = \frac{\log_2[23 \times 35 \times 29 \times 99]}{\log_2[(2^{11} - 1)(2^{12} - 1)(2^{14} + 1)(2^{15} + 1)]}$$

$$= \frac{14.66}{52 \times .694} = 40.6 \text{ per cent}$$

is much larger than 18.2 percent of the single error detecting $2^s - 1$ type moduli system described before. Further, this system has $2^s + 1$ type moduli which are not as convenient as $2^s - 1$ type. Also, since the coding factors A_i are 23, 35, 29, 99, there is no simpler way of obtaining residues modulo A_i other than by division. These features make the n-single error correcting system less attractive.

VIII. CONCLUSION

8.1 Review of the Results and Conclusion

In the first part we are able to categorize the finite number systems as linear and non-linear and study their advantages and disadvantages. It is shown that the digitwise sum has no meaning in non-linear systems. In particular, the weighted systems, which are in the category of linear homogeneous, obey the digitwise sum rule (2.2) as stated in Chapter II. Very important consideration is given to the relations between the digit weights and carry propagation rules. These have been explained very successfully by means of the quotient module structure that can be given for all weighted systems. This leads to the interesting notion of triangular form of carry matrix for non-redundant systems and the theory of canonical transformations for redundant systems as explained in Chapter V.

For the residue number systems, it is shown that the carry matrix is diagonal, and the range M is a divisor of the product of the moduli. This condition limits the choice of redundancy we can use in the residue systems. Also, since there are d sets of acceptable digit weights in a redundant residue system representing Z_M where $M = < m_1, m_2, \ldots, m_n >$ and $d = \frac{\Pi m_i}{M}$, computation can be done using any suitable set of weights. However, the selection of the set of weights is dependent upon the error checking scheme of the system. This dependency of error checking and the digit weights is explained by means of the example of a residue system with three moduli 6, 10, and 21.

Using the theory of redundant residue systems, methods of error checking in residue arithmetic are derived. Moduli of the type $2^s \pm 1$ (for a positive integer s) and their advantages in residue computation are investigated.

The aspect of selecting suitable moduli and the type of redundancy for reliable and logically superior residue arithmetic is one of great importance. Other problems in residue computation are to find improved methods of magnitude comparison, sign detection, and division. There are attempts by some researchers,[10,11] to use redundancy to obtain improved sign determination methods. Unless some breakthroughs are obtained in these problems, the residue computer still remains as a special purpose machine. While the abstract mathematical structure described in the earlier parts is expected to enhance the understanding of the general properties of the weighted systems, the investigation presented in the latter parts of this dissertation is expected to help in the logical design of reliable and improved residue arithmetic units.

APPENDIX

Theorem 3:

The n independent linear congruences expressed below as

$$\begin{bmatrix} c_{11} & c_{12} & \cdot & c_{1n} \\ c_{21} & c_{22} & \cdot & c_{2n} \\ \cdot & & & \\ c_{n1} & c_{n2} & \cdot & c_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \\ x_n \end{bmatrix} \equiv \begin{bmatrix} 0 \\ 0 \\ \\ 0 \end{bmatrix} \pmod{M}$$

have solutions $x_i = P_i$ where $(P_1, P_2, \ldots, P_n, M) = 1$ if only M divides the determinant of the nxn matrix above.

Proof: The congruences can be written as n equations

$$c_{11}\, x_1 + c_{12}\, x_2 + \ldots + c_{1n}\, x_n = k_1\, M$$

$$c_{22}\, x_1 + c_{22}\, x_2 + \ldots + c_{2n}\, x_n = k_2\, M$$

$$c_{n1}\, x_1 + c_{n2}\, x_2 + \ldots + c_{nn}\, x_n = k_n\, M$$

Let

$$\Delta = \begin{vmatrix} c_{11} & c_{12} & \cdot & c_{1n} \\ c_{21} & c_{22} & \cdot & c_{2n} \\ \cdot & & & \\ c_{n1} & c_{n2} & \cdot & c_{nn} \end{vmatrix}$$

and a minor Δ_{ij} of Δ be the $(n-1)$ by $(n-1)$ determinant obtained by deleting the i-th row and j-th column from Δ.

Then

$$x_1 = \frac{k_1 M \Delta_{11} + k_2 M \Delta_{21} + \dots + k_n M \Delta_{n1}}{\Delta}$$

$$= \frac{M C_1}{\Delta} \text{ for some integer } c_1 .$$

Similarly,

$$x_i = (-1)^{i-1} \frac{k_1 M \Delta_{1i} + k_2 M \Delta_{2i} + \dots + k_n M \Delta_{ni}}{\Delta}$$

$$= \frac{M C_i}{\Delta} \text{ for some integer } c_i .$$

Thus, we have

$$x_i \Delta = M C_i \quad \text{for } i = 1,2,\dots,n .$$

Let

$$(x_1,x_2,\dots,x_n) = k .$$

Since

$$(x_1,x_2,\dots,x_n,M) = 1$$

$$(k,M) = 1$$

$$(x_1 \Delta, x_2 \Delta, \dots, x_n \Delta) = k \Delta .$$

$$(M C_1, M C_2, \dots, M C_n) = M C \quad \text{for some integer } C .$$

Therefore,

$$k \Delta = M C .$$

Therefore M divides $k \Delta$.

Since $(M,k) = 1$, M divides Δ, thus proving the theorem.

BIBLIOGRAPHY

1. Garner, H. L., "The Residue Number System," IRE Trans. on EC, Vol. EC-8, No. 2, June 1959.

2. LeVeque, W., Theory of Numbers, Vol. 1, Addison Wesley Publishing Co. 1956.

3. Rozenberg, D. P., "Algebraic Properties of Residue Number Systems," IBM 61-907-176.

4. Jacobson, N., Lectures in Abstract Algebra, Vol. 2, Chapter 3, Van Nostrand Co., Inc., 1952.

5. Garner, H. L., et al., "Residue Number Systems for Computers," ASD Technical Report 61-483, The University of Michigan Technical Note, ORA 04879-6-T, September 1962.

6. Garner, H. L., "Finite Non-Redundant Number System Weights," Information Systems Laboratory, The University of Michigan Technical Note, ORA 04879-6-T, September 1962.

7. Garner, H. L., "Error Checking and the Structure of Binary Addition," Ph.D. Thesis, Chapter V, The University of Michigan, 1958.

8. Diamond, J. M., "Checking Codes for Digital Computers," Proc. of the IRE, 43, (1955) 457-488.

9. Brown, D. T., "Error Detecting and Correcting Binary Codes for Arithmetic Operations," IRE TRANS. EC-9, (1960) 333-337.

10. Aiken, H., et al., "Modular Number Systems," Harvard University Computational Laboratory, July 1960.

11. "Modular Arithmetic Techniques," Technical Documentary Report No. ASD-TDR-62-686, January 1963, Lockheed Missiles and Space Co., Sunnyvale, California.

12. Arnold, R. F., "Linear Number Systems," The University of Michigan, Technical Note 04879-8-T, October 1962.

13. Peterson, W. W., "Error Correcting Codes," The MIT Press and John Wiley & Sons, Inc., New York, (Jan. 1961) 236-244.