ON SOME APPLICATIONS OF ALGEBRA AND GEOMETRY IN CODING THEORY.

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTERS DEGREE OF THE UNIVERSITY OF NAIROBI IN PURE MATHEMATICS.

BEN OBIERO

August 16, 2011

Declarations

I the undersigned declare that this dissertation is my original work and to the best of my knowledge has not been presented for the award of a degree in any other University/College.

Ben Obiero

I56/78951/2009

Signature _____ Date 17/08/2011

Declaration by Supervisor

This project has been submitted for examination with my approval as supervisor

Mw. Achola Claudio

Signature Achola Date 19/08/2011

dedication

To Jerry my dearest son.

Acknowledgement

I would wish to express my sencerest appreciations; first to God almighty without whom nothing hold. To Mw. Achola who has been more than a guardian to me throughout my life as a student, right from the time I was an undergraduate through to graduate school. In you I see a Mentor! To Prof. Pokhariyal the "father" of Geometry, you widenned my perspective of Pure Mathematics. Your ever open doors I cannot fail to point out!

It is also my honour to recognize E.A.U.M.P, and A.M.M.S.I, particularly Prof. Weke and Prof. Ogana. I couldn't have had the chance of writting this dissertation at this particular time if not for you; the late Prof. Owino, I am sure this is what you expected of me after two years in graduate school.

Long live E.A.U.M.P and I.C.T.P your summer schools have had impact in my studies and part of the motivation towards this dessertation was drawn from them. To Mr. Ongaro, Mr. Kikwai, Dr. Nkubi and Ms. Tanui F.J. This project is one of the products of your "inhuman" nature to share whatever knowledge you have with anyone who is willing to learn. Your efforts are bearing fruits. Tire not!

And finally to my parents Mr. and Mrs. Obiero, my brothers and friends; Sam, Dan, Zachary, Michael, Shem and Mark; and also to my classmates Beth, Isaiah, David and Amos, your support both moral and financial is highly appreciated.

Abstract

The fundamental problem that led to the development of the theory of error correcting codes was that of having a reliable communication over unreliable channels. A communication channel can be as simple as the air between the voicebox of one adressing an audience and the ears of the listeners. Copper wires connecting telephones or modems can also be considered as channels. In the case of data storage, say in a magnetic tape or disc, the magnetized field in the magnetic tape or disc is the channel.

One property of these channels is their capacity to distort the information. For instance, the copper wires connecting telephones may get heated resulting in background interruptions. Magnetisation on the tape may re-align over time, or the head of the drive reading the tape or disc may be ill positioned and the right magnetisation be misread! Such distotortions to our good information will be referred to as *noise*.

There are two main ways of handling noise; physical means and system means. Under physical means, one "targets" the cause of noise and seeks to eliminate it. For instance if the drive head misreads the magnetisation, then a better drive is used as a replacement for the "now faulty drive". As for system means, one "sandwiches" the channel between two devices; an encoder and a decoder (see figure 0.1) so that any form of noise can be detected and possibly corrected. The theory of error correcting codes is involved in this.

For our purpose we will consider an abstract communication channel called the Binary Symmetric Channel (BSC). In BSC, information to be transmitted is encoded as a string of 0's and 1's. An error is then considered as an interchange between the binary digits in the sent and received information symbols.

fig.0.1 A simple Communication Channel.

Below is a diagramatic representation of a binary symmetric channel.

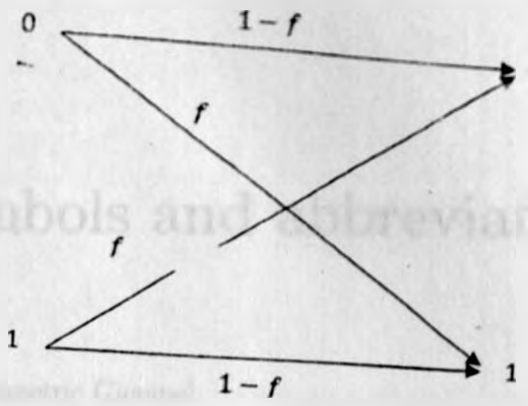

fig.0.2 A Model of BSC.

Until 1948, encoding with minimal error was done at the expense of information rates. Information symbols to be transmitted were repeated several times with more efficiency gained by higher order repetitions. If n, say is the order of repetition, then each bit represented $1/n$ of the information and this value aproaches zero as the order of repetition becomes very large. Shanons ground breaking work [1] created a platform for the launch of error correcting codes (see history of coding in the introduction). Since then, various mathematical disciplines have lead to the development of this theory. In this Dissertation, we will consider the "contributions" from Algebra and Geometry in Coding Theory.

List of Symbols and abbreviations

- BSC:- Binary Symmetric Channel.
- BCH code:-Bose Chaudhuri and Hocquenghem code.
- Bit:-**Bi**nary digit.
- RS code:-Reed-Solomon code.
- $\mathbb{F}_q[x_1 \cdots, x_n]$:-the ring of polynomials over the field $\mathbb{F}$. in indeterminates $x_1 \cdots, x_n$.
- $(n, k, d)-$code:-A code of length n, dimension k and distance d.
- $\mathbb{A}^n : -$ An $n-$affine space.
- $\mathbb{P}^n : -$ A projective space.
- GRS code:- The Generalised Reed-Solomon code.
- $MDS-$codes:-Maximum distance separable codes.

Contents

Chapter 1

Introduction

Information revolution is in full swing today. The number of web pages on the computers connected to the internet runs from hundreds of millions to billions. With all these terabits per second flying around the world, reliability of the networks becomes a major concern.

Loosing just 0.001% of data on a network link whose capacity is one gigabit per second amounts to loss of 10000 bits per second. On average, a typical newspaper story has 1000 words which can be estimated to be 42000 bits. Thus such a loss as above can be equivalent to loosing one newspaper story every four seconds, or put another way, loosing 900 newspaper stories every hour! Even a small error rate becomes impractical as data rates increases to statospheric levels as they have over time.

It is a fact that all networks corrupt the data sent through them. The question is, can we find a means of ensuring that good data gets through poor networks intact? Coding theory seeks to adress this problem. The major concerns of coding theory are the construction of efficient coding schemes that are capable of:

1. Correcting a relatively large number of errors.
2. Achieving a relatively high rate[1] of information transmission.
3. Attaining relatively simple and economical procedures to encode and decode the messages.

To get a deeper understanding of these ideas we need to abstract a communication system. A block diagram of a simple communication system employing an error-correcting code is shown below.

[1] the concept of rate is explained in the next chapter

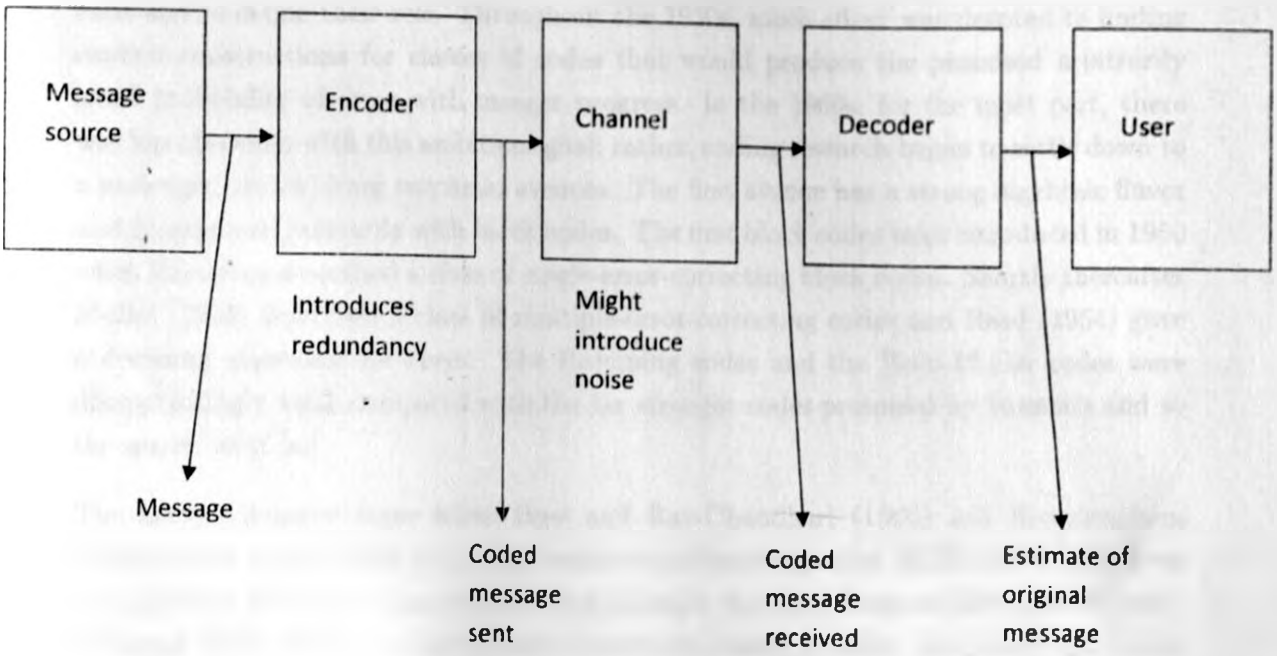

fig.1.1 A Model Channel that uses Error-Correcting Codes.

1.1 A Brief History of Coding

The Ground Breaking Work By Shannon and Its Signifficance

The history of error-correcting codes began in 1948 with the publication of a famous paper by Claude Shannon. Shannon showed that associated with any communication channel or storage channel is a number C (measured in bits per second), called the *capacity*[2] of the channel. The significance of this association is that, whenever the information transmission rate R (in bits per second)of a system is less than C then, by using an error-correcting code, it is possible to design a communication system for the channel whose probability of output error is as small as desired. In fact, an important conclusion from Shannon's theory of information is that it is cheaper and ultimately more effective to use a powerful erro-correcting code at the terminal device of communication system, instead of making a communication channel error-free.

[2]the maximum number of information symbol a channel can convey per unit time.

Advances Over the Years

Shanon gave no idea on how to find suitable codes; his contribution was to prove that they exist and to define their role. Throughout the 1950s, much effort was devoted to finding explicit constructions for classes of codes that would produce the promised arbitrarily small probability of error with meager progress. In the 1960s, for the most part, there was less obsession with this ambitious goal; rather, coding research began to settle down to a prolonged attack along two main avenues. The first avenue has a strong algebraic flavor and is concerned primarily with block codes. The first block codes were introduced in 1950 when Hamming described a class of single-error-correcting block codes. Shortly thereafter Muller (1954) described a class of multiple-error-correcting codes and Reed (1954) gave a decoding algorithm for them. The Hamming codes and the Reed-Muller codes were disappointingly weak compared with the far stronger codes promised by Shannon and so the search went on!

The major advances came when Bose and Ray-Chaudhuri (1960) and Hocquenghem (1959) found a large class of multiple error-correcting codes (the BCH codes), and Reed and Solomon also discovering another class of codes; the Reed Solomon(RS) Codes (1960). Although these remain among the most important classes of codes, the theory of the subject since that time has been greatly strengthened, and new codes continue to be discovered. The discovery of BCH codes led to a search for practical methods of designing the hardware or software to implement the encoder and decoder. The first good algorithm was found by Peterson (1960). Later, a powerful algorithm for decoding was discovered by Berlekamp (1968) and Massey (1969), and its implementation became practical as new digital technology became available. Now many varieties of algorithms are available to fit different codes and different applications.

The second avenue of coding research is more probabilistic in nature. Early research was concerned with estimating the error probability for the best family of block codes despite the fact that the best codes were not known. Associated with these studies were attempts to understand encoding and decoding from a probabilistic point of view, and these attempts led to the notion of sequential decoding. Sequential decoding required the introduction of a class of nonblock codes of indefinite length, which can be represented by a tree and can be decoded by algorithms for searching the tree, hence *the convolutional* codes.

In 1970's, Goppa defined a class of codes that is sure to contain good codes, though without saying how to identify the good ones. The 1980s saw encoders and decoders appear frequently in newly designed digital communication systems and digital storage systems. A visible example is the *compact disk*, which uses a simple Reed-Solomon code for correcting double byte errors. Reed-Solomon codes also appear frequently in many *magnetic tape drives* and *network modems*, and now in *digital video disks*. Meanwhile, mathematicians took the search for good codes based on the Hamming distance into

the subject of algebraic geometry and there started a new wave of theoretical progress that continues to grow. Algorithms for decoding of large nonbinary block codes defined on algebraic curves have been explored. In particular, decoders for the codes known as *hermitian* codes are now available and these codes may soon appear in commercial products. At the same time, the roots of the subject are growing even deeper into the rich soil of mathematics.

1.2 Objective

In this Dissertation, I seek to;

- Survey the Theory of Error Correcting Codes along two main fronts;
  - Algebra.
  - Geometry.
- Establish if any the interrelationship between the codes discussed.

The probabilistic account will not be tackled and any interested reader is referred to [19].

In the next chapter, I discuss some Algebraic concepts which will prove necessary for the introduction of the notion of codes. I then use these concepts to introduce the notion of codes, in Chapter 3 giving simple examples. In Chapter 4, I give a summary of some results in Algebraic Geometry necessary for code description. I do this carefully to capture the transition between "Algebraic" coding and "Geometric" coding. As a matter of fact it will be seen that some of the most important codes that were earlier discussed from the point of view of Algebra are actually instances of Geometric codes! As a conclussion to the dissertation I try to establish any possible inclussion relationship among the classes of codes that are featured in the write-up.

Chapter 2

Some Algebraic Concepts

2.1 Algebraic Systems

In order to fully comprehend the theory of error correcting codes, we need an understanding of structure of some sort which will allow for construction of codes as well as enable the practical instrumentation of the codes. We start by considering a well defined collection S of objects. Such a collection will be called *a set*. On this set, we define a rule that assign to any pair of elements, an element in S and call it *a binary operation*. A binary operation is normally denoted by symbols like $*$, $+$, $\cdot$, $\oplus$, $\otimes$, e.t.c. If for any pair of elements, $a, b \in S$, a binary operation $*$ is such that;

$$a * b = b * a \tag{2.1.1}$$

then $*$ is said to be *commutative* otherwise $*$ is *non-commutative*. For the case when $a * (b * c) = (a * b) * c$ for all $a, b, c \in S$ $*$, is said to be *associative*. A set together with atleast a binary operation is called *an algebraic system*.

2.2 Groups and subgroups

The simplest algebraic system comsist of a set S together with an associative binary operation. Such a system is referred to as a *semigroup*.

Definition 2.2.0.1 (Group). *A **Group** is a set G together with a binary operation $*$ defined on it and is such that the following conditions hold:*

i). the binary operation $*$ is associative.

*ii). G contains a unique element e, such that for any element a in G. $a * e = e * a = a$. This element is called the **identity element** of the group G.*

iii). For any element a in G there exists a unique element a' in G such that;

$$a * a' = a'a = e$$

the element a' is the ***inverse*** *of a and vice versa.*

If in addition equation 2.1.1 holds in G, then G is a called *commutative* or *abelian* group. A group with finite number of elements is said to be finite.

Notation 2.2.0.2. *It is sometimes written $(G, *)$ to mean a group G together with the binary operation $*$.*

Definition 2.2.0.3. *A multiplicative group G is said to be* **cyclic** *if there is an element of $a \in G$ which generates the group. In such a case any element say $b \in G$ is expressible as $b = a^j$ for some integer j. We call the element $a \in G$ the generator of G and denote $G = \langle a \rangle$*

Definition 2.2.0.4 (Equivalence Relation). *A subset R of $S \times S$ is called* ***an equivalence relation*** *on a set S if it has the following properties;*

a). $(s, s) \in R \quad \forall s \in S$ *(reflexivity)*

b). $(s, t) \in R \Rightarrow (t, s) \in R$ *(symmetry)*

c). $(s, t) \in R \quad (t, u) \in R \Rightarrow (s, u) \in R$ *(trensitivity)*

Equivalence relation R on a set S pattitions S i.e. the relation induces a representation of S as a union of non-empty mutually disjoint subsets of S.
the collection of all elements of S equivalent to s forms an *equivalence class* of s, denoted by $[s] = \{t \in S \mid (s, t) \in R\}$

Definition 2.2.0.5. *For any arbitrary integers a and b and a positive integer n we say that a is congruent to b modulo n and write $a \equiv b (mod \quad n)$, if the difference $a - b$ is a multiple of n. i.e. $a = b + kn$ for some integer k.*

Congruence mobulo n partitions the set of integers into the follwing equivalence classes.

$$\begin{aligned} [0] &= (\cdots, -2n, -n, 0, n, 2n, \cdots) \\ [1] &= (\cdots, -2n+1, -n+1, 1, n+1, 2n+1, \cdots) \\ &\vdots \\ [n-1] &= (\cdots, -2n-1, -n-1, -1, n-1, 2n-1, \cdots) \end{aligned} \tag{2.2.1}$$

We can define on the set $R = \{[0], [1], \cdots, [n-1]\}$ a binary operation '$+$' by;

$$[a] + [b] = [a + b]$$

Then the set R together with the binary operation forms a group.

Subgroups

Definition 2.2.0.6. *If a subset H of a group G forms a group with respect to the binary operation of G, then H is called a* ***subgroup*** *of G.*

Definition 2.2.0.7. *Let $\langle G, * \rangle$ be a group and $\langle H, * \rangle$ a subgroup of G. Let $a \in G$. Then the set of elements $a * H := \{a * h \mid h \in H\}$ is called the* ***left coset*** *of H determined by $a \in G$. We can immitate this construction for the right coset.*

Clearly, for a commutative case, the right and left cosets are identical, and since we are primarily concerned with commutative groups, no distinctions will be made. Coset decomposition partitions a Group into distinct orbits.

2.2.1 Homomorphism

If $\langle G, * \rangle$ and $\langle G', \cdot \rangle$ are two groups, then a map

$$f : G \to G'$$

is called a *homomorphism* of G into G' if $\forall a, b \in G$

$$f(a * b) = f(a) \cdot f(b)$$

If in addition, f is onto then f is called an *epimorphism.*
A one-to-one homomorphism of G onto G' is called an *isomorphism* of G onto G'. In this case G and G' are isomorphic. We call f an *automorphism* if $G = G'$

2.2.2 Rings and Fields

Definition 2.2.2.1. *A* ***ring*** *is a set R with two binary operations namely; "addition" denoted by $+$ and "multiplication" denoted by $\cdot$, such that the following properties hold:*

i). $\langle R, + \rangle$ is an abelian group.

ii). $\langle R, \cdot \rangle$ is a semigroup.

iii). multiplication is distributive over addition.

Such a ring is sometimes denoted by $\langle R, +, \cdot \rangle$.

If $a \cdot b = b \cdot a \quad \forall a, b \in R$ then R is a *commutative ring.* Similarly we say that R is a *ring with identity* if it has a multiplicative identity.

Definition 2.2.2.2. *A commutative ring with identity each of whose non-zero elements has a multiplicative inverse is called a* ***field.***

A non-commutative ring each of whose non-zero elements has a multiplicative inverse is called a *division ring* or a *skew field.* The minimum number of elements a field can ever have is two, since the field must have the additive identity and multiplicative identity. The addition and multiplication table of such a field is as below:

Table 2.1: Addition table for a field with two elements

+	0	1
0	0	1
1	1	0

Table 2.2: Multiplication table for a field with two elements

·	0	1
0	0	0
1	0	1

2.2.3 Vector Spaces

Definition 2.2.3.1. *Let $\mathbb{F}$ be a field. A vector space over $\mathbb{F}$ is a system $\langle V, +, \cdot \rangle$ in which the following axioms are valid.*

1). $\langle V, + \rangle$ is an abelian group.

2). For any $a \in \mathbb{F}$ and an element $v \in V$, $a \cdot v$ is an element of V.

3). For an elements $u, v \in V$, and $a, b \in \mathbb{F}$
$a \cdot (u + v) = a \cdot u + a \cdot v$
$(a + b) \cdot v = a \cdot v + b \cdot v$

4). For any $v \in V$ and any $a, b \in \mathbb{F}$

$$(a \cdot b) \cdot v = a \cdot (b \cdot v)$$

where $a \cdot b$ is the usual product of two scalars and $b \cdot v$ is the multiplication of the vector v by a scalar b according to the composition laws of the vector space.

5). Let 1 be the unit element of $\mathbb{F}$ then for for any $v \in V$, $1 \cdot v = v$

Definition 2.2.3.2. *A subset S of V is called a* ***vector subspace*** *of V if $\langle S, +, \cdot \rangle$ is a vector space.*

Theorem 2.2.3.3. *The set of all linear combinations of a set of vectors $\{v_1, \cdots, v_k\}$ of a vector space V is a subspace of V.*

Definition 2.2.3.4. *A set $S = \{v_1, \cdots, v_k\}$ in a vector space V is said to be* ***linearly independent*** *if and only if $\sum_{i=1}^{k} c_i v_i = 0$ implies that all c_i's are all zeros.*

Definition 2.2.3.5. *If the set $S = \{v_1, \cdots, v_k\}$ generates a vector space V, then the set is said to* ***span*** *V and we write $V = span\,(v_1, \cdots, v_k)$.*
Such a set S if linearly independent is called a ***basis*** *for V, and the number of vectors in S is called the* ***dimension*** *of the vector space V.*

Theorem 2.2.3.6. *If a set of k vectors $v_1, \cdots, v_k$ spans a vector space that contains a set of m linearly independent vectors $u_1, \cdots, u_m$, then $k \geq m$.*

Theorem 2.2.3.7. *If two sets of linearly independent vectors span the same space, then there are the same number of vectors in each set.*

Theorem 2.2.3.8. *If V is a $k-$dimensional vector space, then any set of k linearly independent vectors in V is a basis for V.*

Remark 2.2.3.9. *Theorem 2.2.3.8 above points to the fact that the basis of a vector space is not unique, and any two bases have same cardinality.*
Notice also that in light of theorems2.2.3.6, and 2.2.3.7, we can view a basis of a vector space as a ***maximal*** *linearly independent set in V, or as a* ***minimal*** *spanning set of V.*

Definition 2.2.3.10. *An* ***inner product*** *or* ***dot product*** *of two $n-$tuples is a field element and is defined as follows:*

$$(a_1, a_2, \cdots, a_n) \cdot (b_1, b_2, \cdots, b_n) = (a_1 b_1 + a_2 b_2 + \cdots + a_n b_n)$$

Such a relation is commutative with $\cdot$ distributive over $+$. If the inner product of two vectors is zero, then the vectors are said to be ***orthogonal****.*

Theorem 2.2.3.11. *The set of all $n-$tuples orthogonal to a subspace V_1 of $n-$tuples forms a subspace V_2 of $n-$tuples.*

The subspace V_2 in Theorem 2.2.3.11 is called the *null space* of V_1
Any vector which is orthogonal to every vector of a set which spans V is in the null space of V.

2.2.4 Matrices

An $m \times n$ matrix is an ordered set of mn elements in a rectangular array of m rows and n columns:

$$(a_{i,j}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Generally the elements a_{ij} of a matrix may be considered to be elements of any ring, but for our purpose we will consider the matrices with elements in a field.

Definition 2.2.4.1. *The set of all linear combinations of rows (respectively columns) of a matrix is the* ***row space*** *(respectively* ***column space****) of the matrix. The dimension of the row space always equals the dimension of the column space. This dimension is called the* ***rank*** *of the matrix*

For any matrix, we can perform the following operations;

1. interchange any two rows.

2. multiply a row by a non-zero field element.

3. add a scalar multiple of one row to another row.

The process can be carried out so that the resulting matrix has a zero below every leading entry. We can further make each leading entry to have a zero above it as well and make the leading entry to be unity. A matrix of this form will be said to be in *standard form.*

Theorem 2.2.4.2. *If one matrix is obtained from another by a succession of operations* $1, 2$ *and* 3 *above, then both matrices have the same row space.*

The *transpose* of an $m \times n$ matrix M is an $n \times m$ matrix, denoted by M^T, whose rows are the columns of M, and thus whose columns are the rows of M.

Two $m \times n$ matrices can be added, element by element:

$$(a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$$

With this definition it is easily verified that matrices form an Abelian group under addition.

An $n \times k$ matrix (a_{lj}) and a $k \times m$ matrix (b_{ij}) can be multiplied to give an $n \times m$ product matrix (p_{ij}) by the rule;

$$p_{ij} = \sum_{l=1}^{k} a_{il} b_{ij}$$

The null space of the row space of a matrix is called the *null space* of the matrix. A vector is in the null space of a matrix if it is orthogonal to each row of the matrix. If the $n-$tuple v is considered to be a $1 \times n$ matrix, v is in the null space of an $m \times n$ matrix M if and only if $vM^T = 0$.

Theorem 2.2.4.3. *If the dimension of a subspace of* $n-$*tuples is* k*, the dimension of the null space is* $n - k$*.*

2.3 Some Additional Theory on Rings

Let R be a ring and $a \in R$. If $\exists b \neq 0$ in R such that $ab = 0$ then we call b a *zero divisor* of a and conversely.If R has no zero divisors i.e. $ab = 0 \Rightarrow a = 0 \quad or \quad b = 0$ for $a, b \in R$ then R is an integral domain. A commutative integral domain with a unit element is called a *Unique factorisation domain* (UFD) if;

- every non-unit of R is a finite product of irreducible factors.
- every irreducible element is prime.

An example of a UFD is the ring of integers.
Consider now a subring J of R in which the product ar or ra is well defined for $a \in J$ and $r \in R$. Then J is called an *ideal* of R. For a commutative case (i.e where R is commutative), J is generated by one element of $a \in R$. We write $J = \langle a \rangle$ and call J a *principal ideal.* A commutative ring R in which every ideal is principal is referred to as a *principal ideal domain*(PID). Any two elements a and b in a PID have a gcd which can be obtained as the generator of the ideal $\langle a, b \rangle$. Since ideals are normal subgroups of the additive group of a ring, it follows that an ideal J of a ring R defines a partition of R into disjoint cossets called *residue classes* modulo J. We denote the residue class of an element $a \in R$ modulo J by $[a] = a + J$. The set of residue classes of R modulo J forms a ring with respect to the operations;

- $(a + J) + (b + J = (a + b) + J$
- $(a + J)(b + J) = ab + J$

Such a ring is called *residue class ring* or *factor ring* of R modulo J and is denoted by R/J.

Definition 2.3.0.4. *An ideal M of R is said to be* ***maximal*** *if $M \subseteq J \subseteq R \Rightarrow J = M$ or $J = R$.*

The following is a useful characterisation of a maximal ideal.

Theorem 2.3.0.5. *Let R be a commutative ring with identity. Then an ideal M of R is maximal if and only if R/M is a field.*

Definition 2.3.0.6. *An integral domain R is said to be a Euclidean Ring if for every $a \neq 0$ in R, there is a defined nonnegative integer $\nu(a)$ such that;*

i). for all $a, b \in R$ both nonzero$\nu(a) \leq \nu(ab)$

ii). for any $a, b \in R$, both nonzero there exist $t, r \in R$ such that

$$a = tb + r$$

where $r = 0$ or $\nu(r) < \nu(b)$.

Theorem 2.3.0.7. *let R be a Euclidean Ring. Then;*

- *R is a principal ideal domain.*
- *any two elements $a, b \in R$ have a greatest common divisor d. Moreover,*

$$d = \lambda a + \mu b$$

for some $\lambda, \mu \in R$.

Definition 2.3.0.8. *Let R be a commutative ring with unit element (identity). An element $a \in R$ is a unit in R if there exists an element $b \in R$ such that $ab = 1$.*

Note 2.3.0.9. *Do not confuse a* ***unit*** *with a* ***unit element****! A unit in a ring is an element whose inverse is also in the ring.*

Definition 2.3.0.10. *In the Euclidean ring R, a non unit π is said to be a* ***prime*** *element of R if whenever $\pi = ab$ where a, b are in R then one of a or b is a unit in R.*

Lemma 2.3.0.11. *Let R be a Euclidean Ring. Then every element in R is either a unit in R or can be written as the product of a finite number of prime elements of R.*

2.3.1 Polynomials

Definition 2.3.1.1. *Let R be an arbitrary Ring. A* ***polynomial*** *is an expression of the form;*

$$f(x) = \sum_{k=1}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

where n is a nonnegative integer, $a_i : 0 \leq i \leq n$ are elements of R and x is a symbol not belonging to R called ***an indeterminate*** *over R. We refer to $a_n \neq 0$ the leading coefficient, $a_0 \neq 0$, the constant term. The integer n is called the* ***degree*** *of the polynomial, denoted by $deg(f(x))$. By convention, $deg(0) = -\infty$. Here 0 stands for the zero polynomial. If R has identity and the leading coefficient of $f(x)$ is 1, then $f(x)$ is said to be* ***monic****.*

We shall consider the polynomials

$$f(x) = \sum_{k=1}^{n} a_i x^i$$

$$g(x) = \sum_{k=1}^{n} b_i x^i$$

to be equal if and only if $a_i = b_i$ for $0 \leq i \leq n$
We can add polynomials in a natural way. The sum of two polynomials $f(x)$ and $g(x)$ is the polynomial $s(x)$;

$$s(x) = f(x) + g(x) = \sum_{k=1}^{n} (a_i + b_i) x^i$$

We can also multiply two polynomials. Let

$$f(x) = \sum_{k=1}^{n} a_i x^i$$

$$g(x) = \sum_{k=1}^{m} b_j x^j$$

Then the product of $f(x)$ and $g(x)$ is the polynomial $p(x)$;

$$p(x) = f(x)g(x) = \sum_{k=1}^{n} C_k x^k$$

where $C_k = \sum_k a_i b_j, \quad i + j = k$.
With these operations, the set of of all polynomials over R with one indeterminate x forms a ring. We call this ring the polynomial ring and denote it by $R[x]$.

Theorem 2.3.1.2. *Let R be a ring. Then;*

i). $R[x]$ is commutative if and only if R is commutative.

ii). $R[x]$ is a ring with identity if and only if R has an identity.

iii). $R[x]$is an integral domain if and only if R is an integral domain.

Remark 2.3.1.3. *From now on we will be concerned with the case where R is a Euclidean Ring. In fact for coding purposes we will in future be interested in polynomials over fields, more specifically finite fields.*

Theorem 2.3.1.4 (Division Algorithm). *Let $g \neq 0$ be apolynomial in $R[x]$. Then for any $f \in R[x]$ there exist polynomials $q, r \in F[x]$ such that;*

$$f = qg + r$$

where$deg(r) < deg(g)$

Remark 2.3.1.5. *The fact that $R[x]$ permits a division algorithm implies that every ideal of $R[x]$ is principal. Infact $R[x]$ is a Euclidean Ring if we consider the degree function. Therefore we can have the following generalisation of the GCD.*

Theorem 2.3.1.6. *Let $f_1, f_2 \cdots, f_n$ be polynomials in $R[x]$ not all of which are 0. Then there exists a uniquely determined monic polynomial $d \in R[x]$ with the following properties;*

i). d divides each f_i, $1 \leq i \leq n$

ii). any polynomial $c \in R[x]$ dividing each $f_j, 1 < j < n$, divides d.

Moreover, d can be expressed in the form;

$$d = b_1 f_1 + \cdots + b_n f_n$$

where $b_1, b_2 \cdots, b_n \in R[x]$

Theorem 2.3.1.7. *Let $f_1, f_2 \cdots, f_n$ be polynomials in $R[x]$ not all of which are 0. Then there exists a uniquely determined monic polynomial $m \in R[x]$ with the following properties;*

i). m is a multiple of each f_j, $1 \leq j \leq n$

ii). any polynomial $b \in R[x]$ that is a multiple of each f_j, $1 \leq j \leq n$, is a multiple of m.

The polynomial m is called the least common multiple of $f_1 \cdots f_n$

Irreducible Polynomials

When $g(x)$ divides $f(X)$ without a remainder, then $g(x)$ is a *factor* of $f(x)$. A polynomial $f(x)$ over a field $\mathbb{F}$ is said to be *irreducible* over $\mathbb{F}$ if $f(x)$ can not be expressed as a product of two polynomials both over $\mathbb{F}$ and of degree lower than that of $f(x)$. We sometimes refer to an irreducible polynomial as *prime* polynomial.The ring of polynomials over a finite field has an irreducible polynomial of every degree.

Note 2.3.1.8. *Irreducibility of polynomials depends on the field under consideration. For example $x^2 - 2 \in \mathbb{Q}[x]$ is irreducible over $\mathbb{Q}$. However over $\mathbb{R}$. $x^2 - 2 = (x + \sqrt{2})(x - \sqrt{2})$.*

2.4 Field Extension

Recall that a field $\mathbb{F}$ is a commutative ring with identity whose nonzero elements form a multiplicative group. Suppose κ is a subset of $\mathbb{F}$ and κ is itself a field with respect to the binary operations of $\mathbb{F}$, then we call κ a *subfield* of $\mathbb{F}$. For example, the field $\mathbb{R}$ of real numbers can be considered as a subfield of the field $\mathbb{C}$ of complex numbers.

Definition 2.4.0.9 (Extension Field). *Let κ be a subfield of $\mathbb{F}$ and M be a subset of $\mathbb{F}$. Then the intersection of all subsets containing both κ and M is called an* ***extension field*** *of κ obtained by adjoining the elements of M. It is denoted by $\kappa(M)$.*
If $M = (\theta_1, \cdots, \theta_n)$ then $\kappa(M) := (\theta_1, \cdots, \theta_n)$. If $M = \theta$, then (θ) is called a ***simple extention.***

Definition 2.4.0.10. *Let $\mathbb{F}$ be an extention field of κ. If $\mathbb{F}$ considered as a vector space over κ is finite dimensional then $\mathbb{F}$ is called a* ***finite extension*** *of κ. We call this dimension the* ***degree*** *of $\mathbb{F}$ over κ.*

Definition 2.4.0.11. *Let κ be a subfield of $\mathbb{F}$ and $\theta \in \mathbb{F}$. If θ satisfies a polynomial equation*

$$a_n x^n + \cdots + a_1 x + a_0 = 0 \tag{2.4.1}$$

in $\kappa[x]$, then θ is said to be ***algebraic*** *over κ. otherwise θ is* ***trancedental****. We call an extension $\mathbb{F}$ over κ algebraic if all the elements of $\mathbb{F}$ are algebraic over κ.*

Definition 2.4.0.12. *Let θ be an algebraic element over κ. Then the unique monic polynomial $g \in \kappa[x]$ generating the ideal $J = f \in \kappa[x]; f(\theta) = 0$ of $\kappa[x]$ is called the* ***minimal polynomial*** *of θ over κ. It is irreducible and if any polynomial $h(x) \in \kappa[x]$ is such that $h(\theta) = 0$, then $g(x) \mid h(x)$.*

Note 2.4.0.13. *Any element of an extension field of degree m over κ has a minimum polynomial of degree m or less.*

A polynomial $f \in \kappa[x]$ is said to *split* in an extension field $\mathbb{F}$ if f is expressible as a product of linear factors in $\mathbb{F}[x]$ i.e. there exists elements $\alpha_1, \alpha_2, \cdots, \alpha_n \in \mathbb{F}$ such that;

$$f(x) = a(x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n)$$

The field $\mathbb{F}$ is called the *splitting field* of f over κ, if f splits in $\mathbb{F}$ and $\mathbb{F} = \kappa(\alpha_1, \alpha_2, \cdots, \alpha_n)$, that is if it is the smallest field containing the roots of f.

Theorem 2.4.0.14 (Existence and Uniqueness of Splitting Field). *If κ is a field and f is any polynomial of positive degree in $\kappa[x]$, then there exists a splitting field of f over κ. Any two splitting fields of f over κ are isomorphic under an isomorphism which keeps the elements of κ fixed and maps roots of f into each other.*

2.5 Structure of Finite Fields

A field is said to be finite if it has finite number of elements. The residue class ring $\mathbb{Z}/p\mathbb{Z}$ of integers modulo a prime p is an example of a finite field. This field is sometimes denoted by $\mathbb{F}_p$. It is a prime field [1] with p elements.

[1] a field with no proper subfields

Proposition 2.5.0.15. *Let κ be a field with q elements. Then there exist a prime p such that;*

i). $\mathbb{F}_q \subseteq \kappa$.

ii). $q = p^n$ *for some integer* n.

iii). $\alpha^q = \alpha \quad \forall \alpha \in \kappa$.

Theorem 2.5.0.16 (Existence and Uniqueness of Finite Fields). *For every prime p and every positive integer n there exists a finite field with p^n elements. Any finite field with $q = p^n$ elements is isomorphic to the splitting field of $x^q - x$ over $\mathbb{F}_p$*

Lemma 2.5.0.17. *The field $\mathbb{F}_{p^m}$ is a subfield of $\mathbb{F}_{p^n}$ if and only if m divides n.*

Proposition 2.5.0.18. *For every finite field $\mathbb{F}_q$ the multiplicative group $\mathbb{F}_q^*$ of non-zero elements of $\mathbb{F}_q$ is cyclic.*

Definition 2.5.0.19. *A generator of the cyclic group F_q^* is called a* ***primitive*** *element of $\mathbb{F}_q$.*

2.6 Roots

Since the set of roots of an irreducible polynomial over a finite field will be essential in description of codes, we take a look at some information that can be derived from it.

Lemma 2.6.0.20. *Let $f \in \mathbb{F}_q[x]$ be an irreducible polynomial over $\mathbb{F}_q$ of degre m. Then $f(x)$ divides $x^{q^m} - x$ if and only if m divides n. Moreover the splitting field of $f(x)$ over $\mathbb{F}_q$ is given by $\mathbb{F}_{q^m}$ Here $q = p^n$.*

An irreducible polynomial in $\mathbb{F}_q[x]$ of degree m has a root a in $\mathbb{F}_{q^m}$. Furthermore, all the roots of f are simple ([14],27) and are given by the m distinct elements $\alpha, \alpha^q, \cdots \alpha^{q^m}$ of $\mathbb{F}_{q^m}$. Any two irreducible polynomials in $\mathbb{F}_q[x]$ of the same degree have isomorphic splitting fields.

Definition 2.6.0.21. *Let $\mathbb{F}_{q^m}$ be an extension of $\mathbb{F}_q$ and let $\alpha \in \mathbb{F}_{q^m}$. Then the elements $\alpha, \alpha^q, \cdots \alpha^{q^m}$ are called the* ***conjugates*** *of α with respect to F_{q^m}.*

2.7 Arithmatic in $\mathbb{F}_2$

We now look at a suitable field that we will prove useful in code construction. Since our binary symbols are strings of $1's$ and $0's$ taken from $\mathbb{Z}_2$ perhaps we could use it. However it has only two digits, thus if each character is to be represented in the field, we will need

a bigger field. This field can be obtained by considering the set on integers modulo a power of 2, say 2^m. Thus we can consider the field $\mathbf{Z}_{2^m}$ where m is to be determined.

But we soon realize that ordinary arithmetic in $\mathbf{Z}_{2^m}$ does not meet the criterion for inverses. For example in $\mathbb{Z}_{2^4}$ (containing interges $0, 1, ...15$), 2 has no inverse. Thus, $\mathbf{Z}_{2^4}$ is not a field with standard multiplication as $\mathbf{Z}_2$ is. Ordinary arithmetic fails because $\mathbb{Z}_{16} \cong \frac{\mathbb{Z}}{\langle 16 \rangle}$ and the generator of $\langle 16 \rangle$ is a composite number. Remember $\mathbf{Z}_p \cong \frac{\mathbb{Z}}{\langle p \rangle}$ is a field for p a prime.

We thus need a different operation at least for the sake of multiplication. This operation is readily offered by polynomials. We know that $\mathbb{Z}_p(\alpha) \cong \frac{\mathbb{Z}_p[x]}{\langle p(x) \rangle}$ when $p(x)$ is a minimal polynomial with root α. Its degree n is such that

$$p^n = o(\mathbb{Z}_{p(\alpha)}) \tag{2.7.1}$$

We have thus successfully replaced the non prime 2^m with a prime polynomial. This isomorphism asserts the existence of the required field. In general we consider the ring $\mathbb{F}_2[x]$ of polynomials over $\mathbb{F}_2$. In order to construct a field with 16 elements, Equation 2.7.1 asserts that the prime polynomial need to be of degree 4 since $2^4 = 16$. In addition, this polynomial must be primitive in $\mathbb{F}_{2^4}$ These two conditions lead us to the choice of between two polynomials, namely

$$x^4 + x^3 + 1 \quad and \quad x^4 + x + 1$$

Where $\mathbb{F} = \mathbb{Z}_2$ $dim[\mathbb{Z}_2(\alpha) : \mathbb{Z}] = 3$ with the basis $\{0, 1, \alpha\}$.
So, we begin by the basis elements of the extension field and perform all the arithmatic modulo $p(x) = x^4 + x^3 + 1$ raising α to powers successively identifies α^2 *and* α^3 as members of the extension field. We note the fact that $\alpha^4 = \alpha^3 + 1$ and use it as an identity for reducing each power of α greater than three. All the elements of $\mathbb{Z}_{2^4}$ generated by the polynomial $p(x) = x^4 + x^3 + 1$ are tabulated below.

element	Polynomial representation	binary reoresentation
0	0	(0,0,0,0)
1	1	(0,0,0,1)
α	α	(0,0,1,0)
α^2	α^2	(0,1,0,0)
α^3	α^3	(1,0,0,0)
α^4	α^3+1	(1,0,0,1)
α^5	$\alpha^3+\alpha+1$	(1,0,1,1)
α^6	$\alpha^3+\alpha^2+\alpha+1$	(1,1,1,1)
α^7	$\alpha^2\alpha+1$	(0,1,1,1)
α^8	$\alpha^3+\alpha^2+\alpha$	(1,1,1,0)
α^9	α^2+1	(0,1,0,1)
α^{10}	$\alpha^3+\alpha$	(1,0,1,0)
α^{11}	$\alpha^3+\alpha^2+1$	(1,1,0,1)
α^{12}	$\alpha+1$	(0,0,1,1)
α^{13}	$\alpha^2\alpha$	(0,1,1,0)
α^{14}	$\alpha^3+\alpha^2$	(1,1,0,0)
$\alpha^{15}=1$	1	(0,0,0,1)

Table 2.3: elements of $\mathbb{Z}_{2^4}$ generated by the polynomial $p(x)=x^4+x^3+1$

We can also use the primitive polynomial $x^4 + x + 1$ to obtain the same elements of $\mathbb{Z}_{2^4}$.

Table 2.4: elements of $\mathbb{Z}_{2^4}$ generated by the polynomial $p(x) = x^4 + x + 1$

element	Polynomial representation	binary reoresentation
0	0	(0,0,0,0)
1	1	(1,1,1,1)
α	α	(0,0,1,0)
α^2	α^2	(0,1,0,0)
α^3	α^3	(1,0,0,0)
α^4	$\alpha + 1$	(0,0,1,1)
α^5	$\alpha^2\alpha$	(0,1,1,0)
α^6	$\alpha^3 + \alpha^2$	(1,1,0,0)
α^7	$\alpha^3 + \alpha + 1$	(1,0,1,1)
α^8	$\alpha^2 + 1$	(0,1,0,1)
α^9	$\alpha^3 + \alpha$	(1,0,1,0)
α^{10}	$\alpha^2\alpha + 1$	(0,1,1,1)
α^{11}	$\alpha^3 + \alpha^2 + \alpha$	(1,1,1,0)
α^{12}	$\alpha^3 + \alpha^2 + \alpha + 1$	(1,1,1,1)
α^{13}	$\alpha^3 + \alpha^2 + 1$	(1,1,0,1)
α^{14}	$\alpha^3 + 1$	(1,0,0,1)
$\alpha^{15} = 1$	1	(0,0,0,1)

To examine why a primitive polynomial is needed to generate all the field elements, we use the non primitive polynomial $g(x) = x^4 + x^3 + x^2 + x + 1$. Mechnically,these fields are

Table 2.5: elements of $\mathbb{Z}_{2^4}$ generated by the polynomial $p(x) = x^4 + x^3 + 1$

element	Polynomial representation	binary reoresentation
0	0	(0,0,0,0)
1	1	(1,1,1,1)
α	α	(0,0,1,0)
α^2	α^2	(0,1,0,0)
α^3	α^3	(1,0,0,0)
α^4	$\alpha^3 + \alpha^2 + \alpha + 1$	(1,1,1,1)
α^5	1	(1,1,1,1)
α^6	α	(0,0,1,0)
α^7	α^2	(0,1,0,0)
α^8	α^3	(1,0,0,0)
α^9	$\alpha^3 + \alpha^2 + \alpha + 1$	(1,1,1,1)
α^{10}	1	(1,1,1,1)
α^{11}	α	(0,0,1,0)
α^{12}	α^2	(0,1,0,0)
α^{13}	α^3	(1,0,0,0)
α^{14}	$\alpha^3 + \alpha^2 + \alpha + 1$	(1,1,1,1)

easy to generate. Multiplication of an element by α result in a left shift of the previous binary value. If a '1' is shifted out of the fourth degree, it constitutes a polynomial of the fourth degree and the result is reduced by subtracting the generator polynomial from the result. However, the subtraction in the binary arithmatic is merely exclussive 'OR' function. So an encoder for the field elements is easy to build in logic circuits.

2.8 Cyclotomic polynomials

For $b \neq 0$ in a field $\mathbb{K}$, the *exponent* of b is the smallest positive integer n (if it exists) such that $b^n = 1$.That is b is a root of $x^n - 1$ but not of $x^d - 1$ for any smaller d. The polynomial $\Psi_n \in \mathbb{K}[x]$ such that $\Psi_n(b) = 0$ if and only if b is of exponent n is called the n^{th} *order cyclotomic polynomial* of b.

Corollary 2.8.0.22. *The function $x^n - 1$ has no repeated roots in $\mathbb{K}[x]$ if and only if the characteristic of $\mathbb{K}$ does not divide n.*

Proposition 2.8.0.23. *For $n > 1$ define inductively the n^{th} order cyclotomic polynomial $\Psi_n(x)$ as;*

$$\Psi_n = \frac{x^n - 1}{lcm \quad of \quad all \quad x^d - 1 \quad 0 < d < n; \quad d|n}$$

with the lcm monic, then we have the following properties;

(i) Ψ_n *is a monic polynomial.*

(ii) for $\alpha \in \mathbb{K}$ *(scalar)* $\Psi_n(\alpha) = 0$ *if and only if* $\alpha^n = 1$ *and* $\alpha^t \neq 0$ *for all* $0 < t < n$

(iii) $gcd(\Psi_m(x), \Psi_n(x)) = 1$ *for* $m < n$ *with* $m \nmid c, \quad n \nmid c$ *where* c *is the characteristic of the field* $\mathbb{K}$.

(iv) The degree of , $\Psi_n(x) = \varphi(x)$ *-the Euler's phi-function.*

(v) $\Psi_n(x) = \frac{x^n - 1}{\prod_{1 \leq d \leq n \quad d|n} \Psi_d(x))}$

(vi) $x^n - 1 = \prod_{1 \leq d \leq n: \quad d|n} \Psi_d(x))$

We conclude by using the polynomial $x^{63} - 1$ and its factorisation over $\mathbb{F}_2$ to illustrate some concepts already discussed. We pick a primitive element α of $\mathbb{F}_{64} = \mathbb{F}_{2^6}$. Then all the field elements can be expressed as a power of this element α, and they are:

$$0, 1, \alpha, \alpha^2, \alpha^3, \cdots, \alpha^{62}$$

Each of the elements in $\mathbb{F}_{64}^*$ are roots of $x^{63} - 1$, i.e. we have the factorisation

$$x^{63} - 1 = a(x - 1)(x - \alpha)(x - \alpha^2) \cdots (x - \alpha^{62})$$

$\mathbb{F}_{64}^*$ is cyclic by corollary 2.5.0.18, and α is primitive, $\alpha^{63} = 1$. We then have that α^{21} is of order 3 and so is α^{42}. By Lagranges theorem, the order of each element must divide 63, which is the order of the group. As such, the possible orders of the integers m are $1, 3, 7, 9, 21, 63$. Morover we pick the smallest such m so that $(\alpha^{21})^9 = 1$, yet α^{21} is considered of order 3 and not of order 9. We have the orders of the roots as follows;

$$\alpha^{21}, \alpha^{42} \quad are \quad of \quad order \quad 3$$

$$\alpha^9, \alpha^{18}, \alpha^{27}, \alpha^{36}, \alpha^{45}, \alpha^{54} \quad are \quad of \quad order \quad 7$$

$$\alpha^7, \alpha^{14}, \alpha^{28}, \alpha^{35}, \alpha^{49}, \alpha^{56} \quad are \quad of \quad order \quad 9$$

There are 21 elements of order 21. These are multiples of 3 that are divisible by neither 7 nor 9. The rest of the elements are of arder 63. The cyclotomic polynomials are;

$$\Psi_1 = x - 1$$
$$\Psi_3 = \frac{x^3 - 1}{x - 1}$$

and so on.

Then $x^{63} - 1 = \Psi_1\Psi_3\Psi_7\Psi_9\Psi_{21}\Psi_{63}$

Upto cyclotomic polynomials, we have the following factorizations;

$$x^{21} - 1 = \Psi_{21}\Psi_7\Psi_3\Psi_1$$
$$x^9 - 1 = \Psi_9\Psi_3\Psi_1$$
$$x^7 - 1 = \Psi_7\Psi_1$$
$$x^3 - 1 = \Psi_3\Psi_1$$
$$x - 1 = \Psi_1$$

where $\Psi_i = \Psi_i(x)$

Thus:

$$\Psi_1 = x - 1 \quad degree \quad 1$$
$$\Psi_3 = \frac{x^3 - 1}{\Psi_1} \quad degree \quad 2$$
$$\Psi_7 = \frac{x^7 - 1}{\Psi_1} \quad degree \quad 6$$
$$\Psi_9 = \frac{x^9 - 1}{\Psi_3\Psi_1} \quad degree \quad 6$$
$$\Psi_{21} = \frac{x^{21} - 1}{\psi_7\Psi_3\Psi_1} \quad degree \quad 12$$
$$\Psi_{63} = \frac{x^{63} - 1}{\Psi_{21}\Psi_9\psi_7\Psi_3\Psi_1} \quad degree \quad 36$$

Remark 2.8.0.24. *The order e of an element β of the extension field $\mathbb{F}_q$ divides $q^k - 1$ but no smaller number of the form $q^n - 1$ where k is the degree of the minnimum polynomial of β.*

$$e = 3, \quad 3 \mid 2^n - 1 \quad for \quad which \quad we \quad have \quad n = 2$$

Therefore the number of elements of order 3 should have a minimum polynomial over $\mathbb{F}_q$ with degree 2 because they are elements of $\mathbb{F}_{2^2}$. Thus $\Psi_3(x)$ is irreducible over $\mathbb{F}_2$ similarly, for $e = 7 \quad 7 \mid 2^3 - 1$, so that $n = 3$

Thus the elements of order 7 belong to $\mathbb{F}_{2^3}$ and have a minimin function of degree 3 over $\mathbb{F}_2$. All the other elements $\mathbb{F}_{2^6}$ are elements of no subfield, and therefore all their minimum polynomials have degree 6 over $\mathbb{F}_{2^6}$.

Note also that because all the roots of an irreducible polynomial have the same order, if one root of an irreducible polynomial $p(x)$ has order i, all the roots have order i and are therefore roots of $\Psi_i(x)$ and thus $p(x)$ divides $\Psi_i(x)$. Therefore, $\Psi_9(x)$ must be irreducible, $\Psi_{21}(x)$ has two factors of degree 6, and $\Psi_{63}(x)$ has six factors of degree 6.

Chapter 3

Some Coding Theory

3.1 Linear Block Codes

With the backgroung information in Algebra, we are now ready to handle some coding theory. We introduce the notion of a code and quickly give it nice mathematical structure to obtain an object amenable to mathematical analysis.

Definition 3.1.0.25. *We Let $\mathbb{Z}^+$ denote the set of even integers and $\mathbb{F}$ a finite set. This finite set constitute our alphabet and its elements are called* ***letters****. Then*

$$V = \{x_1 \cdots x_n \mid x_i \in \mathbb{F}, \quad 1 \leq i \leq n; \quad n \in \mathbb{Z}^+\}$$

is the set of all possible size n-tuple of letters (repetition allowed) from $\mathbb{F}$. A subset C of V is referred to as a ***Code*** *and the elements of C are called* ***codewords****. For a fixed $n \in \mathbb{Z}^+$, we refer to a subset C of V a* ***blockcode*** *of length n if each of its elements is n-tuple of elements of $\mathbb{F}$. Without loss of Generality we take the finite set $\mathbb{F}$ to be a finite field. If C has k information symbols (k is the dimension of C as a vector space over $\mathbb{F}$, then we will sometimes refer to C as an (n, k) code. The quantity $R = \frac{k}{n}$ gives the* ***information rate.***

The choice of the field $\mathbb{F}$ is quite deliberate. The design of electronic circuitry requires that we we work with the field $\mathbb{F}$ which consist of elements $0, 1$. This is the field $\mathbb{Z}_2$. Thus the information is transmitted as blocks of digits of uniform length comprising of $0's$ *and* $1's$.

What is the motivation behind the description of a code as par Definition 3.1.0.25. The answer follows from the following argument. We concider the simplest technique used in error detection; the parity check, in which a single 0 or 1 is added at the end of the data block so that the block has an even number of 1's. If during transmission a sigle error (in one place) is committed then the received block of data will have an odd number of $1's$. The receiver can then request for a re-transmission. However, in satellite communication,

retransmission is prohibitively expensive and often time consuming. A better strategy would then involve encoding the data in a way that allow the receiver to detect and to correct errors!

A very intuitive strategy would involve the introduction of redundant symbols. One way of doing this would involve transmitting each binary digit (bit) of the original message a specified number of times, say r. The decoder on receiving the possibly garbled message then decodes by replacing the r-tuple by the modal bit. Suppose the encorder transmits 00000 but 00101 is received instead, then the decoder on realising that there are more 0's than 1's decodes the block as 00000. Such codes are called *repetition codes*. It is clear that the code above detect up to 4 errors and correct 2 of them.

Remark 3.1.0.26. *i). In general, repetition codes are can detecting the presence of up to $r-1$ errors and correcting up to $\lfloor \frac{r-1}{2} \rfloor$ [1] errors in each $r-$tuple*

ii). Repetition codes can be viewed as (n,k) codes, where $n = kr$, and $R = \frac{k}{n} = \frac{1}{r}$. This shows that the information rate decreases with increase in r.

Another way of introducing ridundancy is by sphere packing. In this case the information is transmitted as sets of points in higher dimensional space, say an $n-$dimensional space. If k is the number of information symbols (in our case binary digits) to be transmitted, then $n-k$ symbols are adjoined to it in a more systematic way than mere repetition. This is the concept of parity check. This notion can be represented diagramatically as;

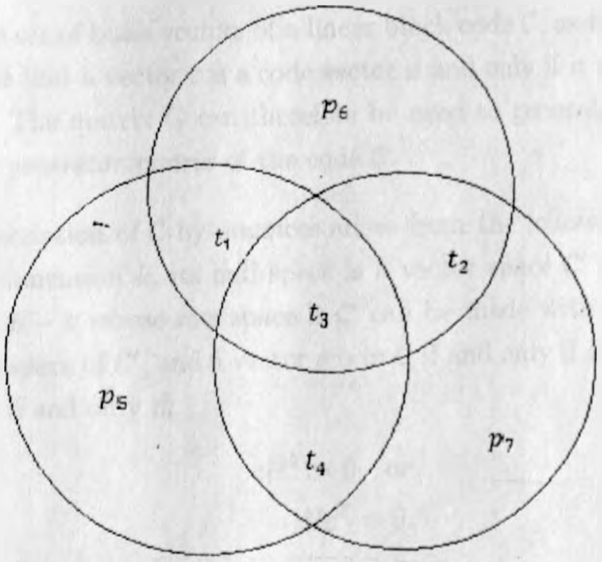

fig.3.1 Sphere Packing.

The idea is to have an even number of i's in each sphere. An important decoding scheme

[1] $\lfloor x \rfloor$ is the smallest integer $\leq x$

we will see in a later section (syndrome decoding) will involve finding, mathematically, the bits that have been eroneously changed on transmission resulting in odd number of 1's in certain spheres.

Since words to be encoded are of fixed length k, we can think of them as elements of $\mathbb{F}^k$ and the code as elements of $\mathbb{F}^n$. The encoding process is defined as a one-to-one map

$$E : \mathbb{F}^k \rightarrow \mathbb{F}^n$$

The image $C = E(\mathbb{F})$ is called the code. With this map, we can associate the decoding operation

$$D : \mathbb{F}^n \rightarrow \mathbb{F}^k$$

such that

$$D \cdot E = I_{\mathbb{F}^k}$$

By defining the domain of this function E, we have our codewords hence the information to be encoded. But how do we define the domain of the function? For large block codes without proper internal structure, it may be difficult to even determine codewords, let alone defining the code. This motivates the idea of additional structure which will facilitate characterization of codewords. We make some restriction to our map so that it is linear. We then have that a code is a vector subspace of $\mathbb{F}^n$ of dimension k. We can now think of E as a linear transformation. Matrix description then follows automatically.

By considering the set of basis vectors of a linear block code $\mathcal{C}$, as the rows of a matrix G, say, we then realise that a vector c is a code vector if and only if it is a linear combination of the rows of G. The matrix G can therefore be used to generate the code $\mathcal{C}$. We call such a matrix *the generator matrix* of the code $\mathcal{C}$.

An alternative description of $\mathcal{C}$ by matrices arises from the following argument. Since $\mathcal{C}$ is a sub-space of dimension k, its null space is a vector space $\mathcal{C}'$ of dimension $n - k$. A matrix H of rank $n - k$ whose row space is $\mathcal{C}'$ can be made with a basis for $\mathcal{C}'$ as rows. Then $\mathcal{C}$ is the null space of $\mathcal{C}'$, and a vector c is in $\mathcal{C}$ if and only if it is orthogonal to every row of H, that is, if and only if;

$$\begin{aligned} cH^T &= 0 \quad or \\ Hc^T &= 0. \end{aligned} \tag{3.1.1}$$

If $c = (c_1, c_2, c_3, \cdots, c_n)$. Denote the element in the i^{th} row and j^{th} column of H by h_{ij} , then Equation 3.1.1 implies that for each i (that is, each row of H),

$$\sum_j c_i h_{ij} = 0 \tag{3.1.2}$$

Thus, the components of C must satisfy a set of $n - k$ independent equations. These equations are called *generalized parity checks*, since in the binary case they are simply checks for even parity on certain sets of symbols in the code word. That is, for each row

of H, the number of $1's$ in c that corresponds to $1's$ in that row of H is even (for the binary case) if and only if c satisfies Equation 3.1.2. The matrix H is called a *parity-check matrix* of C.

Definition 3.1.0.27. *Two codes C and C' are said to be* ***equivalent*** *if they differ only in the arrangement of symbols.*

Equivalent codes have same probability of errors.

By performing elementary row operations on the matrix G, we obtain a matrix $G' = [I_k : A]$. Then G and G' generates the same code. We say that G' is the *standard form.* An easy calulation shows that the parity check matrix H in standard form is $\left[-A^T \mid I_{n-k}\right]$ We now consolidate the concepts by a rather detailed example.

Example 3.1.0.28. *Consider a binary $(5,3)$ code. The vectors,*

$$(0,0,0,0,0)(1,0,0,1,1),(0,1,0,1,0),(1,1,0,0,1),(0,0,1,0.1),$$
$$(1,0,1,1,0),(0,1,1,1,1) \quad \textit{and} \quad (1,1,1,0,0) \tag{3.1.3}$$

forms a vector space say C. The code C is a rowspace of;

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

or

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix}$$

The two matrices are combanitorially equivalent (one of the matrices can be obtained from the other by a combination of row operations and column permutations) hence generates the same code. The former is most preferable for generating a $(5,3)$ code.

Any message $x_1, x_2, x_3 \in \mathbb{F}^3$: $\mathbb{F} = \mathbb{F}_2$ is encoded as a $5-$tuple

$$\begin{aligned}(x_1, x_2, x_3, x_4, x_5) &= (x_1, x_2, x_3)G \\ &= (x_1, x_2, x_3, x_1 + x_2, x_1 + x_3)\end{aligned}$$

hence;

$$\begin{aligned}x_4 &= x_1 + x_2 \\ x_5 &= x_1 + x_3\end{aligned}$$

(a total of $n-k$ equations) are the parity check equations. Thus for $(x_1, x_2, x_3, x_4, x_5)$ to be an encoded message c, it must satisfy the parity check equations;

$$\begin{aligned}(-x_1) + (-x_2) + x_4 &= 0 \\ (-x_1) + (-x_3) + x_5 &= 0\end{aligned}$$

In matrix terms;

$$\begin{pmatrix} -1 & -1 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_5 \end{pmatrix} = 0$$

i.e. $Hx^T = 0$ *where;*

$$H = \begin{pmatrix} -1 & -1 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 & 1 \end{pmatrix}$$

which has the form $\left[-A^T \mid I_{n-k}\right]$

3.1.1 Decoding Process

A code is of no use if the encoded message does not get to the intended party in a manner that can be understood. The information is considered conveyed if the encoded message is correctly decoded back to the original message. Several techniques have been employed to decode information. The simplest one is the nearest neighbourhood decoding. Syndrome decoding has also proved to be an important decoding scheme. It is important to note that there are other more complex algorithms for decoding. We are not going to discuss them here. An iterested reader is referred to [8]. [6] for more on decoding.

Before illustrating the decoding process, let us look at more code parameters.

Definition 3.1.1.1. *The* ***weight*** wt *of a codeword* $x \in \mathcal{C}$ *is defined as;*

$$wt(x) = |\{i : x_i \neq 0\}|$$

and is simply the number of non-zero components of the codeword

Definition 3.1.1.2. *For any two codewords* $x, y \in \mathcal{C}$ *the* ***distance*** d *between them is gived by;*

$$D(x, y) = |\{i : x_i \neq y_i\}|$$

and is the number of components in which they differ. It is a distance function Clearly $d(x, y) = wt(x - y)$

We can also define the distance between a codeword and a code as $D(x, C) = min\,\{d(x, c) \mid c \in \mathcal{C}\}$

Theorem 3.1.1.3. *Let* $\mathcal{C}$ *be a code such that for all* $x, y \in \mathcal{C} \subset \mathbb{F}_q^n$ *satisfy* $d(x, y) \geq d$ *for* $d \geq 1$. *Any* $d-1$ *or fewer errors in a received word can be detected. If* $d \geq 2t + 1$ *for* $t \geq 1$, *then* t *or fewer errors can be detected.*

Proof. Given x and suppose that e is an error of x. If no more than $d - 1$ of bits are changed then $d(x, e) \leq d - 1$ so that $e \in B_{d-1}(x)$ (where $B_{d-1}(x)$ is an open sphere of radius $d - 1$ centered at x) which contradicts the minimum distance. Thus, an error can

be detected.
If $d \geq 2t+1$ then for all $z \in \mathbb{F}_q^n$ by triangle inequality

$$\begin{aligned} d(x,z) + d(z,y) &\geq d(x,y) \\ &\geq 2t+1 \end{aligned} \tag{3.1.4}$$

This implies that $d(x,z) > t$ or $d(y,z) > t$ so that $B_t(x) \cap B_t(y) = 0$ so that $B_t(x) = \{x\}$ for all x. Suppose x has an error e' so that x appears as $x' = x + e'$ to the receiver. Since C is a vector space, $x, e', x' \in C$ and more importantly if x' and x differ by no more than $t-1$ errors (that is, fewer than t errors are introduced), then $x, x' \in B_t(x) = \{x\}$. Therefore x' lies within a sphere of radius t about a codeword x, then x' is the zero codeword. We have thus recovered our original message.

Maximum-likelihood decoding

Consider the (n,k) code C with generator matrix of the form $G = [I_k \mid A]$, so that the original message $x_1 \cdots, x_k$ is encoded as $x = x_1 \cdots x_k x_{k-1} \cdots x_n = (x_1 \cdots x_k)G$. The original message comprises of the first $k-$components of the encoded message.
Suppose after transmission the message received is $\tilde{x} = \tilde{x_1} \cdots \tilde{x_n}$. If $\tilde{x}G \in C$, then the received message is decoded as the first $k-$components of $\tilde{x}$. Two cases may arise; either $\tilde{x} = x$ or $\tilde{x} \neq x$. Whichever the case $\tilde{x_1} \cdots, \tilde{x_k}$ is presumed to be the original message (even though enough errors may have occurred during transmission like in the former case). If $\tilde{x} \notin C$ then an error has occured. In this case $B_r(x) \neq \{x\}$. Then $\tilde{x}$ is decoded as x, with x chosen so that $d(\tilde{x}, x)$ is minnimised (and $\tilde{x} \neq x$). Since $B_t(x') = \{x, \tilde{x}\}$, then this algorithm will choose x, i.e. the nearest neighbor, as the candidate for decoding!

Note 3.1.1.4. *Since we need to look through all $x \in C$. this operation takes $o(q^k)$ running time.*

Syndrome Decoding

Syndrome decoding algorithm is a faster algorithm (in comparison with the maximum-likelihood decoding) that makes use of the vector space structure inherent in C. Suppose $x = wG$ is a codeword for $w \in \mathbb{F}_q^k$ and $e \in \mathbb{F}_q^n$ is an error that is introduced in the transmission of x such that the receive message is decoded as $x' = x + e$. Then

$$x'H^T = (x+e)H^T = xH^T + eH^T = 0 + eH^T = eH^T$$

so that xH^T depends only on $eH^T \in F_q^{n-k}$ The elements eH^T are called *syndromes*. For efficient processing, each of the elements are computed before the decoding process, and the coset representative with the smallest number of non-zero entries is chosen to be the coset leader. The algorithm works as follows:

- If $x \in \mathbb{F}_q^n$ is received, $s = xH^T$ is computed and compared to the coset leader l associated to s.
- If l is unique, replace x by $x' = x - l$. which is an element of C since C is a vector space.
- If l is not unique, report an error.
- If fewer than t errors occur in x, then x' represents the unique codeword closest to the received word. Thus, $E^{-1}(x')$ is returned.

Remark 3.1.1.5. *Unlike the Nearest Neighbor Decoding Algorithm, this requires only $o(q^{n-k})$ running time. While this running time is better than nearest neighbor algorithm, it is still "slow". In fact, decoding algorithms are known to be NP-complete so our present algorithm is only relatively fast.*

Example 3.1.1.6. *Consider the code $C = \{00000, 10110, 01111, 11001\}$. This is an $(n, k, d) = (5, 2, 3)$ binary code, containing four coded messages, with $t = [(d-1)/2] = [(3-1)/2] = 1$. Hence, it can detect whether 1 up to $d - 1 = 2$ errors have occurred, and can correct $t = 1$ error. The code has*

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

and

$$H = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

as generator matrix and parity check matrix.

For the decoding table,

Table 3.1: Standard array and syndromes for a $(5, 2, 3)$ binary code.

Syndromes	Coset Leader			
000	00000	10110	01111	11001
110	10000	00110	11111	01001
111	01000	11110	00111	10001
100	00100	10010	01011	11101
010	00010	10100	01101	11011
001	00001	10111	01110	11000
011	01100	11010	00011	10101
101	01010	11100	00101	10011

Note 3.1.1.7. *Each $n-$tuple with weight atmost t will be the unique vector of the smallest weight in its coset and will thus be the coset leader. Two or more $n-$tuples of weight atmost t can not appear in the same coset.*

In table below we illustrate how the encoded message can be recovered from the received message when the weight of the error vector is at most t. and what can go wrong when its weight is greater than t.

Table 3.2: Error correction for the $(5, 2, 3)$ binary code having generator matrix $G = (I_2 \mid A)$.

Example	x	$\bar{x}$	e	i	$\bar{x} - i$
1	10110	10100	00010	00010	10110
2	11001	10101	01100	01100	11001
3	10110	10101	00011	01100	11001
4	01111	11011	10100	00010	11001

In this case $x - i$ is assumed to be the encoded message that was sent.

In example 1. $wt(e) = 1 \leq t$ and the theorem holds. In example 2., $wt(e) > t$ but since $l = e$, the encoded message is recovered. In example 3., we aren't so lucky as $l \leq e$. In example 4., $\bar{x}$ is in a coset whose leader has weight atmost t and so we cannot recover the message.

3.2 Cyclic Codes

In order to construct good codes and to have practical decoding algorithm, we need to consider codes with more algebraic structure. In this section, we describe a class of codes with more algebraic structure, cyclic codes. We will realise that many of the important linear block codes are equivalent to cyclic codes. Once again we will consider a code to be a binary code unless stated otherwise.

Definition 3.2.0.8. *A linear code C is said to be* ***cyclic*** *if $C' = (c_n, c_1, c_2, \cdots, c_{n-1})$ is a codeword whenever $C = (c_1, c_2, \cdots, c_n)$ is.*

We recall that a code C is a subspace of $\mathbb{F}_q^n$. Since vectors in a vector space V can be represented as polynomials in an indeterminate say x, i.e. we can define a homomorphism (indeed an isomorphism) say,

$$\Pi; a \longrightarrow a(x)$$

where $a \in V$ is given by $a = (a_0, a_1, \cdots a_{n-1})$ and $a(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}$. If we consider the correspondence between the vectors a and the class $a(x)$ in the factor ring of polynomials in x with coefficients from $\mathbb{F}$ modulo the ideal of

$$R = \frac{\mathbb{F}[x]}{x^n - 1}$$

generated by $x^n - 1$, multiplication by $|x|$ in this ring corresponds to performing a cyclic shift in our vectorspace. Therefore the cyclic code is merely an Ideal of R and conversely.

Since R is a principal ideal ring, if $g(x) \in \mathbb{F}[x]$ is a monic polynomial of least degree, then $g(x)$ generates an ideal in R. Thus $g(x)$ divides $x^n - 1$ since $(x^n - 1, g(x)) \in \langle g(x) \rangle$. Recall that the gcd of two polynomials is a linear combination of the two polynomials with the coefficients from the field (by Theorem 2.3.1.6) i.e $(x^n - 1, g(x)) = \alpha(x^n - 1) + \beta g(x) \mid \alpha, \beta \in \mathbb{F}[x]$ and $(x^n - 1, g(x)) \in \langle g(x) \rangle$ (since R is a P.I.D). This unique polynomial is the generator of the ideal or for our case the code. It is a polynomial of degree $n - k$ where k is the dimension of the cyclic code $\mathcal{C}$ generated by $g(x)$.

Proposition 3.2.0.9. *If $\mathcal{C}$ is a cyclic code of length n with the generator polynomial*

$$g(x) = g_1 + g_2 x + g_3 x^2 + \cdots + g_k x^{k-1}$$

then the generator matrix of $\mathcal{C}$ is the $(n - k + 1) \times n$ matrix given by,

$$A = \begin{pmatrix} g_1 & g_2 & g_3 \cdots g_k & 0 & 0 & 0 \cdots 0 \\ 0 & g_1 & g_2 \cdots g_{k-1} & g_k & 0 & 0 \cdots 0 \\ 0 & 0 & g_1 \cdots g_{k-2} & g_{k-1} & g_k & 0 \cdots 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 \cdots * & * & * \cdots & g_k \end{pmatrix}$$

Proof. The rows of G are easily seen as linearly independent. Hence if we show that any codeword can be written as a linear combination of these rows, then we are home. Now by definination a vector $C = (c_0, c_1 \cdots c_{n-1})$ is a codeword *iff* the corresponding polynomial $c(x) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}$ is of the form $c(x) = g(x)f(x)(mod(x^n - 1))$ for some polynomial f which can be taken to be of degree $\leq n - 1$. But this means (in the obvious notation) that;

$$\begin{aligned} c(x) &= g(x)(f_0 + f_1 x + \cdots + f_{n-1} x^{n-1}) \\ &= \sum_{i=0}^{n-1} f_i x g(x)(mod(x^n - 1)) \end{aligned}$$

and this in turns is exactly the statement that

$$c = f_0 \mathbf{g} + f_1 \mathbf{g}^{(1)} + \cdots + f_{n-1} \mathbf{g}^{(n-1)}$$

where $\mathbf{g} = (g_1, g_2 \cdots g_n)$ and $\mathbf{g}^{(k)}$ denotes the cyclic shift of $\mathbf{g}$ by k places. We can thus say that that a cyclic code is completely specified by a polynomial $g(x)$ that divides $x^n - 1$.

We can alternatively say that the code is a nullspace of the ideal generated by;

$$h(x) = \frac{x^n - 1}{g(x)}$$

The polynomial $h(x)$ is referred to as the parity check polynomial of the code $\mathcal{C}$ generated by $g(x)$ hence $h(x) = (x^n - 1) \mid g(x)$ it can be used as a generator polynomial of another cyclic code. Such a code is equivalent to the dual code of $\mathcal{C}$. In the context of cyclic code,

it is simply referred to as *the dual code* of the cyclic code C. The generator matrix of this code is given by;

$$H = \begin{pmatrix} 0 & 0 & \cdots & 0 & h_k & h_{k-1} & \cdots & h_1 & h_0 \\ 0 & 0 & \cdots & h_k & h_{k-1} & h_{k-2} & \cdots & h_0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ h_k & h_{k-1} & \cdots & h_0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

Example 3.2.0.10. *consider the polynomial $x^7 - 1$ over $GF(2)$. The polynomial splits into;*

$$(x^7 - 1) = (x - 1)(x^3 + x + 1)(x^3 + x^2 + 1)$$

consider a case where $g(x) = x^3 + x^2 + 1$. This polynomial generates a $(7, 4)$ code $(i.e. n = 7, \quad deg(g(x)) = 3 \quad therefore, \quad k = 7 - 3 = 4)$

$$G = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

For the parity check polynomial;

$$\begin{aligned} h(x) &= \frac{x^7 - 1}{g(x)} \\ &= (x - 1)(x^3 + x + 1) \\ &= x^4 + x^3 + x^2 + 1 \end{aligned} \tag{3.2.1}$$

The code C can be defined as the null space of the ideal $\langle h(x) \rangle$. In this case;

$$H = \begin{pmatrix} 0 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Since cyclic codes are linear codes we can put the generator matrix and the parity check matrix in the standard form. Needless to say, the standard form is much more useful in encoding and decoding. Rather than first computing the generator matrix and the parity check matrix, we can make use of the generator polynomial to generate the two matrixes in a single algorithm. To achieve this, we let r_i to be the remainder after dividing x^i by $g(x)$ i.e. $x^i = g(x)q_i(x) + r_i(x)$. Then $x^i - r_i(x) = g(x)q_i(x)$ is a code vector. If these polynomials; for $i = n - 1, n - 2, \cdots, n - k$ are taken as rows of the generator matrix, then

$$G = [I_k, -R]$$

where I_k is a $k \times k$ identity matrix and $-R$ is a $k \times (n - k)$ matrix whose j^{th} row is the vector of coefficient of $-r_{n-j}(x)$. The parity check matrix therefore has the form;

$$H = [R^T, I_{n-k}]$$

The j^{th} row of H^T is the vector coefficient of $r_{n-j}(x)$, even for $j \leq n - k$.

Example 3.2.0.11. *Consider again the binary cyclic code generated by* $g(x) = x^3 + x^2 + 1$

$$\begin{aligned}
x^6 &= g(x)(x^3 + x^2 + x) + x^2 + x \\
x^5 &= g(x)(x^2 + x + 1) + \quad x + 1 \\
x^4 &= g(x)(x + 1) \quad + x^2 + x + 1 \\
x^3 &= g(x)(1) \quad\quad + x^2 \quad + 1 \\
x^2 &= g(x)(0) \quad + x^2 \\
x &= g(x)(0) \quad\quad + x \\
x^0 &= g(x)(0) \quad\quad\quad + 1
\end{aligned} \tag{3.2.2}$$

then

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

and

$$H = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Remark 3.2.0.12. *We have previously seen that all that is important about a code is the generator matrix and the parity check matrix. Since for cyclic codes, these can be obtained from the generator polynomial, than cyclic codes are completely specified by the generator polynomial.*

We now look at a theorem which give a suggestion on the minimum distance of a cyclic code.

Theorem 3.2.0.13. *[19] Let* $g(X)$ *be the generator polynomial of a cyclic code of length* n *over* $\mathbb{F}_q$ *and let* $\alpha^{e_1}, \alpha^{e_2} \ldots, \alpha^{e_{n-k}}$ *be the roots of* $g(X)$*, possibly in an extension field, where* α *is an element of order* n*. The minimum distance of the code is greater than the largest number of consecutive integers modulo* n *in the set* $e = (e_1, e_2, \ldots e_{n-k})$

Remark 3.2.0.14. *The bound on the minimum distance follows from the fact that any* $2t$ *or fewer columns are independent, from a van der Monde argument.*

As a consequence of the above theorem, we have the following.

Corollary 3.2.0.15. *[19] A cyclic code with roots* $\alpha^{e+1}, \ldots, \alpha^{e+j(d_0-2)}$ *and possibly others, where* α *is an element of order* n*, has minimum distance* d_0 *or greater provided* $(j, n) = 1$

Remark 3.2.0.16. *Note that Theorem 3.2.0.13 suggest some sort of a bound on the minimum distance and does not give the actual minimum distance. Thus even though Cyclic codes are completely specified by the generator polynomial, the information on the minimum distance is not quite clear just by having the generator polynomial. However, by choosing wisely the generator polynomial, the information on the minimum distance can be gained.*

We conclude this section on cyclic codes by considering a class of cyclic code whose generator polynomial is "carefully chosen."

3.2.1 Special Classes of Cyclic Codes

BCH Codes

The BCH abbreviations stand for the discoverers Bose and Chaudhuri(1960) and independently Hocquenghem(1959). These Codes form a large class of multiple random error-correcting codes.

The class of *BCH* codes is rather parametrised in the sense that the generator polynomial, is carefully chosen, the creterion of which lies in Theorem 3.2.0.13, and Corollary 3.2.0.15. Such codes therefore enjoy simplified decoding algorithm. Apart from simplicity in decoding, codes of this nature enjoy flexibility which allows control over block lengths and acceptable error threshold. Thus a custom code can be designed to a given specification(subject to mathematical constraints)

Definition 3.2.1.1 (BCH Codes). *Let $\alpha \in \mathbb{F}_{q^m}$. For any specified m_0 and d_0, (see Theorem 3.2.0.13, and Corollary 3.2.0.15) the code C generated by $g(X)$ is a BCH code if and only if $g(X)$ is the polynomial of the lowest degree over $\mathbb{F}_{q^m}$ for which $\alpha^{m_0}\alpha^{m_0+1}, \ldots, \alpha^{m_0+d_0-2}$, are roots. We call the code C primitive if $\alpha^n = q^m - 1$ and narrow sense if $m_0 = 1$. The lenth of C is the lcm of the orders of the roots.*

Remark 3.2.1.2. *The most important BCH codes are the binary codes obtained by letting α be a primitive element of $\mathbb{F}_{2^m}$ and letting $m_0 = 1$ and $d_0 = 2t_0 + 1$. Then $\{f(x)\}$ is a code vector if and only if;*

$$\alpha, \alpha^2, \alpha^3, \cdots, \alpha^{2t_0}$$

are roots of $f(x)$. However, every even power of α is a root of the same minimum function as some previous odd power of α. Therefore, an equivalent statement is that $\{f(x)\}$ is a code vector if and only if;

$$\alpha, \alpha^3, \cdots, \alpha^{2t_0-1}$$

are roots of $f(x)$. Thus, the generator polynomial of the code is

$$g(X) = LCM(m_1(x), m_3(x), \cdots, m_{2t_0-1})$$

where $m_i(x)$ is the minimum polynomial of α^i Therefore $g(x)$ has degree at most mt_0, and the code has at most mt parity checks. Hence the corollary;

Corollary 3.2.1.3. *For any positive integers m and $t_0 < \frac{n}{2}$, there is a BCH binary code of length $n = 2^m - 1$ which corrects all combinations of t_0 or fewer errors and has no more than mt_0 parity check symbols.*

Example 3.2.1.4. *Consider a case when you anticipate to correct upto 3 errors. Then* $t_0 = 3$. *Let us choose* $m = 4$, *then* $n = 2^m - 1$. *We thus construct a code using* $x^{15} - 1$ *over* $\mathbb{F}_2$. *Let* α *be a primitive element in* $\mathbf{F}_{2^4}$. *We make use of a primitive polynomial we already know;* $p(x) = 1 + x + x^4$

$$\begin{aligned} m_1(x) &= 1 + x \\ m_3(x) &= 1 + x + x^2 + x^3 + x^4 \\ m_5(x) &= 1 + x + x^2 \\ g(x) &= LCM(m_1(x)m_3(x)m_5(x)) \\ &= m_1(x)m_3(x)m_5(x) \\ &= 1 + x + x^2 + x^4 + x^5 + x^8 + x^{10} \end{aligned}$$

The code is a (15, 5) cyclic code.

Reed-Solomon Codes

In 1960, Irving Reed and Gus Solomon published a paper in the Journal of the Society for Industrial and Applied Mathematics [17]. This paper described a new class of error-correcting codes that are now called Reed-Solomon (RS) codes. These codes have great power and utility, and are today found in many applications from compact disc players to deep-space applications. This class of codes has very close relation with the BCH codes and more often considered as a non binary instance BCH code ([22],[17]).

As with all BCH codes, we consider m consecutive roots, but the roots are taken from the ground field itself and not an extension field of it. Thus taking α for the primitive root in $\mathbb{F}_q$, the other roots are α^i, and all have minimal polynomial of degree equal to unity, $(x - \alpha^i)$. Thus the generating polynomial is;

$$g(x) = \prod(x - \alpha^i)$$

We have $(n - k) = m \quad d = m + 1 = n - k + 1$ for the distance. Notice that this is the maximum possible value for d since there are only $(n - k)$ rows in H and thus any $(n - k + 1)$ columns of H are linearly independent. Where the length of the code is q, the error correcting capability t is $t = \lfloor \frac{q-k+1}{2} \rfloor$. Singleton, in 1964, showed that this was the best possible error correction capability for any code of the same length and dimension [21]. Codes that achieve this "optimal" error correction capability are called *maximum distance separable* (MDS).

We now look at a sample construction of an RS code. For convenience we choose $m_0 = 0$ for the sequence

$$\alpha^{m_0}, \alpha^{m_0+1}, \cdots, \alpha^{m_0+m-1}$$

We take $d^* = 5$ so that the RS code is constructed over $\mathbf{F}_{2^3}$. Let $1, \alpha, \alpha^2, \alpha^3$ where α is a primitive in $\mathbb{F}_{2^3}$ i.e. α can be taken as a root of $x^3 + x + 1$. This gives:

$$\begin{aligned} \alpha^3 &= \alpha + 1 \\ \alpha^4 &= \alpha^2 + \alpha \\ \alpha^5 &= \alpha^2 + \alpha + 1 \\ \alpha^6 &= \alpha^2 + 1 \end{aligned} \tag{3.2.3}$$

so that

$$\begin{aligned} m(x) &= (x+1)(x+\alpha)(x+\alpha)(x+\alpha) \\ &= (x^2 + x + \alpha^3 x + x + \alpha)(x^2 + x + \alpha^5 x + x + \alpha^5) \\ &= x^4 + x + \alpha^2 x^3 + x + \alpha^5 x + x + \alpha^6 \\ &== x^4 + \alpha^2 x^3 + \alpha^5 x + x + \alpha^6 \end{aligned} \tag{3.2.4}$$

This gives a $(7, 3)$ RS code on $\mathbb{F}_{2^3}$ with $d = n - k + 1 = 5$,

$$G = \begin{pmatrix} 1 & 0 & 0 & \alpha & \alpha^3 & \alpha^6 & \alpha^6 \\ 0 & 1 & 0 & 1 & \alpha^4 & \alpha^2 & \alpha \\ 0 & 0 & 1 & \alpha^2 & \alpha^5 & \alpha^5 & \alpha^6 \end{pmatrix}$$

and

$$H = \begin{pmatrix} \alpha & 1 & \alpha^2 & 1 & 0 & 0 & 0 \\ \alpha^3 & \alpha^4 & \alpha^5 & 0 & 1 & 0 & 0 \\ \alpha^6 & \alpha^2 & \alpha^5 & 0 & 0 & 1 & 0 \\ \alpha^6 & \alpha & \alpha^6 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Chapter 4

Some Algebraic Geometry in Coding Theory

4.1 Introduction

The use of Algebraic Geometry codes had its inception when Goppa ([16].[18]) made the crucial observation that codes can be constructed by evaluating a set of rational functions at the points on an algebraic curve. In making this step, many of the tools needed to determine the important parameters of the code, or bounds on them, such as the code length, dimension, and minimum distance, already existed in the elegant theorems of Algebraic Geometry, notably the Hasse-Weil theorem and the Riemann-Roch theorem.

The theory of Algebraic-Geometry codes makes use of the relatively deep and fundamental results of algebraic geometry. For our purpose we will bypass most of the underlying Algebraic Geometry and provide only a brief overview of those concepts from Algebraic Geometry needed to appreciate the development. The theorems in this chapter and the proofs there of follow from [3],[4], [5], [10], [20] where a much rigorous and complete survey of algebraic-geometry codes has been documented. In the next section we set the platform for description of Algebraic Geometry Codes. We consider another construction of RS codes redefine BCH codes in such a manner that makes natural the extension of constructions to codes from algebraic curves. The mathematical background required to understand the application of Algebraic Geometry to Coding is outlined in Section 4.2. Most of the facts here are stated without proof, but effort is made to convey the intuition as much as possible. Section 4.3 uses the ideas developed in Section 4.2 to outline the construction of codes that are derived from algebraic curves. We will see that Goppa codes are instances of Algebraic-Geometry codes and so are RS codes, and some instances of BCH Codes.

4.1.1 Transition

One construction of a Reed-Solomon (RS) code over the finite field $\mathbf{F}_q$ is as follows; Let $\{\alpha_0, \alpha_1, \cdots, \alpha_{n-1}\}$ be a set of n distinct elements from $\mathbb{F}_q$ and let $L \subset \mathbf{F}_q[x]$ denote the set of polynomials of degree less than $k \leq n$. Define the code C by,

$$C = \{(f(\alpha_0), f(\alpha_1), \ldots, f(\alpha_{n-1})); f \in L\}$$

which has length n and dimension k, since a monomial basis easily leads to a generator matrix of rank k. Since a polynomial of degree less then k has at most $k-1$ zeros, each codeword has weight at least $n-(k-1) = n-k+1$. As it is easy to construct polynomials with exactly this many zeros, this is the minimum distance of the code, so the code is MDS.

Further, let $\{v_0, v_1, \cdots v_n\}$ be a set of nonzero, not necessarily distinct elements from $\mathbf{F}_q$.

$$C' = \{(v_0 f(\alpha_0), v_1 f(\alpha_1), \cdots v_n f(\alpha_n)) \mid f \in L\}$$

The code has the same parameters as the previous code and is referred to as the *generalized RS (GRS) code* with vector $\{v_0, v_1, \cdots v_n\}$.

To prepare for a definition of Goppa codes, the definition of BCH codes is first recast. Consider the computation

$$\begin{aligned}
(x^n - 1)\sum_{i=0}^{n-1} \frac{c_i}{x - \alpha^{-i}} &= \sum_{i=0}^{n-1} c_i (x^n - 1)\frac{1}{x(1 - x^{-1}\alpha^{-i})} \\
&= \sum_{i=0}^{n-1} c_i \frac{x^n - 1}{x}.(1 + x^{-1}\alpha^{-i} + x^{-2}\alpha^{-2i} + \cdots) \\
&= \sum_{i=0}^{n-1} c_i \sum_{j=0}^{n-1} x^j (\alpha^{-i})^{n-1-j} \\
&= \sum_{j=0}^{n-1} x^j \sum_{i=0}^{n-1} c_i (\alpha^{j+i})^i
\end{aligned} \tag{4.1.1}$$

For $j = 1, 2, \cdots, d-2$, the inner summation is zero, by definition |Since $c(\alpha^j) = 0; \quad 1 \leq j \leq d-2$|. Thus

$$(x^n - 1)\sum_{i=0}^{n-1} \frac{c_i}{x - \alpha^{-i}} = x^{d-1} f(x)$$

for some polynomial $f(x)$, i.e., the summation is divisible by x^{d-1} (since the result is a polynomial with no constant term of degree atleast $d-1$). Thus

$$\sum_{i=0}^{n-1} \frac{c_i}{x - \alpha^{-i}} \equiv 0 \mod x^{2t}$$

Consequently, a word $(c_1, c_2, \ldots, c_{n-1}; c_i \in \mathbb{F}_q)$, is a codeword if and only if it satisfies the above equation. The construction yields either an RS or BCH code depending on the field of definition. Notice that the polynomial x^{2t} has a zero of order $2t$ at $x = 0$.

The passage from the above definition to that of Goppa codes will involve nothing more than replacing the sequence of n^{th} roots of unity with an arbitrary set of distinct elements and the polynomial x^{2t} with a more general polynomial $g(x)$.

Note 4.1.1.1. *Note that this is not the generator polynomial used in the BCH construction. It is conventional to use $g(x)$ in both cases.*

Definition 4.1.1.2. *Let $L = \alpha_0, \alpha_1, \cdots, \alpha_{n-1}$ be a set of n distinct elements in $\mathbb{F}_{q^m}$ and $g(x)$ be a monic polynomial such that $g(\alpha_i) \neq 0. i = 0, 1, \ldots, n-1$. Then the* ***Goppa code*** *$\Gamma_{(L,G)}$ is the set of words $(c_0, c_1, \ldots, c_{n-1}) \in \mathbb{F}_{q^n}$ such that*

$$\sum_{i=0}^{n-1} \frac{c_i}{x - \alpha^{-i}} \equiv 0 \mod g(x)$$

The polynomial $g(x)$ is referred to as the ***Goppa polynomial.***

Comparing to the previous formulation, if $g(x) = x^{2t}$ and, $L = \{\alpha^{-i}, 0 \leq i \leq n-1\}$, where α is a primitive n^{th} root of unity, then $\Gamma_{(L,g)}$ is a BCH code with designed distance d, although it is noted that not all BCH codes are Goppa codes [26]. It is also noted that $\Gamma_{(L,g)}$ is a subfield subcode of the dual of a generalized RS code.

To put the transition to codes from algebraic curves in perspective, it will be of interest to recast the definition of Goppa codes. Consider a polynomial corresponding to a codeword $(c_0, c_1, \ldots c_{n-1})$

$$\begin{aligned} f(x) &= \sum_{i=0}^{n-1} \frac{c_i}{(x - \alpha_i)} = \frac{\omega(x)}{\lambda(x)} \\ \lambda(x) &= \prod_i (x - \alpha_i) \in \mathbb{F}_{q^m}[x] \end{aligned} \tag{4.1.2}$$

and $deg\omega(x) < deg\lambda(x) = n$. Then

$$c_i = f(x)(x - \alpha_i) \mid_{x=\alpha_i}$$

is obtained by cancelling the simple pole in $f(x)$ at α_i and evaluating the result at α_i, i.e. it is the residue of $f(x)$ at α_i. Let

$$\chi_j(x) = \prod_{i=1, i \neq j}^{n} (x - \alpha_i) = \frac{\lambda(x)}{(x - \alpha_i)}$$

and let

$$f(x) = \frac{\omega(x)}{\lambda(x)} = \frac{g(x)q(x)}{\lambda(x)}$$

since by definition $g(x)|f(x)$. Note that the residue of $f(x)$ at α_i can be expressed as

$$Res_{\alpha_i}(f) = \frac{\omega(x)(x - \alpha_i)}{\lambda(x)} \mid_{x=\alpha_i} = \frac{g(\alpha_i)}{\chi_i(\alpha_i)} q(\alpha_i)$$

which is zero only if $q(\alpha_i) = 0$ as $g(\alpha_i), \chi_i(\alpha_i) \neq 0$, by definition. Now define a vector space L of rational functions $f(x)$, such that;

i $f(x)$ has zeros where $g(x)$ has zeros, with multiplicity at least those of $g(x)$

ii $f(x)$ has poles only contained in the set L and in that case only poles of order one.

Consider the set of n-tuples C' over $\mathbb{F}_{q^m}$ defined by,

$$C' = (Res_{\alpha_0} f, Res_{\alpha_1} f, \ldots, Res_{\alpha_{n-1}} f)$$

where the residue of a rational function is defined in the usual manner. It is seen immediately that the Goppa code is the subfield subcode of this set over $\mathbb{F}_q$. The two important perspectives to be drawn from this section, perspectives that will survive the transition to codes from algebraic curves intact, are the notions of defining codewords in the first instance, as the evaluation of a rational function at a fixed set of distinct places, and in the second instance, as the set of residues of a rational function at a fixed set of places. In the setting of Algebraic Geometry, the fixed set of places will be drawn from the points on a curve in an Algebraic Geometry. The determination of code parameters, however, will depend in crucial ways on the theory of algebraic curves. The next section will serve as an overview of this theory, in preparation for Section 4.3, which considers classes of codes that use these notions for their construction.

4.2 Some Basic Algebraic Geometry

We introduce the basic notions of Algebraic Geometry, in order to extend the construction and properties of codes discussed in the previous section to Algebraic-Geometry codes, to be discussed in the next section. We will give no proofs but refer to the standard textbooks ([7],[4]). The central concepts required are outlined in the most elementary way, with illustrative examples. In other words, our aim is to attempt to convey the key concepts required to appreciate the application of Algebraic Geometry to coding.

4.2.1 Affine and Projective Varieties

Definition 4.2.1.1. *Let $\mathbb{F}_q$ be the finite field with q elements and $\bar{\mathbb{F}}_q$ its algebraic closure. The n-dimensional **affine space** $\mathbb{A}^n$ is the set*

$$\mathbb{A}^n = \{(a_1, a_2, \ldots a_n) | a_i \in \bar{\mathbb{F}}_q\} \tag{4.2.1}$$

*An element $P \in \mathbb{A}^n$ is called an **affine point**, and if $P = (a_1, a_2, \cdots a_n)$ with $a_i \in \bar{\mathbb{F}}_q$ then the elements a_i are called the **coordinates** of P, the point. If $\mathbb{K}$ is a subfield of $\bar{\mathbb{F}}_q$ that contains $\mathbb{F}_q$ and P is a point with coordinates in $\mathbb{K}$, then P is called a **$\mathbb{K}$-rational point** and the set of $\mathbb{K}$-rational points of $\mathbb{A}^n$ is denoted by $\mathbb{A}^n\mathbb{K}$.*

On the set $\mathbb{A}^{n+1} \setminus \{0, 0, \ldots, 0\}$ an equivalence relation $\equiv$ is given by

$$(a_0, a_1, \ldots a_n) \equiv (b_0, b_1, \ldots b_n) \Longleftrightarrow \exists \lambda \in \mathbb{F}_q \setminus \{0\} \tag{4.2.2}$$

such that $b_i = \lambda a_i \quad i = 0, 1, \cdots, n.$
The equivalence class of $(a_1, a_2, \ldots a_n)$ is denoted by $(a_1 : a_2 : \cdots : a_n)$.

Definition 4.2.1.2. *The n-dimensional* ***projective space*** $\mathbb{P}^n$ *is the set of all equivalence classes* $(a_0 : a_1 : \ldots : a_n) | a_i \in \mathbb{F}_q$, *not all* $a_i = 0$. *An element* $P = (a_0 : a_1 : \cdots : a_n) \in \mathbb{P}^n$ *is called a point and* $(a_0 : a_1 : \cdots : a_n)$ *are called* ***homogeneous coordinates*** *of* P. *If* $\mathbb{K}$ *is a subfield of* $\bar{\mathbb{F}}_q$ *which contains* $\mathbb{F}_q$ *and* P *is a point for which there exist homogeneous coordinates* $(a_0, a_1, \ldots a_n)$, *then* P *is called a* $\mathbb{K}$***-rational point*** *and the set of* $\mathbb{K}$*-rational points of* $\mathbb{P}^n$ *is denoted by* $\mathbb{P}^n(\mathbb{K})$. *The set* $H = (0 : a_1 : \ldots : a_n) \in \mathbb{P}^n$ *is called the* ***hyperplane at infinity*** *and the points* $Q \in H$ *are the points at infinity. The mapping* $\varphi : \mathbb{A}^n \longrightarrow \mathbb{P}^n \setminus \{0\}$ *defined by*

$$\varphi(a_1, a_2, \cdots a_n) = (a_0 : a_1 : \cdots : a_n) \tag{4.2.3}$$

embeds $\mathbb{A}^n$ *in* $\mathbb{P}^n$.

As a matter of notation, Q will be reserved throughout to denote a point at infinity. The one-dimensional projective space $(1 : a_1) | a^i \in \mathbb{F}_q$, also called the *projective line*, consists of the points together with the point at infinity $(0 : 1)$. and this set will be used later for the construction of RS codes, thereby showing that the class of Algebraic Geometry codes is a special case of Reed-Solomon codes. A polynomial $f \in \bar{\mathbb{F}}_q[x_1, \cdots, x_n]$ can be considered as a map $f : \mathbb{A}^n \longrightarrow \bar{\mathbb{F}}_q$ defined by

$$f(P) = f(a_1, a_2, \ldots a_n) \tag{4.2.4}$$

If $f(P) = 0$ we call P a zero of f. More generally, with every $T \subseteq \mathbb{F}_q[x_1, \ldots, x_n]$ we associate the *zero set* of T

$$Z(T) = \{P \in \mathbb{A}^n \mid f(p) = 0 \quad \forall f \in T\} \tag{4.2.5}$$

Definition 4.2.1.3. *A subset* V *of* $\mathbb{A}^n$ *is called an* ***algebraic set*** *if there exists a* $T \subset \mathbb{F}_q[x_1, \ldots, x_n]$ *such that* $V = Z(T)$.

Definition 4.2.1.4. *Let* $V \subset \mathbb{A}^n$ *be an algebraic set. The set*

$$I(V) = \left\{\bar{\mathbb{F}}_q[x_1, \ldots, x_n] | f(p) = 0 \textit{for every } P \in V\right\}$$

is called ***the ideal*** *of* V.

It is easy to see that $\bar{\mathbb{F}}_q[x_1, \ldots, x_n]$ is indeed an ideal of V . The ring $\bar{\mathbb{F}}_q[x_1, \ldots, x_n]$ is Noetherian, that is, every ideal is finitely generated.
Recall that an ideal I with a single generating element is called *principal* and an ideal is *prime* if it is not the whole ring and whenever $ab \in I$ then $a \in I$ or $b \in I$. An ideal I is *maximal* in a set A if there is no proper ideal of A that properly contains I.

Lemma 4.2.1.5 (Hilbert Nullstellensatz[15]). *Every maximal ideal of $\bar{\mathbb{F}}_q[x_1,\ldots,x_n]$ is of the form $\langle x_1 - a_1,\ldots,x_n - a_n\rangle$ with $a_i \in \bar{\mathbb{F}}_q$. For every element $P = (a_1, a_2, \ldots a_n) \in \mathbb{A}^n$ the singleton $\{P\}$ is an algebraic set with ideal $I(P) = \langle x_1 - a_1,\ldots,x_n - a_n\rangle$.*

Definition 4.2.1.6. *An affine variety V in $\mathbb{A}^n$ is an algebraic set where $I(V)$ is a prime ideal.*

Remark 4.2.1.7. *A variety is simply a set of solutions to the polynomial equations $f_i = 0$.*

The set of $\mathbb{K}-$rational points of V is denoted by $V(\mathbb{K})$. If $I(V)$ has a set of generators in $\mathbb{K}[x_1,\ldots,x_n]$, we say that V is defined over $\mathbb{K}$ and write $V|\mathbb{G}$. In this case we associate with the variety $V|\mathbb{K}$ the ideal

$$I(V \mid \mathbb{K}) = I(V) \cap \mathbb{K}[x_1,\ldots,x_n].$$

Definition 4.2.1.8. *Let V be an affine variety. The quotient ring*

$$\bar{\mathbb{F}}_q[V] = \bar{\mathbb{F}}_q[x_1,\ldots,x_n]|I(V) \tag{4.2.6}$$

is called the ***coordinate ring*** *of V. If V is defined over $\mathbb{K}$, the quotient ring $\mathbb{K}[V] = \mathbb{K}[x_1,\ldots,x_n]|I(K|G)$ is called the coordinate ring of $V|K$.*

Remark 4.2.1.9. *The coordinate ring of a variety V can be considered as a set of polynomial functions with values in $\bar{\mathbb{F}}$ defined at every point of V : let $g \in \bar{\mathbb{F}}_q \in [V]$ and $G \in \bar{\mathbb{F}}_q[x_1,\cdots,x_n]$ such that $g = G + I(V)$. Put $g(P) = G(P)$. This definition is independent of the choice of the representative G : if $G' \in \bar{\mathbb{F}}_q[x_1,\ldots,x_n]$, and $G' + I(V) = G + I(V)$, then $G' - G \in I(V)$ and therefore $0 = (G' - G)P = G'(P) - G(P)$ hence $G'(P) = G(P)$. Since the ideal of the variety is a prime ideal the coordinate ring is a domain. The following definition is therefore possible.*

Definition 4.2.1.10. *Let V be an affine variety. The field of fractions of $\bar{\mathbb{F}}_q[V]$, denoted $\mathbb{F}_q(V)$, is called the* ***function field*** *of V. It follows from the definition of the function field $\bar{\mathbb{F}}_q(V)$ that it is a finitely generated extension of $\bar{\mathbb{F}}_q$, that is, there exists elements $x_1,\ldots,x_k \in \bar{\mathbb{F}}_q(V)$ such that $\bar{\mathbb{F}}_q(V) = \bar{\mathbb{F}}_q(x_1,\ldots,x_k)$. The dimension of an affine variety is the transcendence degree of $\bar{\mathbb{F}}_q(V)$ over $\bar{\mathbb{F}}_q$.*

Definition 4.2.1.11. *An affine curve is a variety of dimension 1.*

Remark 4.2.1.12.

- *An affine curve is an algebraic curve since it is defined over an algebraically closed field. We will be considering plane algebraic curves i.e. curves difined over $\bar{\mathbb{F}}_q(x,y)$ for our illustrative examples, though the concepts apply even to curves defined in $n-$dimensional algebraically closed field.*
- *As a matter of notation we will use χ to denote a curve in an algebraic geometry. When it is defined by a polynomial, we will denote the polynomial by F_χ or simply F when the curve is understood.*

Example 4.2.1.13. *Let $F \in \bar{\mathbb{F}}_q[x,y]$ be an irreducible polynomial. Consider the variety*

$$\chi = \{F = 0\} = \{P \in \mathbb{A}^2 \mid C(P) = 0\}$$

Clearly the function field $\bar{\mathbb{F}}_q(\chi)$ has a transcendence degree one, and therefore is an affine curve, and since it is contained in $\mathbb{A}^2$, (every point is defined by a 2-tupple) it is called an affine plane curve.

A point (x_0, y_0) on a curve χ, with equation $F(x,y) = 0$ is said to be *nonsingular* if the partial derivatives do not both vanish at the point. The tangent line at a point is a linear polynomial (i.e. a polynomial of degree one) described by the equation

$$l_{x_0,y_0}(x,y) = F_x(x_0,y_0)(x - x_0) + F_y(x_0,y_0)(y - y_0)$$

where $F_x(x,y)$ and $F_y(x,y)$ are partial derivatives of F with respect to x and y respectively.

Definition 4.2.1.14. *A curve χ is said to be **nonsingular** (or **smooth** or **regular**) if all the points on the curve are nonsingular, otherwise the curve is singular.*

A polynomial F which is the sum of monomials of the same degree is said to be *homogeneous*. A homogeneous polynomial $F \in \bar{\mathbb{F}}_q[x_1, \cdots, x_n]$ is said to have a zero at a point $P = (a_0 : a_1 : \cdots : a_n) \in \frac{\mathbb{P}^n}{\mathbb{F}_q}$ if $F(a_0, a_1, \cdots, a_n) = 0$. Of course this makes sense since $F(\lambda a_0, \lambda a_1, \cdots, \lambda a_n) = \lambda^d F(a_0, a_1, \cdots, a_n)$ if F is homogeneous of degree d. For a polynomial $f(x) \in \mathbb{F}_q[x]$ of degree d, the polynomial $y^d f(\frac{x}{y})$ will be homogeneous of degree d. Conversely, one can reduce a homogeneous polynomial of degree d in n variables to a (nonhomogeneous) polynomial in $n-1$ variables, through a process of *dehomogenization*.

Example 4.2.1.15. *As an example over a finite field $\mathbb{F}_q$, where $q = r^2 = p^{2m}$ consider the Hermitian curve, which is described by the polynomial $F(x,y) = y^r + y - x^{r+1}$. The curve is nonsingular since the derivatives*

$$F_x(x,y) = -(r+1)x^r = x^r$$

and

$$F_y(x,y) = ry^{r-1} + 1 = 1$$

have no roots in common.
The homogeneous form of this curve is given by $F(X,Y,Z) = Y^r Z + Y Z^r - X^{r+1}$

Notation 4.2.1.16. *For non-homogeneous polynomials, we use lower-case letters x, y as the indeterminates while we use capital letters X, Y and Z homogeneous polynomial.*

More generally, with every set T of homogeneous polynomials from $\bar{\mathbb{F}}_q[x_0, x_1, \cdots, x_n]$ we associate the zero set of T

$$Z(T) = \{P \in \mathbb{P}^n | f(P) = 0 \quad \forall f \in T\}.$$

Definition 4.2.1.17. *A subset V of $\mathbb{P}^n$ is called a* **projective algebraic set** *if there exists a set T of homogeneous polynomials such that*

$$V = Z(T)$$

Definition 4.2.1.18. *Let $V \subseteq \mathbb{P}^n$ be a projective algebraic set. The ideal in $\bar{\mathbb{F}}_q[x_0, x_1, \cdots, x_n]$ which is generated by all homogeneous polynomials F with $F(P) = 0$ for every $P \in V$ is called* **the ideal** *of V and is denoted by $I(V)$.*

Definition 4.2.1.19. *A projective variety $V \subseteq \mathbb{P}^n$ is a projective algebraic set such that $I(V)$ is a prime ideal [15].*

The set of $\mathbb{K}-$rational points of V is denoted by $V(\mathbb{K})$. If $I(V)$ has a set of homogeneous polynomials from $\mathbb{K}[x_0, x_1, \cdots, x_n]$ as generators, we say that V is defined over $\mathbb{K}$. In this case we associate with $V/\mathbb{K}$ the ideal

$$I(V/\mathbb{K}) = I(V) \cap \mathbb{K}[x_0, x_1, \cdots, x_n]$$

Definition 4.2.1.20. *Let $V \subseteq \mathbb{P}^n$ be a nonempty projective variety. The quotient ring*

$$\Gamma_h(V) = \bar{\mathbb{F}}_q[x_0, x_1, \cdots, x_n]/I(V)$$

is called the homogeneous coordinate ring of V. If V is defined over $\mathbb{K}$ then

$$\Gamma_h(V/\mathbb{K}) = \mathbb{K}[x_0, x_1, \cdots, x_n]/I(V/\mathbb{K}).$$

An element $f \in \Gamma_h(V)$ is called a *form* of degree d if $f = F + I(V)$ where F is a homogeneous polynomial of degree d.
The function field of V is defined by;

$$\mathbb{F}_q(V) = \left\{ \frac{g}{h} \mid g, h \in \Gamma_h(V) \text{ are forms of the same degree and } h \neq 0 \right\}$$

and

$$\mathbb{K}(V) = \left\{ \frac{g}{h} \mid g, h \in \Gamma_h(V/\mathbb{K}) \text{ are forms of the same degree and } h \neq 0 \right\}$$

The dimension of the projective variety V is the transcendence degree of $\mathbb{F}_q(V)$ over $\mathbb{F}$.

Definition 4.2.1.21. *A projective curve $\chi \in \mathbb{P}^n$ is a projective variety of dimension 1.*

We clarify the connection between projective and affine varieties. For a polynomial

$$F = F(x_1, \cdots, x_n) \in \bar{\mathbb{F}}[x_1, \cdots, x_n]$$

of degree n set

$$F^* = x_0^d F(x_1/x_0, \cdots, x_n/x_0) \in \bar{\mathbb{F}}[x_0, \cdots, x_n].$$

Then F^* is a homogeneous polynomial of degree d in $n+1$ variables.
Consider now an affine variety $V \in \mathbb{A}^n$, and the corresponding ideal $I(V) \in \bar{\mathbb{F}}[x_0, \cdots, x_n]$.
Define the projective variety $\tilde{V} \in \mathbb{P}^n$ as follows:

$$\tilde{V} = \{P \in \mathbb{P}^n | F^*(P) = 0 \quad \forall F \in I(V)\}$$

This variety is called the projective closure of V.

On the other hand, let $\tilde{V} \in \mathbb{P}^n$ be a projective variety and suppose that

$$W = \tilde{V} \cap \{(c_1 : \cdots : c_n) \in \mathbb{P}^n | c_0 \neq 0\} \neq 0$$

Define $\varphi : \mathbb{A}^n \to \mathbb{P}^n$ by

$$\varphi(a_1, \cdots, a_1) = (1 : a_1 : \cdots : a_n)$$

Then

$$V = \varphi^{-1}(W)$$

is an affine variety and

$$I(V) = \left\{ F(1 : x_1 : \cdots : x_n) | F \in I(\tilde{V}) \right\}$$

and the projective closure of V is $\tilde{V}$. If V is an affine variety and $\tilde{V}$ its projective closure, the function fields $\bar{\mathbb{F}}_q(V)$ and $\bar{\mathbb{F}}_q(\tilde{V})$ are isomorphic and V and $\tilde{V}$ have the same dimension.

Example 4.2.1.22. *The projective closure of the hermitian curve $f(x, y) = y^r + y - x^{r+1}$ is a variety with the equation*

$$y^r z + y z^r - x^{r+1} = 0$$

This curve has only one point at infinity, namely $(0 : 1 : 0)$.

4.2.2 Local Ring at a Point

Let V be a variety and $P \in V$. If there exists a representative $f = g/h$ and $h(P) \neq 0$, f is said to be defined at P.
The ring

$$O_P(V) = \{f \in \bar{\mathbb{F}}_q(V) | f \quad is \quad defined \quad at \quad P\}$$

is called the *local ring* at P.
The evaluation of $f \in O_P(V)$ is defined as $f(P) = g(P)/h(P)$ in the affine case and in the projective case let $g = G + I(V)$, $h = H + I(V) \in \Gamma_h(V)$, where G and H are homogeneous polynomials of degree d. Let

$$P = (a_0 : a_1 : \cdots : a_n)$$

Since

$$\frac{G(\lambda a_0, \lambda a_1, \cdots, \lambda a_n)}{H(\lambda a_0, \lambda a_1, \cdots, \lambda a_n)} = \frac{\lambda^d G(a_0, a_1, \cdots, a_n)}{\lambda^d H(a_0, a_1, \cdots, a_n)} = \frac{G(a_0, a_1, \cdots, a_n)}{H(a_0, a_1, \cdots, a_n)}$$

we can have

$$f(P) = G(a_0, \cdots, a_n)/H(a_0, \cdots, a_n)$$

if $H(P) \neq 0$ $O_P(V)$ is indeed a local ring, and its maximal ideal is given by

$$M_P(V) = \{f \in O_P(V) | f(P) = 0\}$$

Definition 4.2.2.1 (Valuation Ring). *A valuation ring of the function field $\bar{\mathbb{F}}_q(V)$ is a ring O with the properties*

i). $\bar{\mathbb{F}}_q \subset O \subset \bar{\mathbb{F}}_q(V)$

ii). *For any* $z \in \bar{\mathbb{F}}_q(V), z \in O, or \quad z^{-1} \in O$

Theorem 4.2.2.2. *Let O be a valuation ring of the function field $\bar{\mathbb{F}}_q(V)$. Then;*

i). *O is local ring and has as unique maximal ideal $\mathcal{P} = O \setminus O^*$ where $O^* = \{z \in O \mid \exists w \in O \quad : wz = 1\}$*

ii). *For $0 \neq x \in \bar{\mathbb{F}}_q(V), \quad x \in \mathcal{P} \quad \Leftrightarrow x^- \notin O$*

iii). *$\mathcal{P}$ is a principal ideal.*

iv). *If $\mathcal{P} = tO$ then any $0 \neq z \in \bar{\mathbb{F}}_q(V)$ has a unique representation of the form $z = t^n u$ for some $n \in \mathbb{Z}, \quad u \in O^*$.*

v). *O is a principal ideal domain. If $P = tO$ and $\{0\} \neq I \subseteq O$ is an ideal then $I = t^n O$ for some $n \in \mathbb{N}$.*

Definition 4.2.2.3. *Let O be a valuation ring of $\bar{\mathbb{F}}_q(V)$ and $\mathcal{P}$ its unique maximal ideal with $\mathcal{P} = tO$. Then $z \in \bar{\mathbb{F}}(V)$ has a unique representation $z = t^n u$ with $u \in O^*, n \in \mathbb{Z}$. We define $\nu_{\mathcal{P}}(z) = n$ and $\nu_{\mathcal{P}}(0) = \infty$.*

Observe that this definition does not depend on the choice of generator t of $\mathcal{P}$.

Definition 4.2.2.4. *A discrete valuation of $\bar{\mathbb{F}}_q(V)$ is a function $\nu : \bar{\mathbb{F}}_q(V) \to \mathbb{Z} \cup \{\infty\}$ with the following properties:*

i). $\nu_{\mathcal{P}}(x) = \infty \quad \Leftrightarrow x = 0.$

ii). $\nu_{\mathcal{P}}(xy) = \nu_{\mathcal{P}}(x) + \nu_{\mathcal{P}}(y)$

iii). $\nu_{\mathcal{P}}(x + y) \geq min\{\nu_{\mathcal{P}}(x), \nu_{\mathcal{P}}(y)\}$ *. with equality holding if* $\nu_{\mathcal{P}}(x) \neq \nu_{\mathcal{P}}(y)$

iv). *There exists an element z such that $\nu_{\mathcal{P}}(z) = 1$.*

v). $\nu_{\mathcal{P}}(a) = 0$ *for any* $0 \neq a \in \bar{\mathbb{F}}_q$.

Theorem 4.2.2.5. *Let χ be a curve (projective or affine) and P a point of χ. P is nonsingular if and only if is a $O_P(\chi)$ is discrete valuation ring.*

If the variety is defined over $\mathbb{K}$ one can also consider the function field $\mathbf{K}(V)$. The definitions and the theorems still hold when one exchanges $\bar{\mathbf{F}}_q$ and $\mathbb{K}$. If ν is a discrete valuation of $\mathbb{K}$ with valuation ring $O$and maximal ideal $\mathcal{P}$ then the pair $(O, \mathcal{P})$ is called a closed point of V and $d = [O/\mathcal{P} : \mathbb{K}]$ is called the degree of the point. If $\mathbb{K} = \bar{\mathbf{F}}_q$ then the closed points correspond to the nonsingular points and all have degree 1.

Example 4.2.2.6. *Consider the projective plane curve with equation $zy^2 + yz^2 = x^3$ over the field $\mathbb{F}_2$. Here $Q = (0 : 1 : 1)$. We can take as a local parameter $t = \frac{x}{z}$. Let $f = \frac{x}{(y-z)}$. We will determine $\nu_Q(f)$. We have;*

$$\frac{x}{(y-z)} = \frac{x^3}{x^2y + x^2z} = \frac{zy^2 + z^2y}{x^2y + x^2z} = \frac{zy}{x^2} = t^{-2}\frac{y}{z}$$

and the second factor is a unit in $O_Q(\chi)$ so $\nu_Q(f) = -2$.

4.2.3 Divisors, the Vector Space L(G) ,and the Theorem of Riemann-Roch

Let χ be a regular projective curve defined over $\mathbb{F}_q$. The free abelian group generated by the points of χ is called the *divisor group* of χ. The elements of this group are called *divisors* of χ. In other words, a divisor D of χ is a formal sum;

$$\sum_{P \in \chi} n_P P \tag{4.2.7}$$

where $n_P \in \mathbb{Z}$ and all but finitely many n_P's are zero. The degree of D is;

$$degD = \sum_{P \in \chi} n_P degP \tag{4.2.8}$$

where n_P is an integer equal to 0 for all but a finite number of points of χ. The *support* of D is defined by $supp(D) := f \in \chi : n_p \neq 0$. Two divisors $D = \sum_{P \in \chi} n_p P$ and $D' = \sum_{P \in \chi} n'_P P$ are added in the natural way

$$D + D' := \sum_{P \in \chi} (n_P + n'_P) P \tag{4.2.9}$$

The zero element of the group divisor is $\sum_{P \in \chi} n_P P$ with $n_P = 0$ for any $P \in \chi$. It will be denoted by 0.
A partial ordering on the group divisor can be defned by $D \leq D' \Leftrightarrow n_P \leq n'_P$ for any $P \in \chi$. If $n_p \geq 0$ for any $P \in \chi$ we call D *positive* or *effective*. The degree of D is the sum of all integers n_P , that is;

$$deg(D) = \sum_{P \in \chi} n_p \tag{4.2.10}$$

We will mainly be concerned with a subgroup of the divisor group. A $\mathbb{K}$-*divisor* is a divisor $D = \sum_{P \in \chi} n_P P$ such that $n_P = n'_P$ whenever $P' = \alpha(P)$ with α in the Galois group of $\bar{\mathbb{K}}$ over $\mathbb{K}$, $\bar{\mathbb{K}}$ being the algebraic closure of $\mathbb{K}$. Note that any divisor whose support is contained in the set of $\mathbb{K}$-rational points of χ is a $\mathbb{K}$-divisor. The set of all $\mathbb{K}$-divisors is a subgroup of the group divisor, and it will be denoted by D_χ.

Definition 4.2.3.1. *Let $f \in \bar{\mathbb{F}}_q(\chi)$. The order of f at a point $P \in \chi$ is defined to be $\nu_P(f)$ where ν_P is the discrete valuation corresponding to the valuation ring $O_P(\chi)$. If $\nu_P(f) > 0$, f is said to have a zero at P, and if $\nu_P(f) < 0$, f is said to have a pole at P.*

Remark 4.2.3.2. *For the sake of simplicity, from now on by the word divisor we will mean a $\mathbb{K}$-rational divisor.*

Given a rational function f, it is natural to associate a divisor to f, in the following way via the valuation map:

$$(f) := \sum \nu_P(f) P$$

Such a divisor is the zero divisor if and only if $f \in \mathbb{K}$, otherwise a pole divisor. $f \in \mathbb{K}$ be written as a difference of two effective divisors i.e $(f) = (f)_0 - (f)_\infty$, where $(f)_0 = \sum_{\nu_P(f)>0} \nu_P(f) P$ is the *zero divisor* of f, and $(f)_\infty = \sum_{\nu_P(f)<0} -\nu_P(f) P$, is the *pole divisor* of f. We call (f) a *principal* divisor.

Definition 4.2.3.3. *Two divisors D and D' are called linearly equivalent if $D - D' = (f)$ for a rational function f.*

Definition 4.2.3.4. *If $G \in D_\chi$, we can define the Riemann-Roch Space associated with G by:*

$$L(G) = \{f \in \mathbb{F}_q(x) | (f) + G \geq 0\} \cup \{0\} \tag{4.2.11}$$

Which kind of functions are admissible in the space so defined? Notice that if

$$G = \sum_{i=1}^{r} n_i P_i - \sum_{j=1}^{s} m_j Q_j$$

with $n_i > 0, m_j > 0$, then $L(G)$ consist of all elements $f \in \mathbb{F}_q$ such that

- f has zeros of order $\geq m_j$ at Q_j, for $j = 1, \ldots, s$, and
- f may have poles only at the places $P_1, \ldots, P_r$, with the pole order at P_i being bounded by n_i, $(i = 1, \ldots, r)$.

Remark 4.2.3.5. *Let $G \in D_\chi$. Then;*

a). $f \in L(G)$ if and only if $\nu_P(f) \geq -\nu_P(G)$ for all P.

b). $L(G) \neq \{0\}$ if and only if there is a divisor G' is equivalent to G.

Lemma 4.2.3.6. *Let $G \in D\chi$. Then we have:*

a). $L(G)$ is a vector space over $\bar{\mathbb{F}}_q$.

b). If G' is a divisor equivalent to G, then $L(G)$ is isomorphic to $L(G')$ as a vector spaces over $\bar{\mathbb{F}}_q$.

Remark 4.2.3.7. *Notice that the divisor of a product of two functions is the sum of the respective divisors, $(f.h) = (f)(h)$, and the divisor of the sum of two functions satisfies $(f+h) \geq min\{(f)(h)\}$, i.e., the minimum coefficient is chosen, point by point.*

The maps

$$\varphi = \begin{cases} L(G) \to \mathbb{F}_q, \\ f \to fg. \end{cases}$$

$$\varphi' = \begin{cases} L(G') \to \mathbb{F}_q, \\ f \to fg^{-1}. \end{cases}$$

are $\bar{\mathbb{F}}_q-$ linear and are inverses of each other. Infact φ is an isomorphism from $L(G)$ to $L(G')$

$L(G)$ is a finite-dimensional vector space over $\mathbb{F}_q(x)$, its dimension is denoted $l(G)$ [3]. The theorem of Riemann says that there exists a nonnegative integer m such that for every divisor G of χ

$$l(G) \geq deg(G) + 1 - m \tag{4.2.12}$$

and the smallest nonnegative integer with this property is called the genus and is denoted by $g(\chi)$ or simply g. We now consider objects of the form fdh where f and h are rational functions, i.e. elements of $\mathbb{F}_q(\chi)$, such that the map which sends h to dh is a derivation. We denote the set of differentials on χ by $\Omega(\chi)$. We can also think of the zeros and poles of differentials. At every closed point P there exists a local parameter that is, a function u such that $\nu_P(u) = 1$, and for every differential ω there exists a function f such that $\omega = fdu$. The valuation $\nu_P(\omega)$ is now by definition $\nu P(f)$, so that ω has a zero of order ρ if $\rho = \nu_P(f) > 0$ and a pole of order ρ if $\rho = -\nu_P(f) > 0$. The divisor of ω is by definition $(\omega) = \sum \nu_P(u)P$. The divisor of a differential is called canonical and always has degree $2g - 2$.

In the same way as we have defined $L(G)$ for functions we now define the vector space $\Omega(G)$ with zeros and poles prescribed by G as;

$$\Omega(G) = \{\omega \in \Omega_\chi \mid \omega = 0 \quad or \quad (\omega) \geq G\}$$

One could then defined the genus as the dimension of the vector space of differentials without poles, that is, of, $\Omega(0)$, where 0 is the divisor with coefficient zero at every closed point. The dimension of $\omega(G)$ is called the *index of speciality* of G and is denoted by $i(G)$.

Remark 4.2.3.8. *The theory of differentials is key in understanding the structure of Riemann-Roch space. Understanding this theory requires much effort. We will not discuss this theory here. For more theory on differentials, one is referred to [10], [12] and [26].*

Theorem 4.2.3.9 (Riemann-Roch). *For a divisor G of a curve of genus g*

$$l(G) = deg(G) + 1 - g + i(G) \tag{4.2.13}$$

Furthermore, $i(G) = l(K - G)$ for all divisors G and canonical divisors K.

Moreover it is a consequence of the Riemann-Roch theorem we have;

Corollary 4.2.3.10. *For any divisor with $deg(G) > 2g - 2$*

$$l(G) = deg(G) + 1 - g$$

Proof. By Riemann-Roch Theorem we have $l(G) = deg(G) + 1 - g + l(K - G)$, where K is a canonical divisor. As $deg(G) > 2g - 1$ and $deg(K) = 2g - 2$, we have $deg(K - G) < 0$. Now $0 \neq f \in L(K - G)$ would mean f has atleast a zero but no pole which is impossible! Hence $L(K - G) = 0$ and $l(K - G) = 0$, and the corollary is provad.

Note 4.2.3.11. *In texts like [20],[26], it is shown that nonsingular curves χ which intersects the line at infinity in a single point, Q say, are important for the construction of Algebraic Geometry codes. Such codes are often referred to as one-point Goppa codes. However, two-point codes have been constructed and studies on them conducted [27]. For our purpose we will only consider the case of one-point codes.*

4.2.4 Counting Points on Curves

There are several approaches in counting the number of points on a curve. Finding these points explicitly may be impractical especially when the dimension of the field is large. Morover, what we may be interested in is whether these points are many or few. Thus we make use of a bound given by Hasse-Weil theorem.

Theorem 4.2.4.1 (Hasse-Weil Bound). *The number $N = N(F)$ of places of $\mathbb{F}_q$ of degree one satisfies the inequality;*

$$|N - (q + 1)| \leq 2gq^{1/2}.$$

Curves that attain this bound with equality are called *maximal* curves. An example of such a curve is the hermitian curve [5].

4.3 Algebraic geometry In Coding

4.3.1 Algebraic Geometry Codes

Recall that $L(G)$ is a $\mathbb{F}_q$−vector space for any rational divisor G on a curve defined over $\mathbb{F}_q$. Recall also that a linear code is simply a vector subspace of $\mathbb{F}^n$ for some positive integer n. But $L(G)$ is a vector space of functions, so it is not immediately a code. In this chapter, we show how the linear property of the Riemann-Roch spaces can be exploited to construct linear codes. Furthermore, we will use the Riemann-Roch theorem to determine the ranks and (designed) minimum distances of these codes. This highlights the importance of the Riemann-Roch theorem to the theory of AG-codes.

We first emulate the construction of RS codes for the case when the sequence of points is obtained from curves in an algebraic geometry.

Definition 4.3.1.1. *Let $P_1 \cdots P_n$ be n distinct $\mathbb{F}_q$-rational points of χ and let $G \in D_\chi$ such that $\nu_{p_i}(G) = 0$ for $i = 1, ..., n$. Let*

$$\begin{aligned} \Phi : L(G) &\longrightarrow \mathbb{F}_q^n \\ f &\longrightarrow (f(p_1), ..., f(p_n)) \end{aligned} \tag{4.3.1}$$

which is an $\mathbb{F}_q$-linear map. Set $D := P_1 + ... + P_n$.
*The **(Geometric)Goppa code** associated with D and G is $C_{D,G} := \Phi(L(G))$.*

Clearly, $C_{D,G}$ is a linear code since $L(G)$ is a vector space.

Lemma 4.3.1.2. *Let $k := dim(C_{D,G})$ and d be the minimum distance of $C_{D,G}$. Then*

i). $k = l(G) - l(G - D)$

ii). $d \geq n - deg(G)$

Proof. i). The map Φ is surjective from $L(G)$ to $C_{D,G}$. Then, by linear algebra, $k = l(G) - dim(ker(\phi))$. We claim that $\ker\phi = L(G - D)$, so that $k = l(G) - l(G - D)$

ii). Let $x = (f(P_1), \cdots, f(P_n))$ such that $w(x) = d$ (for a linear code, the minimum distance is the same as the minimum weight of a non zero codeword).
Then there exist (a codeword with zeros in $n - d$ positions) $n - d$ points, say $P_{i_1}, , \cdots P_{i_{n-d}}$, such that $f(P_{i_j}) = 0$, i.e. $\nu_{P_{i_j}}(f) \geq 1$. Then $f \in L(G-(p_{i_1}, \cdots p_{i_{n-d}}))$ and hence $deg(G) - (n - d) \geq 0$ from which the result follows.

□

proof of claim. It is clear that $f \in \ker\phi$ if and only if $f(P_i) = 0$ for $i = 1, 2, \ldots, n$. therefore $\nu_{P_i}(f) \geq 1$. So $(f) - \sum_{i=1}^{n} P_i \geq 0$. Since $f \in L(G)$ and f has a zero at each of the $P_i s$, we can deduce that $f \in L(G - \sum P_i) = L(G - D)$. Hence $\ker\phi$ being a subspace is given by; $\ker\phi = L(G - D)$, and so $dim\ \ker\phi = l(G - D)$. Thus the code so constructed can be regarded as an $[n, k, d]$ code with parameters k *and* d as in the lemma above, with the length $n = deg(D)$. This follows from the fact that, the points of D are rational, the length of the code n must then be the same as the degree of D. □

Proposition 4.3.1.3. *Let $C_{D,G}$ be a Goppa code with parameters k and d as above. Let g be the genus of the underlying curve.*

i). Let $n > deg(G)$, then $k = l(G)$. Furthermore, a generator matrix of $C_{D,G}$ is given by;

$$M := \begin{pmatrix} f_1(P_1) & \cdots & f_1(P_n) \\ \vdots & \vdots & \vdots \\ f_k(P_1) & \cdots & f_k(P_n) \end{pmatrix}$$

where $f_1, \cdots, f_k$ is an $\mathbb{F}_q-$ basis of $L(G)$.

ii). If $n > deg(G) > 2g - 2$, then $k = deg(G) + 1 - g$.

Proof. i). We have that $L(D - G) = \{0\}$ and hence the first part of (1) follows from lemma 4.3.1.2 and the Riemann-Roch theorem. To see that M is a generator matrix of $C_{D,G}$ we have to show that the rows $x_1, \cdots, x_k$ of M are $\mathbb{F}_q-$linearly independent. Suppose that $\sum_{i=1} a_i x_i = 0$ with $a_i \in \mathbb{F}_q$. Then $\sum_{i=1}^{k} a_i f_i(P_j) = 0$ for $j = 1, \cdots, n$. Then $\sum_{i=1}^{k} a_i f_i \in L(G - D)$ and so $a_i = 0$ for each i. This completes the proof of (1).

ii). The claim follows from (1) and Corollary 4.2.3.10.

□

We now consider the other construction of AG codes, the residue construction of Goppa codes. We did not develop the theory of differentials necessary for a proper account of the construction of such codes, however, we will resort to a definition which will still be sufficient for our purpose. It can be noted that the canonical construction of $C^*_{D,G}$ does not play a part in the description of the theory covered here. A more canonical construction of $C^*_{D,G}$ can be looked up in ([12], 138), from it follows most of our arguments as well as results unproven. We make use of a the fact that $C^*_{D,G}$ is the dual code of the code $C_{D,G}$ already described.

Definition 4.3.1.4. *Let D and G be as before. Then*

$$C^*_{D,G} := \left\{ (f_1, f_2, \cdots, f_n) \in \mathbb{F}^n \mid \sum_{i=1}^{n} f_i \varphi(P_i) = 0 \forall \varphi \in L(G) \right\}$$

is a linear code of length $n = d(D)$, *rank* $k = n - l(G) + l(G - D) \ldots *$ *and minimum distance* $d \geq d(G) - (2g - 2)$ *where* g *is the genus of* χ.

Clearly $n = d(D)$ as in the previous argument. Since $C^*_{D,G}$ is the dual of $C_{D,G}$, we have

$$dim C^*_{D,G} + dim C_{D,G} = n \tag{4.3.2}$$

By Lemma 4.3.1.2, $dim C_{D,G} = l(G) - l(G - D)$. Substituting this in Equation4.3.2 above we obtain the required result. For the minimum distance, it is clear that if $d(G) \leq 2g - 2$.Suppose $d < d(G) - (2g - 2)$. Let $C = (c_1, c_2, \cdots, c_n) \in C^*_{D,G}$ be a word of minimum weight. Consider the set I of indices of c where $c_i \neq 0$. Clearly, $|I| = d$, so we have,

$$|I| = d\left(\sum_{i \in I} P_i < d(G) - (2g - 2)\right)$$

which yields;

$$d\left(G - \sum_{i \in I} P_i\right) > 2g - 2$$

By Riemann-Roch we have;

$$\left(G - \sum_{i \in I} P_i\right) = d(G) - d + 1 - g$$

Now let $j \in I$, then we have;

$$l\left(G - \sum_{i \in I} P_i + P_j\right) = d(G) - (d - 1) + 1 - g > l\left(G - \sum_{i \in I} P_i\right)$$

So there exists $\varphi \in L\left(G - \sum_{i \in I} P_i + P_j\right)$ but $\varphi \notin L\left(G - \sum_{i \in I} P_i\right)$ Since $P_i \in supp(D)$ and $P_i \notin supp(G)$, we must have $\nu_{P_i} \geq 1$ forall $i \neq j$, implying that $\varphi(P_i) = 0$ for $i \neq j$. Similarly since $\nu_{P_j} < 0$, we must have $\nu(P_j) \neq 0$. By $D - \sum_{i \in I} P_i \leq G$, we have $\varphi \in L(G)$, and by the definition of $C_{D,G}$ we have

$$c_1\varphi(P_1) + c_2\varphi(P_2) + \cdots + c_n\varphi(P_n) = 0$$

If $i \in I$ then $\varphi(P_1) = 0$ and if $i \notin I$ then $c_i = 0$, so the above equation reduces to $c_j\varphi(P_j) = 0$. But $j \in I$ which implies $c_j \neq 0$, and this is a contradiction, since $\varphi(P_j) \neq 0$. Therefore we must have $d \geq d(G) - (2g - 2)$ if $d(G) > 2g - 2$.

Theorem 4.3.1.5. *Suppose* $d(G) > 2g - 2$. *The code* $C^*_{D,G}$ *has rank* $k = n - d(G) + g - 1 + l(G - D)$.

Proof. By Riemann-Roch, $l(G) = d(G) + 1 - g$ if $d(G) > 2g - 2$. Substitute into the equation (*) above, we get the result.

Definition 4.3.1.6 (Designed Minimum Distance, Minimum distance). *The designed minimum distance of $C^*_{D,G}$ is defined to be $d^* := d(G) - (2g-2)$. We sometimes denote d^* as $d^*(C^*_{D,G})$ to emphasise that the code is $C^*_{D,G}$. Define $t^* := \left\lfloor \frac{d^*-1}{2} \right\rfloor$, and let $d(C^*_{D,G})$ denote the true minimum distance of $C^*_{D,G}$.*

Remark 4.3.1.7. *The designed minimum distance is only useful if $d(G) > 2g-2$.*

We now looking at an example of code construction over a harmitian curve. We need a basis for the Riemann-Roch space constructed from a carefully chosen divisor. The construction of this basis requires a little more theory which is not presented here. More can be looked up in [9].

Example 4.3.1.8. *Consider the Hermitian Curve defined by $f = X^3 + Y^2Z + YZ^2$. It is non-singular with genus 1. It has 9 rational points and one point at infinity, $Q = [0:1:0]$. Consider $C^*(D, aQ)$ where D is the sum of all the rational points except Q. The code has designed minimum distance $d^* = a - (2g-2) = a$. So letting $a = 5$ will allow the correction of 2 errors. The codes has rank $8 - a + 1 - 1 = 8 - a = 3$ if $a = 5$. We have $L(5Q) = 1, x, y, x^2, xy$. Define $\mathbb{F}_4 := \mathbb{F}_2[\omega]$ where $\omega^2 + \omega + 1 = 0$. Let:*

$$P_1 = (0:0:1) \quad P_2 = (0:1:1) \quad P_3 = (1:\omega:1) \quad P_4 = (1:\omega^2:1)$$
$$P_5 = (\omega:\omega:1) \quad P_6 = (\omega:\omega^2:1) \quad P_7 = (\omega^2:\omega:1) \quad P_8 = (\omega^2:\omega^2:1) \quad (4.3.3)$$

The code $C^*_{D,5Q}$ has parity check matrix.

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x(P_1) & x(P_2) & \cdots & x(P_8) \\ y(P_1) & y(P_2) & \cdots & y(P_8) \\ x^2(P_1) & x^2(P_2) & \cdots & x^2(P_8) \\ xy(P_1) & xy(P_2) & \cdots & xy(P_8) \end{pmatrix}$$

which evaluates to;

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & \omega & \omega & \omega^2 & \omega^2 \\ 0 & 1 & \omega & \omega^2 & \omega & \omega^2 & \omega & \omega^2 \\ 0 & 0 & 1 & 1 & \omega^2 & \omega^2 & \omega & \omega \\ 0 & 0 & \omega & \omega^2 & \omega^2 & 1 & 1 & \omega \end{pmatrix}$$

4.4 Summary

Goppa codes prove to be the easiest to construct and implement as Algebraic Geometry readily provide tools for construction and analysis of codes. The Riemann-Roch Theorem, for instance, readily gives enough information as to the code length, the rank and the minimum distance are known. All that remains is to choose an appropriate curve, based

on theories deeply founded in Algebraic Geometry. The Structural properties of curves, mainly the number of rational points and genus qualifies or disqualifies a curve from being a preferable candidate for code construction. For example, curves with small genus and large numbers of points are of interest in constructing codes. The quantity $N_q(g)$, the largest number of points over $\mathbb{F}_q$ for any curve of genus g, has been studied by several authors (e.g., [13]) and the results find implications in coding theory.

Most recent contributions into the area coding theory have involved the search of new maximal curves [25]. Several authors have also studied construction of function fields via some restriction, and the end product has been the discovery of new codes [23]. The search for decoding algorithm has also formed a an opening for more research in coding theory as evident in [25]. All in all the theory of coding has become more a mathematical problem than a problem in information theory. Hermitian codes for example has been extensively studied and digital systems bearing it aplications may soon be in the market!

4.5 Conclusion

As a conclussion of this dissertation, we try to establish the relationship if any among some classes of codes already discussed. That;

$$\begin{aligned} Linear \quad Codes \supset Cyclic \quad Codes \supset BCH \quad Codes \\ \supset Goppa \quad codes \\ \supset Reed - Solomon \quad codes \end{aligned} \tag{4.5.1}$$

is obvious. What remains is to establish how the BCH codes and the Goppa codes relate. Consider the construction of the RS and extension to GRS in Section 4.1.1. We can view the code in a different manner by the following argument;

We know that the projective line of order q can be described by giving the points coordinates (x_1, x_2) where (x_1, x_2) and $(\lambda x_1, \lambda x_2)$ are the same point $P \in \mathbb{F}_q$. If $a(x, y)$ and $b(x, y)$ are homogeneous polynomials of the same degree, then it makes sense to study the rational function $a(x, y)/b(x, y)$ on the projective line (since a change of coordinates does not change the value of the fraction). We pick the point $Q := (1, 0)$ as a special point on the line. The remaining points have as coordinates $(0, 1)$ and $(\alpha^i, 1)$, $(0 < i < q - 1)$, where α again denotes a primitive element of $\mathbb{F}_q$. We now consider those rational functions $a(x.y)/y^l$ for which $l < k$ (and of course $a(x, y)$ is homogeneous of degree l). This is a vector space (say K) of dimension k. Looked at in this manner, the description of RS codes given above amounts to numbering the points of the line in some fixed order (say $P_0, P_1, \cdots, P_{q-1}$) and taking as codewords $(f(P_0, \cdots, f(P_{q-1})$, where f runs through the space K. The functions have been chosen in such away that we can indeed calculate their values in all the points P_i's; this is not so for Q. Thus we conclude based on our previous discussions that the simplest examples of algebraic geometry codes are Generalized

Reed-Solomon, constructed by replacing the projective line by a projective curve in some projective space.

We also once again revisit the BCH construction for clues of inter relationship. Let α be a primitive n^{th} root of unity in an extension field of $\mathbb{F}_q$, say $\mathbb{F}_{q^m}$. $n \mid q^m - 1$ and let $g(x) \in \mathbb{F}_q[x]$ be the polynomial of smallest degree with zeros

$$\left\{\alpha^i; i = 1, 2, \cdots 2t\right\}$$

for some integer $t \geq 1$. Let the degree of $g(x)$, referred to as the generator polynomial of the code, be $n - k$. The maximum number of distinct cyclotomic cosets of these elements is $2t$, each containing at most m elements. Then

$$\mathcal{C} = \{a(x)g(x) \mid deg(a(x)) < k, \quad a(x) \in \mathbb{F}_q[x]\}$$

is a BCH code of length n, dimension $k \geq n - 2tm$,, and minimum distance $d \geq 2t + 1$. By considering the polynomial;

$$h(x) = \prod_{i=1}^{2t}(x - \alpha^i) \in \mathbb{F}_{q^m}[x]$$

then the above code, with $g(x)$ replaced by $h(x)$ and the field of definition, $\mathbb{F}_q$ replaced by $\mathbb{F}_{q^m}$, is an RS code $\mathcal{C}$', of length n, dimension exactly k, and minimum distance exactly $d = 2t + 1$. The BCH code is then a subfield subcode of $\mathcal{C}$,' i.e.,

$$\mathcal{C} = \mathcal{C}' \cap \mathbb{F}_q^n$$

i.e., the set of all codewords in with all coordinates in the field $\mathbb{F}$. Therefore
BCH Codes $\subset$ GRS Codes.

Diagramatically, we can have the representation below (not to scale).

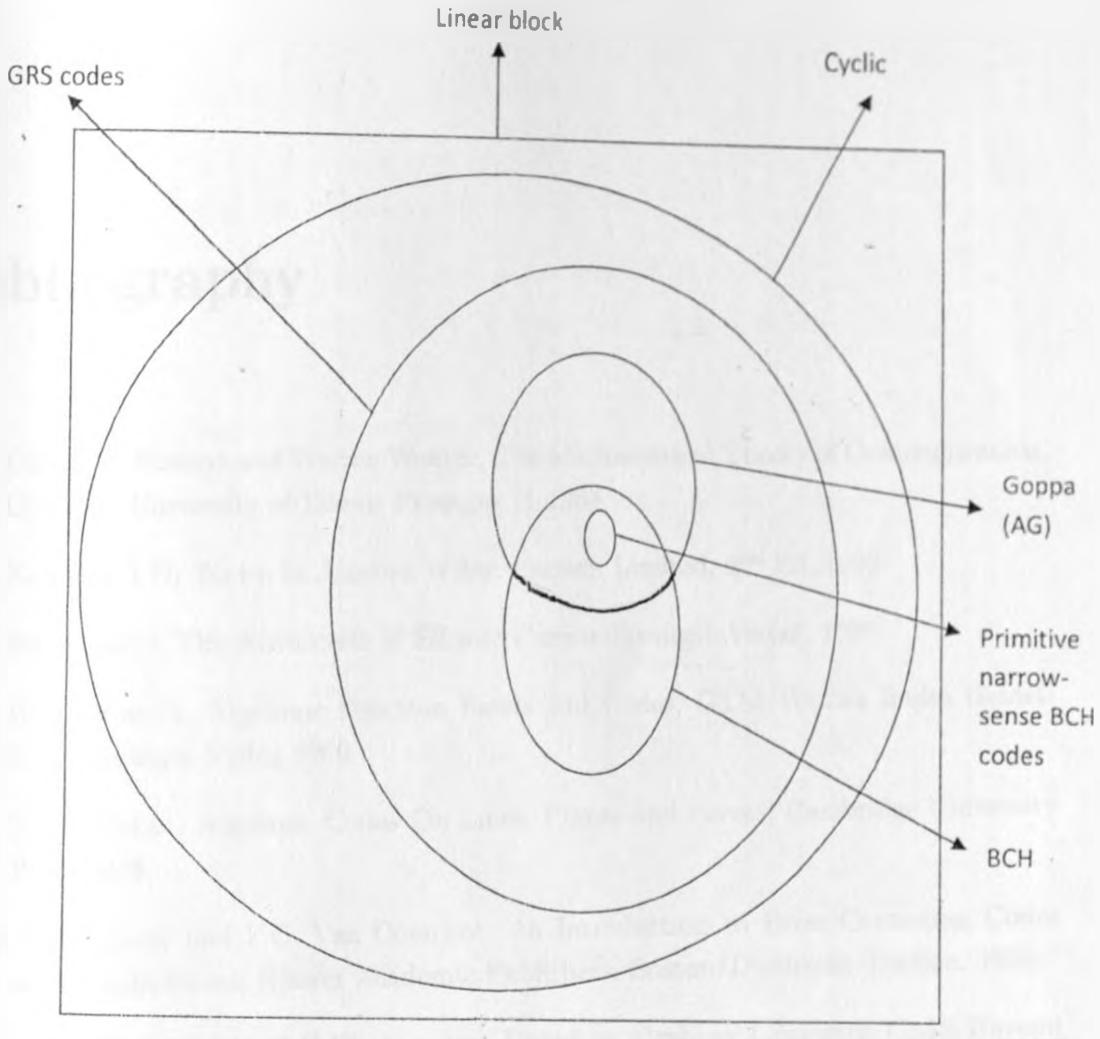

fig.4.1 Diagramatic representation of Inclussion of Some Codes.

Since we did not exhaust all the codes discovered so far, it would be interesting if one does a detailed work on codes thereby comming up with a diagramatic representation as above, possibly drawn to scale, and fitting in most if not all of the classes of codes.

Bibliography

[1] Claude E. Shanon and Warren Weaver, The Mathematical Theory of Communication, Chicago: University of Illinois Press,pp 71 1963.

[2] Herstein, I.N, Topics In Algebra Wiley, Eastern Limited, 2^{nd} Ed. 1993

[3] Silverman,J. The Arithmetic of Elliptic Curves. Springer-Verlag, 1986.

[4] H.Stichtenoth, Algebraic Function Fields and Codes, GTM Vol.254 Berlin Heidelberg: Springer Verlag 2009.

[5] R. E. Blahut, Algebraic Codes On Lines, Planes and curves, Cambridge University Press 2008.

[6] A. Vanstone and P.C. Van Oorschot, An Introduction to Error Correcting Codes with Applications, Kluwer Academic Publishers, Boston/Dordrecht/London, 1989.

[7] M. A. Shokrollahi and H.Wassermann, Decoding Algebraic Geometric Codes Beyond the Error Correction Bound, Berkeley, CA, 1997.

[8] F.J. Mac.Williams and N.J.A. Sloane, The Theory of Error-Correcting Codes, amsterdam .new york.oxford, 1993.

[9] Høholdt, T. van Lint, J.H. and Pellikaan R., Algebraic Geometry Codes, Handbook of Coding Theory, vol. 1, pp. 871-961 (Pless V.S., Huffman W.C. and Brualdi R.A. Eds.). Elsevier, Amsterdam, 1998.

[10] H. Niederreiter and C. Xing, Algebraic Geometry in Coding Theory and Cryptography, Princeton University Press,2009.

[11] V.S. Pless and W.C. Huffman, Handbook of Coding Theory Volume 1 Amsterdam, The Nethelands, Elsevier Science B.V. 1998.

[12] O. Pretzel, Codes and Algebraic Curves, New York. Oxford University Press, Inc.1998.

[13] V.G. Drinfeld and S.G. Vlădut, Number of points of an algebraic curve, Func. Anal., vol. 17, pp. 53-54, 1993.

[14] G.-C. Rota, Encyclopedia of Mathematics and its Applications, Finite fields, Cambridge University Press Vol 20, 1997.

[15] Ian Blake, Chris Heegard, Tom Hϕholdt and Victor Wei, Algebraic-Geometry Codes. IEEE Transactions on Information Theory, Vol. 44, No. 6, October 1998.

[16] V. D. Goppa, Codes on algebraic curves, Sov. Math.Dokl., vol.24. pp. 170-172, 1981. translation from Dokl. Akad. Nauk S.S.S.R., vol. 259, pp. 1289-1290, 1981.

[17] Reed I.S and G. Solomon, Polynomial Codes over Certain Finite Fields, Journal of the Society for Industrial and Applied Mathematics, Volume 8, 1960.

[18] V.D. Goppa, Codes with divisors. Probl. Inform. Transm., vol. 13, pp. 22-27, 1977.

[19] W. Wesley Peterson E.J. Weldon, Jr, Error-Correcting Codes, Second Edition, 1972.

[20] A. Tsfasman and S.G. Vlădut, Algebraic-Geometric Codes,The Netherlands: Kluwer, 1991.

[21] R.C. Singleton, Maximum Distance Q-nary Codes, IEEE Transactions on Information Theory, Volume IT-10, pp. 116-118,1964.

[22] Michael Purser, An Introduction to Error-Correcting Codes, Norwood, MA 02062, 1995.

[23] C. Xing and S Ling Yeo, New Linear Codes and Algebraic Function Fields Over Finite Fields IEEE Transactions on Information Theory, Volume 53, NO. 12, December, 2007.

[24] San Ling Chaoping Xing,Coding Theory; A First Course, Cambridge University Press, 2004.

[25] S.Fanali and M.Giulietti; One-Point AG Codes on the GK Maximal Curves, IEEE Transactions on Information Theory, Volume 56, NO. 1, January, 2010.

[26] J.H. van Lint, Introduction to, Coding Theory Third Edition, Springer-Verlag Berlin Heidelberg 1999.

[27] Iwan,Duursma and Radoslav; Improved Two-Point Codes on Hermitian Curves, IEEE Transactions on Information Theory, Volume 57, NO. 7, July, 2011.