Error-Correcting Codes Over Galois Rings

by

Gregory Reid Holdman

A thesis submitted in partial fulfillment of the requirements
for graduation with Honors in Mathematics.

Whitman College
2016

Certificate of Approval

This is to certify that the accompanying thesis by Gregory Reid Holdman
has been accepted in partial fulfillment of the requirements for
graduation with Honors in Mathematics.

_____

Patrick Keef, Ph.D.

Whitman College
May 11, 2016

Contents

Abstract

The theory of error-correcting codes has historically been most useful in the context of linear codes. Such codes may be viewed as vector spaces over Galois fields carrying with them many familiar and well-studied properties. A generalization of Galois fields is the concept of Galois rings. It is therefore natural to consider codes over Galois rings to study which properties such codes maintain in the move to a more general setting. This thesis will present two separate expositions on coding theory and Galois rings. After this, the intersection of these topics will be considered: codes over Galois rings.

List of Figures

1 Introduction

Coding theory began with an issue with which many of us are all too familiar: losing our work on the computer. In 1947, while working at Bell Laboratories, Richard W. Hamming had sporadic access to a computer [1]. While running a program, the computer had the inconvenient quirk of skipping to the next program upon the detection of an error. Many years after pioneering the development of error-correcting codes, Hamming recalled its inception:

> Two weekends in a row I came in and found that all my stuff had been dumped and nothing was done...And so I said, 'Damn it, if the machine can detect an error, why can't it locate the position of the error and correct it?' [1]

As a problem-solver, Hamming decided to put an end to this by developing coding theory. His idea was to add digits to the binary numbers in such a way that would add redundancy to the message. That way, if a message read by the computer had an error or two, it would still be possible to know what the intended message was.

When communicating, information is inevitably lost in the process of transmission. People in conversation on a busy street find it hard to hear each other. They may raise their voices to ensure their words are heard correctly. Or perhaps a painting may degrade over time. Curators make attempts to preserve the colors. It may be impossible to get rid of these losses entirely and so we perform actions that can at least guard against them. Coding theory is a mathematical discipline with the same goal in mind, but applied to strings of symbols.

A model for an application of coding theory is shown in Figure 1. Device 1 has a binary message to send to device 2. An encoder alters that message in some way to guard against loss of information. During transmission, it experiences some noise that flips one of the digits. The decoder on the other end is still able to determine what the original message was. Codes such as this one are designed in a rigorous way that guarantees that a single error in transmission can be corrected.

Coding theory uses mathematics to add redundancy in a manner much more efficient than naive approaches. Such a naive approach would be to send a message multiple times. For example, suppose one wishes to send the message "HAMMING." Suppose also that

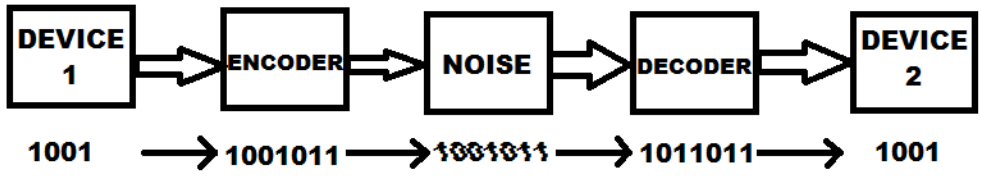

Figure 1: A model for how a code can make it easier to transmit messages over noisy channels.

during transmission of this message that one of the letters is transmitted incorrectly. Perhaps the message is received as "HUMMING." One may add redundancy by sending a message many times. It is unlikely that A will be the letter that is transmitted incorrectly every time, let alone that it will switch to U again. Maybe the second time we receive the message "RAMMING." The last five letters were the same both times, so those are probably correct. The first two were different each time, but it is much less likely that both of them transmitted incorrectly at the same time. From this, we can be fairly certain that the message is "RUMMING" or "HAMMING." A few more transmissions will make this more clear.

This method of adding redundancy is costly and ineffective. It requires that we use many times more computer power and memory to have a good chance of sending our message correctly. The genius of Hamming and others was to develop methods that made it much more likely to transmit a readable message while using less than twice the energy and information it would take to send the message alone. These savings are huge compared to the naive approach.

Historically, codes have been defined using a certain type of number system called a *field*. More specifically, codes use finite fields, also called *Galois fields*. This has been

wildly successful because of the close relationship between fields and with the well-studied concept of a *vector space*. Every electronically transmitting device you own uses a code which is defined using a Galois field. In particular, the most common Galois field in use is $\mathbb{Z}_2$ which is the binary field consistsing of 0 and 1. Recent work has shown that it is possible to extend codes to more general types of number systems called *rings*. A field is a special type of ring, and a direct generalization of finite fields are *Galois rings*. These types of rings have proven useful of late in defining codes over rings.

In this thesis, we will first present separate expositions on coding theory and Galois rings. Section 3 will cover the theory of error-correcting codes focusing on concepts that have been shown to generalize to Galois rings. Section 4 will build the theory of Galois rings leading up to the Classification of Galois Rings. Neither of these sections requires knowledge of the other, so they may be read in either order. After presenting these concepts, we will then cover the connections between these two topics.

2 Review of Algebra

Throughout this thesis, we will assume that the reader is familiar with undergraduate abstract algebra. This includes the definition of a group, normal subgroup, ring, ideal, and field among other related concepts and results. We will not distinguish between left and right ideals because all rings considered will be commutative. Despite this assumption of the reader's knowledge, this section will include some of the more relevant results both to remind the reader and refer back to later. Proofs from this section can be found in any standard undergraduate textbook on algebra such as Gallian's [3].

Lets first recall ideas related to maximal ideals. If R is a ring with ideal M, then M is called *maximal* if for any other ideal I of R such that $M \subseteq I \subseteq R$ we have $I = M$ or $I = R$. The following result will feature prominently in this thesis:

Theorem 1. *If R is a ring and M is an ideal of R then M is maximal if and only if R/M is a field.*

A ring need not have a multiplicative identity, called a *unity*, but if it does, the ring is called a *unitary ring*. A unitary ring has *units* which are elements with multiplicative

inverses. We point out here the important and often used fact that if a unitary ring R has an ideal I that contains a unit u, then necessarily $I = R$. This is easily shown from the definition of an ideal. We know $(u^{-1})u = 1 \in I$. Since $1 \in I$, then for all $r \in R$ we have $r1 = r \in I$ proving $R \subseteq I$ and hence $I = R$.

Throughout this thesis, the ideal generated by $a \in R$ will be denoted $(a) = \{ra | r \in R\}$. Note that this is in contrast to $\langle a \rangle = \{k \cdot a | k \in \mathbb{Z}\}$ which denotes the subgroup generated by a. Suppose R is a unitary ring. We will use the notation n to denote the element of R that is given by n summands of the element 1. That is, $n = n \cdot 1 = 1 + 1 + ... + 1$ with n terms. This will be important when we speak of the ideal generated by an element p where p is prime, denoted (p).

An important theorem that the reader ought to remember is the First Isomorphism Theorem for Rings. We include it here for reference.

Theorem 2. *Let ϕ be a ring homomorphism from the ring R to the ring $\bar{R}$. Then the mapping from $R/\ker\phi$ to $\phi(R)$, given by $a + \ker\phi \mapsto \phi(a)$, is an isomorphism. That is, $R/\ker\phi \cong \phi(R)$.*

Note that there is an analogous theorem for groups.

A definition with which the reader may not be familiar is that of a *local ring*.

Definition 1. If a ring has a unique maximal ideal, then it is called a *local ring* [4].

If a R is a local ring with maximal ideal M, then R/M is a field called the *residue field* of R. Galois rings are a specific type of local ring which we will cover later.

The characteristic of a ring is simply the smallest number of times any element of a ring added to itself will return 0. The *characteristic* of a ring with unity is the additive order of its unity. This is easily motivated by noting that for any element $a \in R$, if n is the additive order of 1, then $n \cdot a = a(n \cdot 1) = a(0) = 0$. Finite fields, also known as *Galois fields*, are rings with unity, and in particular they have the following property.

Theorem 3. *The characteristic of a Galois field is a prime p.*

On top of that, a Galois field always has prime power order p^k, where p is its characteristic and $k \geq 1$. It turns out that we can classify the Galois fields up to isomorphism.

Theorem 4. *For each prime p and each positive integer k, there is, up to isomorphism, a unique finite field of order p^k denoted $GF(p^k)$.*

Sometimes for brevity we will simply use $GF(q)$ to refer to the finite field of order q. The reader should always remember that q must be a power of a prime.

Perhaps we wish to have a field of order p^k with which we could perform operations. We can construct these fields as quotients of polynomial rings. We first look at the polynomial ring $\mathbb{Z}_p[x]$. For $f(x) \in \mathbb{Z}_p[x]$, the ideal $(f(x))$ is maximal if and only if $f(x)$ is irreducible over $\mathbb{Z}_p$ [3]. If $f(x)$ is irreducible over $\mathbb{Z}_p$ and of degree k, then $\mathbb{Z}_p[x]/(f(x))$ is a field of order p^k.

A fact that will be used briefly in the Classification of Galois Rings is as follows.

Proposition 1. *Suppose $\alpha \in GF(p^k)$. Let $u(x) \in \mathbb{Z}_p[x]$ be the minimal polynomial of α over $\mathbb{Z}_p$. Then $u(x)$ does not have repeated roots.*

Proof. Recall that for any polynomial over any field, $u(x)$ has repeated roots if and only if it shares a root with its derivative $u'(x)$. We will therefore show that this is the case.

Suppose by contradiction that $u(x)$ were to share the root α with its derivative. Let $u(x) = u_0 + u_1 x + ... + u_\ell x^\ell$. Then $u'(x)$ has smaller degree, yet α is still a root, so we must have $u'(x)$ be equivalently zero. Since $u(x) \in \mathbb{Z}_p[x]$, it must be the case that

$$u(x) = u_0 + u_p x^p + u_{2p} x^{2p} + ... + u_{jp} x^{jp},$$

where $jp \leq \ell$. It is easy to check that the derivative is identically zero. Now we know α is a root, so

$$0 = u(\alpha) = u_0 + u_p(\alpha)^p + u_{2p}(\alpha)^{2p} + ... + u_{jp}(\alpha)^{jp} = u_0 + u_p(\alpha^p) + u_{2p}(\alpha^p)^2 + ... + u_{jp}(\alpha^p)^j.$$

That is, α^p is the root of $v(x) = u_0 + u_p x + u_{2p} x^2 + ... + u_{jp} x^j$. However, α^p is the image of α under the Frobenius automorphism $\phi(a) = a^p$. Since automorphisms preserve degrees, we find a contradicion because α and α^p should have the same degree over $\mathbb{Z}_p$. $\square$

As was stated previously, a more detailed and rigorous proof of any statements from this section can be found in Gallian's text [3] or any other undergraduate text on abstract algebra.

3 Coding Theory

3.1 Defining Codes

Any method of written communication uses sequences of symbols to convey a message. In English, we have 26 letters to choose from. Computers communicate with binary representations of numbers. But English and binary numbers are very different in their structure. We can not always "add" two English words to get a new one. The development of codes is an attempt to rigorize when certain mathematical structure is applicable to these strings of symbols and what the consequences are.

Codes can be defined to be very general in their structure. Suppose that when writing a codeword, we can choose from nonempty finite set of symbols $\mathcal{A}$. Then $\mathcal{A}$ is called an *alphabet*. A *word* over the alphabet $\mathcal{A}$ is simply a finite sequence of elements from $\mathcal{A}$. Let $\mathcal{A}^*$ be the set of all words over $\mathcal{A}$. Then we can define the notion of a *code*:

Definition 2. Let $\mathcal{A}$ be an alphabet and let C be a subset of $\mathcal{A}^*$. Then C is a *code* over $\mathcal{A}$ [4].

To be even more specific, a *q-ary code* is a code whose alphabet has cardinality q. The word q-ary is simply a generalization of the words binary and terniary. We may even stipulate that all words be of the same length n, in which case we have a code of length n. To gain some intuition for the definition of a code and to highlight its subleties, we include some examples.

Example 1. Let $\mathcal{E}$ be the standard English alphabet of 26 letters. Then the written English language is a 26-ary code over $\mathcal{E}$. The set of words {for, you, and, the, dog} is a code of length 3 over $\mathcal{E}$. Note that this sentence does not use every letter in the English language but rather the subset {f,o,r,y,u,a,n,d,t,h,e,g}$\subset \mathcal{E}$ making it a 12-ary code.

Example 2. Let $\mathcal{G}$ be the standard German alphabet. German uses the same letters as English and more, namely ß. Written symbolically, $\mathcal{E} \subsetneq \mathcal{G}$. The written English and German languages are both codes over $\mathcal{G}$ but German is not a code over $\mathcal{E}$.

Example 3. Let $\mathbb{Z}_2 = \{0, 1\}$ be an alphabet. Then the set of all strings of four 0s and 1s is a binary code of length 4 over $\mathbb{Z}_2$.

A little thought reveals that codes are not unique. That is, we could make a code that is "the same" in many different ways.

Example 4. If tomorrow every English speaker in the world agreed that the letter A should be used in place of the letter B and vice versa, then the English lbngubge would still aehbve exbctly the sbme wby, though it mby ae difficult to get used to.

Example 5. We could exchange the position of the symbols. Suppose we had the code

$$\{0000, 0001, 1000, 1001\}.$$

If we exchange the last two symbols in each codeword, that is, permute the coordinates, then we obtain

$$\{0000, 0010, 1000, 1010\}.$$

It is not hard to convince oneself that these are essentially the same thing. It simply does not matter when we write each digit. It turns out that codes with symbols exchanged or coordinates swapped have the same properties, and so we can define the notion of *code equivalence*.

Definition 3. Two codes are *equivalent* if we can obtain one from the other by

1. any permutation of the letters of the alphabet in any fixed coordinate,
2. any permutation on the coordinate positions.

It is not hard to show that this is an equivalence relation and that we get equivalence classes of codes. In practical applications, we simply use the class representative that best fits the task at hand.

As examples 1 and 2 demonstrate, the definition of a code is general enough that its symbols need not have a mathematical structure of their own. For example, there is no binary operation that "combines" the letters A and B to get another letter, and thus no operation that combines any two words to produce another word. This gives us the freedom to add different types of structure, which can add to the power of coding theory. Taking example 3, we do indeed have some sort of structure imposed on the symbols: addition modulo 2. We will see that we are able to combine these codewords as vectors and find useful results. In this thesis we will cover some of the more important ways of adding structure to codes. We will view them as vector spaces which requires the use of Galois fields. We will also touch on recent work that looks at codes over Galois rings, which gives a different type of structure.

3.2 Introducing Linear Codes

In this section, we discuss the most developed and most commonly utilized codes: linear codes over Galois fields.

Definition 4. Let $V = GF(q)^{(n)}$ be the vector space of dimension n over $GF(q)$. Then a *linear code* C is a subspace of V. The code C has some dimension $k \leq n$ and the dimension n of V is also referred to as the *length* of C [2].

In the context of linear codes, there may be some confusion on the use of the words "codeword" and "vector". A "vector" is simply any element of the larger space V while a "codeword" is an element of the subspace, or code, C.

Example 6. View the code in example 3 as a vector space where each digit is a component of a vector (e.g. we may add 0001 and 1001 as $(0,0,0,1) + (1,0,0,0) = (1,0,0,1)$). Then we have a linear code over $\mathbb{Z}_2$. We could also take the subspace $C = \{(0,0,0,0), (0,0,0,1), (1,0,0,0), (1,0,0,1)\}$. This would also be a linear code.

What is particularly useful about linear codes is that, being a vector space, they have a basis. Suppose we only had the basis elements $(0,0,0,1)$ and $(1,0,0,0)$ of C from example 6. Then we know that $(1,0,0,1)$ is also a codeword in C because of closure under addition. In

general, if we store all basis elements of a linear code, then we can find all other codewords simply by taking linear combinations of the basis vectors. As this shows, if we have a k-dimensional binary code, then instead of storing all possible 2^k codewords, we need only store the k basis elements. This provides massive savings on storage space for large k.

A linear code C of dimension k is always a subspace of a vector space V of dimension n. The number n may also be called the *length* of a linear code. An (n, k) *code* is a linear code of length n and dimension k. It is important to note that in the future, saying a code is (n, k) will imply that it is linear.

Listing the basis vectors of a linear code is so useful that the list has its own name. In fact, it is both a list and a matrix:

Definition 5. [2] Let C be an (n, k) code with basis vectors $\mathbf{a}_1$, $\mathbf{a}_2$,...,$\mathbf{a}_k$. The *generator matrix* is the $k \times n$ matrix

$$G = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_k \end{pmatrix}.$$

Lets develop an example by Pless [2] to follow along with this definition. The Hamming $(7, 4)$ code is a binary linear code whose generator matrix can be given by

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

Note that according to the definition of code equivalence, we could permute the columns or exchange 0 and 1 in any of the columns. As written, we have a particularly nice form where $G = (I, A)$, I is the identity matrix and A a 4×3 matrix. In general an (n, k) code whose generator matrix has the form (I, A) where I is the $k \times k$ identity matrix and A is a $k \times n-k$ matrix is said to be in *reduced echelon form* [2]. Given any linear code, it is always possible to find an equivalent code that is in reduced echelon form. For the remainder of the thesis, we will generally only consider the reduced echelon form of a code.

In an (n, k) code, any k columns of the generator matrix which are linearly independent

are called the *information set* and their coordinates are the *information positions*. The rest of the coordinate positions are the *redundancy positions*[2]. Using this terminology, putting an (n, k) code in reduced echelon form is simply putting all of its information positions in the first k positions and reducing to the identity matrix.

Having the generator matrix of a code allows us to *encode* the messages. In the case of the Hamming $(7, 4)$ code, we have a 4-dimensional code suggesting we should be able to encode any 4-digit message. This is even more apparent when looking at the generator matrix. The four left positions are simply a single 1 in each possible position. We can add these basis vectors to get other codewords, so if we wished to encode, say, 1001, we could add the first column and fourth column to find the encoded message $1000011 + 0001111 = 1001100$. Notice that our original message is in the first four digits. We now have an encoded message that we could send. Later, we will look at how to obtain the original message in a process known as *decoding*.

Compared to nonlinear codes, linear codes make it particularly easy to encode messages. In the case of nonlinear codes, we would need to search a list for the codeword with our message in the information set. Any practical code is rather large in size, exponentially larger than the codes considered here, and so nonlinear codes are limited by the efficiency of search algorithms. We will see that linear codes have a similar advantage over nonlinear codes in decoding as well.

3.3 Parity Check Matrix

Consider a vector space V. Just as we are familiar with from geometry, the *inner product* of two codewords $\mathbf{v} = (v_1, v_2, ..., v_n)$ and $\mathbf{w} = (w_1, w_2, ..., w_n)$ in V is given by

$$\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{n} v_i w_i.$$

The two vectors are *orthogonal* or *perpendicular* if $\mathbf{v} \cdot \mathbf{w} = 0$. Note that in the case that the code is binary, the inner product simply counts the number of positions in which $\mathbf{v}$ and $\mathbf{w}$ both have a 1. A useful fact is the following:

Proposition 2. *The inner product is linear in both of its arguments.*

Proof. Well-known from courses on calculus or linear algebra. □

Now we can move on to defining the *dual code*.

Definition 6. Let C be an (n, k) code that is a subspace of V. The *dual code* of C is the set

$$C^{\perp} = \{\mathbf{x} \in V | \mathbf{x} \cdot \mathbf{y} = 0 \text{ for all } \mathbf{y} \in C\}.$$

Since the inner product is linear, the dual code is linear as well. That is, if $\mathbf{v}, \mathbf{w} \in C^{\perp}$ and $\mathbf{x} \in C$, then $(\alpha\mathbf{v} + \beta\mathbf{w}) \cdot \mathbf{x} = \alpha(\mathbf{v} \cdot \mathbf{x}) + \beta(\mathbf{w} \cdot \mathbf{x}) = \alpha(0) + \beta(0) = 0$, so any linear combination of $\mathbf{v}$ and $\mathbf{w}$ is also in $C^{\perp}$. If the dimension of the code C is k, then the dual code has dimension $n - k$ [2].

It turns out that the dual code of the dual code is the code itself, or in symbols, $(C^{\perp})^{\perp} = C$. Since $C^{\perp}$ has dimension $n - k$ and is linear, its dual code will have dimension $n - (n - k) = k$. In addition, it is self-evident that $C \subseteq (C^{\perp})^{\perp}$ because every vector in C is orthogonal to every vector in $C^{\perp}$. But C also has dimension k so it must be equal to $(C^{\perp})^{\perp}$.

Since the dual code is linear, we can determine a generator matrix. The generator matrix of the dual of a code will play a special role in decoding, so it has its own name.

Definition 7. Let C be an (n, k) code and let $C^{\perp}$ be the dual code. A generator matrix H of $C^{\perp}$ is a $(n - k) \times n$ matrix called a *parity check matrix* [2].

In the case of the Hamming $(7, 4)$ code, a parity check matrix is

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

One may easily check that every row vector in H is orthogonal to every row vector in G as defined in Equation 1.

Another reason for writing a code in reduced echelon form is that it is much easier to calculate a parity check matrix. We have the following theorem:

Theorem 5. *[2] If an (n, k) code C has a generator matrix $G = (I, A)$ in reduced echelon form, then a parity check matrix of C is $H = (-A^t, I)$ where A^t is the transpose of A and I is the $(n-k) \times (n-k)$ identity matrix.*

Proof. We will calculate the inner product of the rows of G and H. Each of these calculations is done simultaneously if we simply consider the matrix product GH^t.

$$GH^t = \begin{pmatrix} I, & A \end{pmatrix} \begin{pmatrix} -A \\ I \end{pmatrix} = -A + A = 0$$

□

3.4 Quantifying the Quality of a Code

The goal of coding theory is to correct as many errors as possible, and so the most obvious measure of a code's quality is the number of errors it can correct. Practical considerations may force one to look at other features such as how many codewords a computer would need to store or how much computational power it takes to encode or decode. For now, we will only consider the number of errors that the code can correct.

First, we should define what "distance" is in a code. Let C be a code which is a subspace of V. For any such V, the *Hamming distance* d_H between two vectors is the number of coordinates which have different entries.

Definition 8. The *Hamming distance* between $\mathbf{v}, \mathbf{w} \in V$ is $d_h(\mathbf{v}, \mathbf{w}) = |\{i | v_i \neq w_i, 1 \leq i \leq n\}|$.

Example 7. If $u = (1, 0)$ and $v = (0, 0)$ are in the space $\mathbb{Z}_2^{(2)}$, we have $d_H(u, v) = 1$.

The title of "distance" is accurate in that d_H is a metric on the space of the code. Recall that a metric on a set X is a function $d\colon X \times X \to R^+ \cup \{0\}$ that satisfies the following properties:

1. $d(x, y) \geq 0$ for all $x, y \in X$ and $d(x, y) = 0$ if and only if $x = y$,
2. $d(x, y) = d(y, x)$ for all $x, y \in X$,
3. $d(x, y) \leq d(x, z) + d(y, z)$ for all $x, y, z \in X$.

Proposition 3. *The Hamming distance is a metric*

Proof. The domain and codomain satisfy the form of a metric since $\mathbb{N}$ is a subset of $\mathbb{R}$. We now prove the three properties:

1. The function d_H is a function to the non-negative integers, so the image is always greater than or equal to 0. In addition, $d_H(\mathbf{v}, \mathbf{w}) = 0$ if and only if $|\{i : 1 \leq i \leq m, v_i \neq w_i\}| = 0$ if and only if $v_i = w_i$ for all i if and only if $\mathbf{v} = \mathbf{w}$.

2. We have that $d_H(\mathbf{v}, \mathbf{w}) = |\{i : 1 \leq i \leq m, v_i \neq w_i\}| = |\{i : 1 \leq i \leq m, w_i \neq v_i\}| = d_H(\mathbf{w}, \mathbf{v})$.

3. Consider a third vector $\mathbf{u} \in C$. We prove the triangle inequality by looking at a coordinate at a time. Let $D(\mathbf{v}, \mathbf{w})$ denote the set $\{i : 1 \leq i \leq m, v_i \neq w_i\}$ such that $d_H(\mathbf{v}, \mathbf{w}) = |D(\mathbf{v}, \mathbf{w})|$. Consider coordinate i. The only way to have $d_H(\mathbf{v}, \mathbf{w}) > d_H(\mathbf{v}, \mathbf{u}) + d_H(\mathbf{w}, \mathbf{u})$ is if it is possible to have $i \in D(\mathbf{v}, \mathbf{w})$ but $i \notin D(\mathbf{v}, \mathbf{u}) \cup D(\mathbf{w}, \mathbf{u})$. Hence we will show that $i \in D(\mathbf{v}, \mathbf{w})$ implies $i \in D(\mathbf{v}, \mathbf{u}) \cup D(\mathbf{w}, \mathbf{u})$. If $i \in D(\mathbf{v}, \mathbf{w})$ then $v_i \neq w_i$. But the only way to have $i \notin D(\mathbf{v}, \mathbf{u}) \cup D(\mathbf{w}, \mathbf{u})$ is if $u_i = v_i$ and $u_i = w_i$ in which case $v_i = w_i$ contradicting our assumption. Hence the triangle inequality is proved.

□

The *minimum distance* of a code is the smallest nonzero distance between two codewords. Including distances of zero is meaningless. A code is nonempty and so it has a vector $\mathbf{v}$ for which $d_H(\mathbf{v}, \mathbf{v}) = 0$. If we included distances of zero, then every code would have minimum distance of zero. If an (n, k) code has minimum distance d, then it can be called an (n, k, d) code. The *weight* of a vector is its distance from the zero vector and the *minimum weight* of a code is the smallest weight of any nonzero codeword. Note that if we have a code D with $\mathbf{u}, \mathbf{v} \in D$, then $d_H(\mathbf{u}, \mathbf{v}) = |\{i | u_i \neq v_i\}| = |\{i | u_i - v_i \neq 0\}| = wt(\mathbf{u} - \mathbf{v})$. Of course, $\mathbf{u} - \mathbf{v}$ is not necessarily in the code, but if D is linear, it will be. This leads to the following proposition:

Proposition 4. *In a linear code C with minimum distance d and minimum weight w, we have $w = d$.*

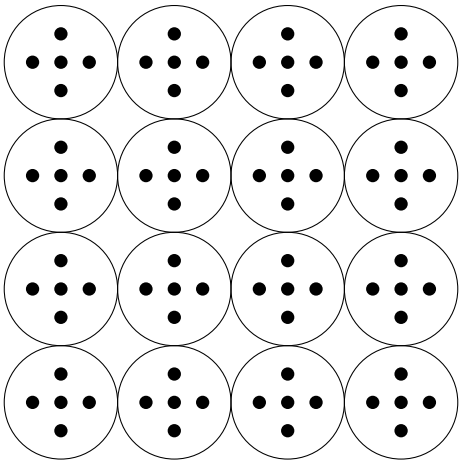

Figure 2: Visualization of a code. The vertices at the centers of circles are codewords while offset vertices are vectors in the larger space. By adding dimensions to the space, we were able to add vectors as a buffer between codewords.

Proof. The minimum weight cannot be less than the minimum distance because C is linear, meaning 0 is in the code. The minimum distance cannot be less than the minimum weight because $d_H(\mathbf{u}, \mathbf{v}) = wt(\mathbf{u} - \mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in C$. Hence $d = w$. □

Consider Figure 2. This is a visualization of a code. We took all of our messages that we wished to send and embedded them in a space with larger dimension so that we could fit more vectors around them. Receiving any vector in a particular circle means we know that the original codeword sent was the one at the center of that circle. This assumes only one error occurred, which is not a bad assumption since if the probability of one error occurring is p, then the probability of two errors occurring is p^2. Note that to move from one codeword to another requires 3 steps. That is, this figure represents a code of minimum distance 3. Note also that moving more than one vertex away from the center of a circle puts us in another. We enter the neighborhood of another codeword. Apparently a minimum distance (or weight) of 3 gives the ability to correct 1 error. Lets make this idea more rigorous. We have this theorem given by Pless in [2] to help us determine the number of errors a code can correct.

Theorem 6. *[2] If d is the minimum weight of a code C, then C can correct $t = \lfloor (d-1)/2 \rfloor$ or fewer errors.*

Proof. This proof is given by Pless [2]. First, define a sphere about a vector $\mathbf{u}$ in the space

V to be $S_r(\mathbf{u}) = \{\mathbf{v} \in V | d_H(\mathbf{u}, \mathbf{v}) \leq r\}$. We will show that spheres of radius t are disjoint.

For a contradiction, suppose these spheres were not disjoint. Then there are code words $\mathbf{u}$ and $\mathbf{w}$ such that $S_t(\mathbf{u}) \cap S_t(\mathbf{w}) \neq \emptyset$. Let $\mathbf{v} \in S_t(\mathbf{u}) \cap S_t(\mathbf{w})$. Then $d_H(\mathbf{u}, \mathbf{w}) \leq d_H(\mathbf{u}, \mathbf{v}) + d_H(\mathbf{v}, \mathbf{w}) \leq 2t$ by the triangle inequality proved in Proposition 3. This implies that $d_H(\mathbf{u}, \mathbf{w}) \leq 2(d-1)/2 = d - 1$. So $d_H(\mathbf{u}, \mathbf{w}) < d$. But $d_H(\mathbf{u}, \mathbf{w}) = wt(\mathbf{u} - \mathbf{w}) \geq d$ (this last statement follows because C is linear, so $\mathbf{u} - \mathbf{v} \in C$). This contradiction implies that the two spheres are disjoint. Since they are disjoint, of all the words in the sphere the code word at the center is the only code word in the sphere at all. □

This theorem tells us how many errors a code can correct, but not how it corrects them. This will be the subject of the next section on decoding.

Example 8. Lets determine how many errors the Hamming $(7, 4)$ code can correct. To do this, we will calculate the minimum weight of the Hamming $(7, 4)$ code. From the generator matrix in equation 1, we can see that the all basis vectors have weight 3 or greater. That is, we know that some nonzero vectors of weight 3 exist, so $d \leq 3$. Now simple calculation shows that adding any two of the rows gives weight at least 3 because there will be two nonzero coordinates among the first four and a nonzero coordinate among the last three. Adding any three rows will certainly give a vector of weight at least 3 because there will be three nonzero coordinates among the first four. Finally, adding all four rows gives that the first four coordinates are nonzero. Since any nonzero vector in the code can be written as a sum just described, we know there are no codwords of weight less than 3, so $d = 3$. Thus, the Hamming $(7, 4)$ code can correct $t = \lfloor (3-1)/2 \rfloor = 1$ error.

3.5 Syndrome Decoding

We now know how many errors a code *can* correct, now lets see *how* it corrects these. The process described here is called *syndrome decoding*.

Let C be an (n, k) q-ary code and let V be the vector space of dimension n containing C. Let H be the parity check matrix with rows $\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_{n-k}$. We have the following definition:

Definition 9. [2] The **syndrome** of a vector $\mathbf{y} \in V$ is

$$\text{syn}(\mathbf{y}) = \begin{pmatrix} \mathbf{y} \cdot \mathbf{h}_1 \\ \mathbf{y} \cdot \mathbf{h}_2 \\ \vdots \\ \mathbf{y} \cdot \mathbf{h}_{n-k} \end{pmatrix}.$$

A priori there are q^{n-k} possible syndromes because $\text{syn}(\mathbf{y})$ has $n-k$ coordinate positions with q possible values. We can think of the syndrome as just multiplying $\mathbf{y}$ by the parity check matrix. Note that if $\mathbf{y} \in C$, its syndrome will be the zero vector. This easily follows from the definition of the parity check matrix. In addition, since the inner product is linear, so is the syndrome ($\text{syn}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\text{syn}(\mathbf{x}) + \beta\text{syn}(\mathbf{y})$).

Pless describes an elegant way of viewing linear codes [2]. Lets view C as a subgroup of V under addition. Note that V is Abelian under addition, so C is a normal subgroup of V. All our standard results about normal subgroups and their cosets apply [3]. This allows us to prove the following theorem.

Theorem 7. *[2] Two vectors in V have the same syndrome if and only if they are in the same coset of C. Moreover, all possible q^{n-k} syndromes exist.*

Proof. We first show that for $\mathbf{v} \in V$ the syndrome is zero if and only if $\mathbf{v} \in C$. The "if" is easy because all vectors in C are orthogonal to the vectors in $C^\perp$. To prove the "only if" statement, lets prove the contrapositive: if $\mathbf{v} \notin C$ then $\text{syn}(\mathbf{v}) \neq 0$. Let $\mathbf{g}_1, ..., \mathbf{g}_k$ be a basis for C. Assuming $\mathbf{v} \notin C$, we know then $\mathbf{v} \notin (C^\perp)^\perp$, therefore $\mathbf{v}$ is not orthogonal to every vector in $C^\perp$. It is therefore the case that $\text{syn}(\mathbf{v}) \neq 0$.

Now to prove the statement of the theorem. Suppose $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{e} + C$. Then $\mathbf{v}_1 = \mathbf{e} + \mathbf{x}_1$ and $\mathbf{v}_2 = \mathbf{e} + \mathbf{x}_2$ for $\mathbf{x}_1, \mathbf{x}_2 \in C$. But $\text{syn}(\mathbf{v}_1) = \text{syn}(\mathbf{e}) = \text{syn}(\mathbf{v}_2)$ proving the "if" direction. Now suppose $\mathbf{v}_1$ and $\mathbf{v}_2$ have the same syndrome. Then $0 = \text{syn}(\mathbf{v}_1) - (v_2) = \text{syn}(\mathbf{v}_1 - \mathbf{v}_2)$ showing $\mathbf{v}_1 - \mathbf{v}_2 \in C$. This is only true if $\mathbf{v}_1$ and $\mathbf{v}_2$ are in the same coset of C.

The code C has dimension k and therefore has q^k codewords. This follows from a simple counting argument of linear combinations of the basis vectors. Every element $\mathbf{v} \in C$ can be written in the form $v_1\mathbf{g}_1 + v_2\mathbf{g}_2 + ... + v_k\mathbf{g}_k$ where the $\mathbf{g}_i$ are basis vectors. Since the code is over a field of cardinality q, there are q^k choices for the coefficients. Now since C is a normal subgroup of V, which has cardinality q^n, there are $q^n/q^k = q^{n-k}$ cosets of C each

Coset	Vectors				Syndrome
C	$(0,0,0,0)$	$(1,0,1,0)$	$(0,1,1,1)$	$(1,1,0,1)$	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$
$\mathbf{e}_1 + C$	$(1,0,0,0)$	$(0,0,1,0)$	$(1,1,1,1)$	$(0,1,0,1)$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$
$\mathbf{e}_2 + C$	$(0,1,0,0)$	$(1,1,1,0)$	$(0,0,1,1)$	$(1,0,0,1)$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
$\mathbf{e}_3 + C$	$(0,0,0,1)$	$(1,0,1,1)$	$(0,1,1,0)$	$(1,1,0,0)$	$\begin{pmatrix} 0 \\ 1 \end{pmatrix}$

Figure 3: An example of a code with its cosets and their syndromes. Coset leaders are in the left column of vectors.

of which has a distinct syndrome by the above paragraph. Thus all syndromes exist. □

After one more definition we'll be ready to describe *syndrome decoding*.

Definition 10. Given a coset $\mathbf{u} + C$, a *coset leader* of the coset is the vector $\mathbf{e}$ of smallest weight in the coset [2].

Note that it is possible to have multiple coset leaders. See the example later in this section.

On to syndrome decoding as described by Pless [2]. We first list all of the q^{n-k} syndromes along with their coset leaders. To do this, we simply take all vectors of weight 1 in V and compute their syndromes. Once this is completed, we move on to vectors of weight 2, only keeping them if their syndrome has not yet shown up. We continue for larger weights until we find all q^{n-k} syndromes for some vector. This will get the vectors of smallest weight for each syndrome and thus will get all coset leaders. Now with these syndromes and coset leaders stored in a list, we receive a vector $\mathbf{y}$ and compute its syndrome. As Theorem 7 tells us, we will then know which coset the received vector is in. We then take the coset leader of that coset $\mathbf{e}$ and decode to $\mathbf{x} = \mathbf{y} - \mathbf{e}$. How do we know that $\mathbf{y} - \mathbf{e}$ is in C? It is clear if that if $\mathbf{y}$ is in the coset $\mathbf{e} + C$, we can uniquely express it as $\mathbf{y} = \mathbf{e} + \mathbf{x}$ for $\mathbf{x} \in C$. Computing $\mathbf{x} = \mathbf{y} - \mathbf{e}$ is simply a rewriting of this equation. Since $\mathbf{e}$ is the vector of minimum weight that can be used to express such a sum, it is sometimes called the *error vector*.

We will now perform an example of syndrome decoding. Suppose we have the code

$$C = \{(0,0,0,0), (1,0,1,0), (0,1,1,1), (1,1,0,1)\} \subseteq \mathbb{Z}_2^{(4)},$$

as seen in Figure 3. A generator matrix is easily seen to be

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

From Theorem 5, we know that a parity check matrix is

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

In the notation from the definition of syndromes, we have $\mathbf{h}_1 = (1,1,1,0)$ and $\mathbf{h}_2 = (0,1,0,1)$. Figure 3 shows the cosets with syndromes calculated from these row vectors. The coset leaders are in the left-hand column of vectors.

Now suppose we receive the vector $\mathbf{y} = (1,1,1,0)$. Looking at Figure 3, we can just locate it on the list and subtract the coset leader, but remember that this is a consequence of using a small code. For large codes it is much less practical to keep all of the vectors and search the list. We will instead decode $\mathbf{y}$ by calculating its syndrome.

$$\text{syn}(\mathbf{y}) = \begin{pmatrix} \mathbf{y} \cdot \mathbf{h}_1 \\ \mathbf{y} \cdot \mathbf{h}_2 \end{pmatrix} = \begin{pmatrix} (1,1,1,0) \cdot (1,1,1,0) \\ (1,1,1,0) \cdot (0,1,0,1) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

So $\mathbf{y}$ is in the coset $\mathbf{e}_2 + C$ where $\mathbf{e}_2 = (0,1,0,0)$. Thus, we decode $\mathbf{y}$ to $\mathbf{y} - \mathbf{e}_2 = (1,0,1,0)$.

3.6 Perfect Codes

Recall from Theorem 6 that a linear code C, where C is a subspace of a larger vector space V, of minimum weight d can correct $t = \lfloor (d-1)/2 \rfloor$ errors. That is to say, spheres of radius t about each codeword are disjoint in the vector space V. Now it may be the case that these spheres contain all vectors in the space or that some lie outside. This leads to the following definition:

Definition 11. A code $C \subseteq V$ of minimum weight d is *perfect* if the union of all spheres about its codewords of radius t (t defined as above) is the vector space V [2].

Perfect codes are, as the name suggests, the best codes possible. Every received vector can be decoded to a codeword in C. Granted if there are too many errors, it will not decode to the correct codeword, but the probability of this occurring is small compared to a single error. If a received vector does not lie in a sphere, we need to decide in some other manner how best to decode it. We will not cover that topic here.

For an (n, k, d) code there is a simple constraint on the parameters n, k, and d for such a perfect code to exist. Instead of constraining d, we will think of constraining t, but it is nearly equivalent as d and t are related.

Theorem 8. *[2] In order for a perfect, t-error-correcting binary (n,k) code to exist, the numbers n, k, and t must satisfy the relation*

$$\left(\binom{n}{0} + \binom{n}{1} + ... + \binom{n}{t}\right) 2^k = 2^n.$$

Proof. First, let C be a perfect, t-error-correcting, binary (n, k) code. Since C is linear, it contains the zero vector. Let $S_t(0)$ be the sphere of radius t about the zero vector. The number of vectors of weight w, where $w \leq t$, in $S_t(0)$ is easily seen to be $\binom{n}{w}$ because there are that many distinct ways to arrange the w 1s in the n coordinates. So there are $\binom{n}{0} + \binom{n}{1} + ... + \binom{n}{t}$ vectors in $S_t(0)$. Now C is a linear space, so adding $\mathbf{c} \in C$ to each element in $S_t(0)$ gives $S_t(\mathbf{c})$. By linearity, $S_t(\mathbf{c})$ has the same cardinality. Since C is t-error-correcting, the spheres of radius t about the codewords are disjoint. At the same time, since C is perfect we know that their union is all of V. So we add the cardinalities of the 2^k disjoint spheres, equate it to the cardinality of V and find the relation stated in the theorem. □

An analogous theorem for q-ary codes in general is possible to prove as well.

Theorem 9. *[2] In order for a perfect t-error-correcting (n, k) code over $GF(q)$ to exist, the numbers n, k and t must satisfy the following equation*

$$\left(\binom{n}{0} + (q-1)\binom{n}{1} + ... + (q-1)^t\binom{n}{t}\right) q^k = q^n.$$

Some perfect codes are trivial. One example of a *trivial perfect code* is when the code is the entire space. This code cannot correct any errors because every possible codeword is in the code.

There are famous examples of nontrivial perfect codes. One example is an infinite family of binary, single-error-correcting codes, the *general binary Hamming codes*.

Definition 12. For each $r \in \mathbb{N}$ the *general binary Hamming code* H_r is defined as the code whose parity check matrix has columns consisting of all nonzero r-tuples. Typically, the code is defined such that we order the columns numerically viewing the r-tuples as binary integers.

This definition is somewhat opaque at first glance, so here is an example.

Example 9. Take H_3. Its generator matrix is actually given by G in the first section. Without knowing that fact we can still find the code using Definition 12. The parity check matrix using the standard ordering is simply

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

Notice that in each column, the top row is viewed as the first digit of a binary number while the second and third represent the rest of the digits. Given the parity check matrix, the code is uniquely defined, so in practice, we know the code. To find the generator matrix G, we would simply calculate a parity check matrix of H.

Note that the length of H_r is $2^r - 1$ (i.e. H_r is a subspace of $GF(2)^{(2^r-1)}$). This is because there are that many nonzero vectors in the r-dimensional column space. Since the dimensions of the dual code and the code itself sum to the dimension of the larger space, we know that the dimension of H_r is $(2^r - 1) - r$. We showed in Example 8 that the minimum weight of H_3 is $d = 3$. In general, H_r has minimum distance 3 [2]. Overall, H_r is a $(2^r - 1, 2^r - r - 1, 3)$ code. This satisfies Theorem 8. After calculating $t = \lfloor (3-1)/2 \rfloor = 1$, we have

$$\left(\binom{2^r-1}{0}+\binom{2^r-1}{1}\right)2^{2^r-r-1}=2^{2^r-1} \implies (1+2^r-1)2^{2^r-r-1}=2^{2^r-1}.$$

Note that this does not prove in itself that H_r is perfect, only that it can be. It turns out, however, that all general binary Hamming codes H_r are perfect codes. This is also true of the *general q-ary Hamming codes*.

Definition 13. For each $r \in \mathbb{N}$ the *general q-ary Hamming codes* H_r over the Galois field $GF(q)$ is defined as the code whose parity check matrix has columns consisting of all nonzero r-tuples, excluding scalar multiples. Again, the convention is to order the columns so they count up from left to right [2].

We didn't need to worry about scalar multiples in the case of binary codes because such vectors have no nonzero scalar multiples other than themselves.

Example 10. Lets consider H_3 over $GF(3)$. There are $3^3-1=26$ nonzero vertical 3-tuples. We can multiply each by 1 or 2 to get a different nonzero scalar multiple, so overall there is a set of 13 nonzero vertical 3-tuples none of which are scalar multiples of the others. Choosing from each pair the 3-tuple representing the smaller number, we have that the parity check matrix of H_3 over $GF(3)$ is

$$H=\begin{pmatrix} 1 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

Since H is the generator matrix of the dual of H_r we see that the dual of H_r has dimension 3 meaning H_r itself will have dimension $(3^3-1)-3=23$.

In general, H_r over $GF(q)$ has length $n=(q^r-1)/(q-1)$ and is a perfect, single-error-correcting $(n, n-r, 3)$ code [2].

Another famous name in coding theory is Marcel Golay. He discovered perfect, *multiple*-error-correcting codes which we will discuss briefly now. Golay discovered a binary $(23, 12, 7)$ code [2]. Note that since the minimum weight is $d=7$, this is code can correct $t=$

$\lfloor (7-1)/2 \rfloor = 3$ errors. Now recall that for any (n, k) code, we may write a generator matrix of an equivalent code of the form $G = (I, A)$ where I is the $k \times k$ identity matrix and A is a $k \times n - k$ matrix. Golay also discovered a *ternary* $(11, 6, 5)$ code with generator matrix $G = (I, A)$ where

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 & 1 \\ 1 & 0 & 1 & 2 & 2 \\ 2 & 1 & 0 & 1 & 2 \\ 2 & 2 & 1 & 0 & 1 \\ 1 & 2 & 2 & 1 & 0 \end{pmatrix}.$$

We end with a nontrivial result on the existence of perfect codes.

Theorem 10. *[2] The only nontrivial multiple-error-correcting perfect codes are equivalent to either the binary (23, 12, 7) code or the ternary (11, 6, 5) code. The only nontrivial single-error-correcting perfect codes have the parameters of the Hamming codes.*

This theorem was proved after many years of work by Tietevainen, van Lint, and Pless [2].

3.7 Cyclic Codes

Cyclic codes are a very useful type of code because of their simple structure and easy representation as ideals of rings. Many familiar codes, like the Hamming codes, can be represented as cyclic codes [2]. The definition is as follows.

Definition 14. An (n, k) code C is a *cyclic code* if whenever $a = (a_0, a_1, ..., a_{n-1})$ is in C then so is $a' = (a_{n-1}, a_0, ..., a_{n-2})$ [2].

The element a' as defined above would be called the *first cyclic shift* of a [2].

Suppose the code C is over a field $F = GF(q)$. Lets correspond the element $(a_0, a_1, ..., a_{n-1})$ in C with the element $a(x) = a_0 + a_1 x + ... + a_{n-1} x^{n-1}$ in the factor ring $R_n = F[x]/A$ where A is some ideal of $F[x]$. Recall that $F[x]$ is a *principal ideal ring* [3]. By definition, this means we can represent A as $A = (f(x))$. Which $f(x)$ should we choose to find an ideal that gives us a cyclic code? We will see that there is an obvious choice. If we want to

use the polynomial operations to find the first cyclic shift, this means we want to move the coefficient a_0 to be the coefficient on x, a_1 to be the coefficient on x^2 and so on. Clearly we must multiply by x.

$$\begin{aligned} a(x)x &= (a_0 + a_1x + ... + a_{n-2}x^{n-2} + a_{n-1}x^{n-1})x \\ &= a_0x + a_1x^2 + ... + a_{n-2}x^{n-1} + a_{n-1}x^n \\ &= a_{n-1}x^n + a_0x + a_1x^2 + ... + a_{n-2}x^{n-1}. \end{aligned}$$

This result will correspond to $(a_{n-1}, a_0, ..., a_{n-2})$ only if we choose $x^n = 1$. That is, $x^n - 1 = 0$ in this ring, so all polynomials are viewed modulo $x^n - 1$. Hence $f(x) = x^n - 1$ and $R_n = F[x]/(x^n - 1)$. From now on we will use the notation $R_n = F[x]/(x^n - 1)$ making sure the field F is clear from context.

From the argument above, we know we can correspond a cyclic code C with some subset of R_n. It turns out that it is not just a subset, but an ideal. We have the following nice theorem from Pless [2] which we state without proof.

Theorem 11. *A subset S in R_n corresponds to a cyclic code C if and only if S is an ideal of R_n [2].*

Now R_n is also a principal ideal ring as well, so a cyclic code C will correpond to the ideal $(g(x))$ for some $g(x) \in R_n$. These next theorems give us some immediate information about $g(x)$. In the theorems, we state that C, by abuse of terminology, not only corresponds to an ideal of R_n, but is itself an ideal.

Theorem 12. *If C is an ideal of $R_n = F[x]/(x^n - 1)$, which by Theorem 11 makes it a cyclic code, let $g(x)$ be a monic polynomial of smallest degree in C. Then $g(x)$ is the unique monic polynomial of smallest degree and $C = (g(x))$ [2].*

We do not offer a proof here. But notice that if $g(x)$ and $h(x)$ were two monic polynomials of minimum degree, then the leading terms cancel in $g(x) - h(x)$. An ideal is a subring, so this polynomial is in the ideal and has smaller degree than either contradicting the minimality of their degree.

Table 1: Showing the correspondence between divisors of $x^3 - 1$, ideals, and codes.

Divisor of $x^3 - 1$	Associated Ideal	Associated Code
1	R_3	$\mathbb{Z}_2^{(3)} = \{(0,0,0),(1,0,0),...,(1,1,1)\}$
$x+1$	$(x+1)$	{(0,0,0),(0,1,1),(1,0,1),(1,1,0)}
x^2+x+1	(x^2+x+1)	{(0,0,0),(1,1,1)}
x^3-1	$(0) = \{0\}$	{(0,0,0)}

Theorem 13. *If C is an ideal of R_n then its unique monic generator $g(x)$ divides $x^n - 1$. Conversely, if $g(x) \in C$ and it divides $x^n - 1$, then $g(x)$ has the lowest degree in $(g(x))$ [2].*

Together, these theorems determine all cyclic codes. From Theorem 11 we know all cyclic codes are ideals of R_n. Theorem 12 says we can correspond an ideal (and hence a code) with a monic polynomial in R_n. Finally, Theorem 13 says exactly which monic polynomials generate ideals, namely those that divide $x^n - 1$.

We know there are no other ideals (and hence no other cyclic codes) for suppose $f(x)$ were monic and did not divide $x^n - 1$. Would $(f(x))$ be an ideal? Yes, it would by its definition, and hence it would correspond to a cyclic code. But by Theorems 12 and 13 it would not be of minimum degree in the ideal. We could rename it $(f(x)) = (g(x))$ in $(f(x))$ where $g(x)$ is of minimum degree and by Theorem 13 does divide $x^n - 1$. For this reason, we call $g(x)$ the *generator polynomial* of the code it generates [2].

Lets determine all binary cyclic codes of length 3. From Theorem 11, we know these correspond to ideals of $R_3 = \mathbb{Z}_2[x]/(x^3-1)$. Now lets determine all polynomials that divide x^3-1. We can easily do this if we factor x^3-1. Note that $x^3-1 = x^3+1 = (x+1)(x^2+x+1)$. The quadratic factor is irreducible over $\mathbb{Z}_2$ because it is of degree 2 and has no roots in $\mathbb{Z}_2$. Table 1 gives each polynomial dividing $x^3 - 1$, its associated ideal of R_3, and its associated code. Assume a product of no factors is 1.

The next theorem is useful and should be fairly intuitive.

Theorem 14. *[2] If C corresponds to $(g(x))$ where $g(x) = g_0 + g_1x + ... + g_{n-k}x^{n-k} \in R_n$ and the degree of $g(x)$ is $n - k$, then the dimension of C is k and a generator matrix is*

$$\begin{pmatrix} g_0 & g_1 & g_2 & ... & g_{n-k} & 0 & 0 & ... & 0 \\ 0 & g_0 & g_1 & ... & g_{n-k-1} & g_{n-k} & 0 & ... & 0 \\ & \vdots & & & & & & & \\ 0 & 0 & 0 & ... & & & & & g_{n-k} \end{pmatrix}.$$

Proof. This is equivalent to showing that the vectors $g(x)$, $g(x)x$, $g(x)x^2$,...,$g(x)x^{k-1}$ in R_n are linearly independent and span C. Suppose they were not linearly independent. Then there is a linear combination of these vectors with some nonzero coefficients which is equal to zero: $a_0(g(x)) + a_1(g(x)x) + ... + a_{k-1}(g(x)x^{k-1}) = (a_0 + a_1x + ... + a_{k-1}x^{k-1})g(x) = 0$. But the degree of $g(x)$ is $n-k$, so the product is a polynomial of degree $k-1+n-k = n-1$, so the polynomial cannot be 0 modulo $x^n - 1$ unless all a_i are 0.

To see that the vectors span, recall that the code is an ideal generated by $g(x)$, so any polynomial can be written $c(x)g(x) = c_0g(x) + c_1xg(x) + ... + c_{k-1}x^{k-1}g(x)$. □

Finally, we move onto the dual code of a cyclic code. Note that if $g(x)$ is the generator polynomial of some code C then $x^n-1 = g(x)h(x)$ for some $h(x)$ because $g(x)$ divides x^n-1. Now $h(x)$ is called the *check polynomial* of C, although it does not necessarily generate the dual code [2]. But a polynomial closely related to it does. This is the *reciprocal polynomial* of $h(x)$ given by $x^kh(x^{-1})$ where k is the degree of $h(x)$.

In Table 1 we saw that $x + 1$ generates a code and $x^3 - 1 = (x + 1)(x^2 + x + 1)$, so the polynomial $x^2[(x^{-1})^2 + (x^{-1}) + 1] = 1 + x + x^2$ generates the dual code. In this case, the reciprocal polynomial is equal to the original, but this is not always so. For example, if we had the polynomial x^3+x+1, then its reciprocal would be $x^3((x^{-1})^3+x^{-1}+1) = 1+x^2+x^3$.

The polynomials of minimum degree are not the only generators of the ideals. There is an easier way to find generators that does not involve factoring x^n-1. It involves *idempotent* elements. An element a of a ring R is idempotent if $a^2 = a$. Lets review these elements in the context of polynomials.

Example 11. Consider the ring $\mathbb{Z}_2[x]/(x^3-1)$. Elements of this ring look like $a_0+a_1x+a_2x^2$ where $a_i \in 0, 1$ and all polynomials are viewed modulo x^3-1. We can recall that this means

$x^3 = 1$. Then the element $x^2 + x + 1$ is idempotent.

$$\begin{aligned}(x^2 + x + 1)^2 =& x^4 + x^3 + x^2 + x^3 + x^2 + x + x^2 + x + 1\\ =& x^4 + 2x^3 + 3x^2 + 2x + 1\\ =& x(x^3) + x^2 + 1\\ =& x^2 + x + 1\end{aligned}$$

Within an ideal, an idempotent element has nice properties.

Lemma 1. *Let R be a ring with ideal I and idempotent element $a \in I$. The element a is a generator of I if and only if it acts as a unity in I.*

Proof. Suppose a is a generator of I. Let $c \in I$. Then $c = ba$ for some $b \in R$. Now $ca = (ba)a = b(a^2) = ba = c$. So for elements of I, a acts like the unity.

Suppose a acts as the unity of I and $c \in I$. To show that $I = (a)$, we must show that $c = ba$ for some b. But this is simple as $c = ca$ because a is a unity in I. Thus a generates I. □

We will now take a short detour to cyclotomic cosets. In the construction of finite fields, we find a concept that naturally occurs.

Definition 15. Consider a number s such that $0 \leq s \leq p^m - 1$. Let r be the smallest number such that $p^{r+1}s = s \pmod{p^m - 1}$. The *cyclotomic coset* of s is $\{s, ps, p^2s, ..., p^r s\}$ where the elements are viewed modulo $p^m - 1$ [2].

In determining the cyclotomic cosets, the following is useful.

Proposition 5. *The cyclotomic cosets partition the set $\{0, 1, 2, ..., p^m - 1\}$.*

Proof. Define the relation $\sim$ on the set such that $x \sim y$ if $x \equiv p^k y \pmod{p^m - 1}$ for some integer k where $0 \leq k \leq m - 1$. We will show that this is an equivalence relation which will imply that it partitions the set into equivalence classes. It is reflexive because $x \equiv p^0 x$. It is symmetric because if $x \sim y$, then $x \equiv p^k y$ which implies $p^{m-k}x \equiv p^{m-k}p^k y \equiv p^m y \equiv y$. Congruence is symmetric so $y \equiv p^{m-k}x$ which means $y \sim x$. Finally, it is transistive. If

$x \sim y$ and $y \sim z$, then $x \equiv p^k y$ and $y \equiv p^l z$. Then $x \equiv p^k(p^l z) \equiv p^{k+l} z$. Note that $k+l$ may be greater than $m-1$ but not as large as $2m-2$. If $k+l > m-1$, we simply factor out p^m and note that $p^m \equiv 1 \pmod{p^m - 1}$. So $x \equiv p^{k+l} z \equiv p^m p^{k+l-m} \equiv p^{k+l-m} z$ meaning $x \sim z$. This proves that $\sim$ is an equivalence relation implying that the cyclotomic cosets partition the set. □

Example 12. Let $p = 2$ and $m = 3$. We will compute all cyclotomic cosets for $2^3 - 1 = 7$. Note that $2 \cdot 0 = 0$. Therefore the cyclotomic coset for 0 is $\{0\}$. This will always be the case. Now, looking at 1, we see that $2 \cdot 1 = 2$, $2^2 \cdot 1 = 4$, and $2^3 \cdot 1 = 1$. Referring to the definition, the corresponding r for 1 is $r = 2$. Thus 1 has cyclotomic coset $\{1, 2, 4\}$. Now by Proposition 5 the cyclotomic cosets partition the set, so we need not calculate the coressponding cosets for 2 and 4. The final cyclotomic coset is easily seen to be $\{3, 2 \cdot 3, 2^2 \cdot 3\} = \{3, 6, 5\}$.

Usually, cyclotomic cosets are given by C_u where u is the smallest number in the coset.

Lets now assume that we are looking only at binary cyclic codes of odd length n. It turns out that this means $x^n - 1$ has distinct irreducible factors [2]. Another consequence is that it is easy to take squares of elements. Recall the following:

Lemma 2. *If R is a ring with prime characteristic p and $x, y \in R$, then $(x+y)^p = x^p + y^p$.*

Proof. By the binomial theorem, $(x+y)^p = x^p + \binom{p}{1} x^{p-1} y + ... + \binom{p}{p-1} x y^{p-1} + y^p$. Now $\binom{p}{k}$ is divisible by p for $1 \leq k \leq p-1$, so these terms are zero because the characteristic is p. This proves the lemma. □

By induction it is easy to show that for any finite sum of elements in R, taking the pth power is simply the sum of the pth powers of the summands. To motivate this, note $(x+y+z)^p = ((x+y)+z)^p = (x+y)^p + z^p = x^p + y^p + z^p$. This fact will be important when squaring polynomials. For if $f(x) = a_0 + a_1 x + ... a_k x^k$, then $f^2(x) = (a_0)^2 + (a_1 x)^2 + ... + (a_k x^k)^2$.

We now have a theorem that will tell us all idempotent generators.

Theorem 15. *A binary polynomial $f(x)$ is an idempotent in $R_n = \mathbb{Z}_2[x]/(x^n - 1)$ if and only if the set S of powers of x that occur with nonzero coefficients in $f(x)$ is a union of cyclotomic cosets for n [2].*

Before we begin the proof, lets attempt to understand the statement of the theorem. We know what a binary idempotent polynomial is, but it will be useful to get an idea of what it means for the set S of powers of x that occur with nonzero coefficients in $f(x)$ to be a union of cyclotomic cosets for n. Consider $n = 2^2 - 1 = 3$. The cyclotomic cosets are $\{0\}$ and $\{1, 2\}$. So if we take the polynomial $x + x^2 = (1)x + (1)x^2$ and square it, we find $((1)x + (1)x^2)^2 = (1)^2x^2 + (1)^2x^4 = x^2 + x$. We see that the coefficient on x changed to become the coefficient on x^2. In general terms a_i, the coefficient on x^i becomes the coefficient on x^{2i} upon squaring the polynomial. Remember that this is because in $\mathbb{Z}_2$ both elements are idempotents. In order for the polynomial to be idempotent, we need $a_i^2 = a_{2i}$ for all i. The polynomial must have the property that a_i nonzero implies a_{2i} is nonzero. Thus for any nonzero coefficient, we have the entire set $\{a_i, a_{2i}, a_{4i}...\}$ is nonzero elements. Now on to the formal proof.

Proof. For the forward direction, let $f(x) = a_0+a_1x+a_2x^2+...+a_{n-1}x^{n-1}$ be an idempotent. To get a contradiction, suppose the set of powers in question S is not a union of cyclotomic cosets. Then there is some nonzero $a_j = 1$ such that $a_{2j} = 0$ where $2j$ is viewed modulo n. Now since R_n has characteristic 2, $f^2(x) = a_0^2 + a_1^2x^2 + a_2^2x^4 + ... + a_{n-1}^2x^{2(n-1)}$. But $f^2(x) = f(x)$ by assumption, so in particular, $a_j^2 = a_{2j}$. But this implies $1^2 = 0$, which is not the case. So by contradiction S is the union of cyclotomic cosets of n.

Now suppose that S is the union of cyclotomic cosets. Using the same definition for $f(x)$, we see that this means if a_i is nonzero, then a_{2i} is nonzero, a_{4i} is nonzero, and so on so that the subscripts together form the cyclotomic coset of i. As before, squaring $f(x)$ gives $a_i^2 = a_{2i}$ for all i (again, $2i$ is viewed modulo n). But the only elements of $\mathbb{Z}_2$ are 0 and 1, and we have $0^2 = 0$ and $1^2 = 1$. Hence, $f(x)$ is idempotent. □

The argument that states that $a_j^2 = a_{2j}$ may be a little difficult to follow, so I will provide a short example.

Example 13. Let $n = 5$, and consider $f(x) = x + x^2$. That is, $a_1 = a_2 = 1$ and $a_0 = a_3 = a_4 = 0$. Note that in particular, $a_2^2 \neq a_4$. This alone shows that $f(x)$ will not be an idempotent, for we ought to have $f^2(x) = (x)^2 + (x^2)^2 = x^2 + x^4 = x + x^2$, but this is not the case.

Example 14. Recall that for $n = 7$, the cyclotomic cosets are $\{0\}, \{1, 2, 4\}$, and $\{3, 5, 6\}$. Looking at the cosets, they can correspond directly to the polynomials $x^0 = 1$, $x + x^2 + x^4$, and $x^3 + x^5 + x^6$ which are easily shown to be idempotent. We can also take the union of any of these cosets and use those powers to make polynomials, but this just corresponds to taking sums of the three polynomials we just wrote down. These idempotent polynomials are $1+x+x^2+x^4$, $1+x^3+x^5+x^6$, $x+x^2+x^3+x^4+x^5+x^6$, and $1+x+x^2+x^3+x^4+x^5+x^6$. Of course, we may also include the empty coset which corresponds to the empty sum giving us the final idempotent polynomial: 0. The "if and only if" in Theorem 15 means that these are the only idempotent polynomials in R_7.

4 Galois Rings

4.1 Defining Galois Rings and Examples

Galois rings are a very natural generalization of finite fields. Not only does each Galois ring contain finite fields, but many of the most important results about finite fields generalize to Galois rings. In particular, we can classify all Galois rings. In the case of finite fields, the classification is up to order. For Galois rings, we must add the stipulation that the characteristics be equal. Lets begin with the definition of a Galois ring.

Definition 16. A ring R is called *Galois* if it is finite, commutative, unitary, local and its maximal ideal is given by (p) where p is prime.

Here we begin to see the hints of why Galois rings are a generalization of finite fields. Not only are they finite, commutative, and unitary, but the element p plays an important role for Galois rings as it does for fields of p-power order. Just as $\mathbb{Z}_p$ is the canonical example of a finite field, the canonical example of a Galois ring is the ring $\mathbb{Z}_{p^n}$ with p prime and $n > 0$. We will prove that this is a Galois ring.

Proposition 6. *The ring $\mathbb{Z}_{p^n}$ is a Galois ring.*

Proof. Since this ring is obviously finite, commutative, and unitary, we must only show that (p) is the unique maximal ideal. Stipulating "unique" will implicitly prove that $\mathbb{Z}_{p^n}$

is local. Consider the homomorphism $\phi\colon \mathbb{Z}_{p^n} \to \mathbb{Z}_p$ that maps $a \bmod p^n \mapsto a \bmod p$. This is clearly onto as we can map the elements less than p to themselves. It is obvious that $\ker\phi = (p)$. Using the First Isomorphism Theorem for Rings gives $\mathbb{Z}_{p^k}/(p) \cong \mathbb{Z}_p$ which is a field. By Theorem 1, we see that (p) is maximal. To show unicity, suppose M were some other maximal ideal not equal to (p). We can first note that $p \notin M$. Suppose it were the case that $p \in M$. Then $(p) \subseteq M$ with $(p) \neq M$ meaning M properly contains (p). Since (p) is maximal, this would imply $M = \mathbb{Z}_{p^n}$. A maximal ideal is a proper ideal by definition, so this contradiction shows $p \notin M$. Now by Theorem 1, $\mathbb{Z}_{p^n}/M$ must be a field. Since M is a proper ideal, $1 \notin M$ and so $1 + M$ is nonzero. Since all fields have prime characteristic, in this case p, this implies $p \cdot 1 + M = M$ so that $p \in M$. This contradiction proves the unicity of (p). All together, we see that $\mathbb{Z}_{p^n}$ is a Galois ring. □

We saw that we can construct finite fields of order p^k from quotients of polynomial rings over $\mathbb{Z}_p$. In the same way, we will be able to construct Galois rings as quotients of polynomial rings over $\mathbb{Z}_{p^k}$.

4.2 Important Properties of Galois Rings

It turns out that a Galois ring with maximal ideal (p) must have characteristic p^k for some k. We will prove this soon, but first here is a lemma.

Lemma 3. *Let R be a finite, commutative, local ring with unique maximal ideal M. If I is a proper ideal of R, then $I \subseteq M$.*

Proof. Suppose I is not a subset of M. Then there is some $a \in I$ such $a \notin M$. Hence, $(a) \nsubseteq M$. If (a) is maximal, then this contradicts the unicity of M. There must be some ideal A_1 such that $(a) \subseteq A_1$. If A_1 is maximal, this again contradicts the unicity of M. Continuing inductively, there must always be a larger ideal. But R is finite, so this chain of ideals must be finite and there is some final A_r which again contradicts the unicity of M. Hence by contradiction $I \subseteq M$. □

We will now use this lemma to prove the following important theorem.

Theorem 16. *Let R be a finite, unitary, commutative, local ring. Then the characteristic of R is p^k for some prime p and positive integer k.*

Proof. Let us first decompose R as a direct sum. We know that under addition, R is a finite Abelian group, so by the Fundamental Theorem of Finite Abelian Groups we may write

$$R \cong \mathbb{Z}_{p_1^{k_{11}}} \oplus \mathbb{Z}_{p_1^{k_{12}}} \oplus ... \oplus \mathbb{Z}_{p_2^{k_{21}}} \oplus ... \oplus \mathbb{Z}_{p_m^{k_{ml}}}.$$

Now, since R is local, it has a unique maximal ideal M which by Lemma 1 means R/M is a field. Clearly this is a finite field, so its characteristic is some prime p. Referring back to the decomposition, it is clear that $p = p_i$ for some i. That is to say, we definitely have elements of order p in R. Suppose not. Then p does not divide $|R|$, so p does not divide $|R|/|M|$. But this is exactly the order of R/M which does have an element of order p, namely 1.

Lets now define $A \cong \mathbb{Z}_{p^{k_1}} \oplus \mathbb{Z}_{p^{k_2}} \oplus ... \oplus \mathbb{Z}_{p^{k_t}}$ and B to be isomorphic to all group summands that do not have orders of powers of p. In this way, we can write $R \cong A \oplus B$ where we know A is not empty. In order to show that the characteristic of R is p^k for some k, we only need to show that B is trivial.

Suppose that B is nontrivial. It is then easy to see that $A \oplus \{0\}$ and $\{0\} \oplus B$ are both proper ideals of R. By Lemma 3 then $A \oplus \{0\}, \{0\} \oplus B \subseteq M$. But since M is closed under addition, this implies that $A \oplus B = R \subseteq M$ which is clearly not the case. Hence, B is trivial and the only group summands of R are those with order a power of p. The characteristic of R is therefore some power of p. □

Note that this theorem does not mention Galois rings, only finite, unitary, commutative, and local rings. A Galois ring R is described by all of these adjectives but also has maximal ideal (p). Knowing p, we know from this theorem that the characteristic of R is p^k for some positive k. Additionally, we can see directly from the decomposition of R into direct sums as shown in the proof that the order of a Galois ring will also be a power of p. From now on, R will be a Galois ring and $M = (p)$ will be its maximal ideal.

Galois rings split nicely into two disjoint subsets: units and elements divisible by p. The units turn out to be exactly those elements which are not in the maximal ideal. In addition,

the multiples of p, elements on M, are nilpotent elements. We prove these statements in the next couple of paragraphs.

Proposition 7. *An element $x \in R$ is a unit if and only if $x \notin M$.*

Proof. In the forward direction, we'll prove the contrapositive. Suppose $x \in M$. We wish to show that x is not a unit. Clearly $(x) \subseteq M$. Hence for all $r \in R$, we have $rx \in M$. Now, $1 \notin M$ because otherwise M would not be a proper ideal. Therefore we cannot have a y such that $xy = 1$. In the other direction, suppose $x \notin M$. Then the ideal (x) is not a subset of M. By lemma 3 however, M contains all proper ideals of R. Hence we must have $(x) = R$. This implies that $1 \in (x)$ so there is a $y \in R$ such that $xy = 1$. □

Proposition 7 tells us about the units of a Galois ring. It turns out that nilpotent elements are important in the study of these rings as well. Recall that $a \in R$ is *nilpotent* if $a^k = 0$ for some $k \in \mathbb{N}$. It is not difficult to see that the nilpotent elements form an ideal. For let N be the set of all nilpotent elements. Then if $r \in R$ and $a \in N$, we have $(ra)^k = r^k a^k = r^k 0 = 0$ proving N is an ideal. As we proved in Lemma 3, $N \subseteq M$. We will now prove $M \subseteq N$ showing $N = M$.

Proposition 8. *The set of nilpotent elements in R is exactly the maximal ideal M.*

Proof. Let N be the ideal of nilpotent elements. We already know $N \subseteq M$. Now we show $M \subseteq N$. Since $M = (p)$, this is equivalent to showing that if an element of R is divisible by p, then it is also nilpotent. But this is simple. Recall that the characteristic of R is p^k for some k, so $p^k = 0$ showing p is nilpotent. Then if a is divisible by p, we find $a = bp$. This implies a is nilpotent because $a^k = (bp)^k = b^k p^k = b^k 0 = 0$. This completes the proof. □

Another useful fact is the following:

Proposition 9. *In any unitary ring R, if u is a unit and a is nilpotent, then $u + a$ is a unit.*

Proof. The inverse of $u+a$ is $(u^{k-1} - u^{k-2}a + ... + (-1)^{k-1}a^{k-1})(u^{-1})^k$ where k is such that

$a^k = 0$.

$$\begin{aligned}(u+a)(u^{k-1} - u^{k-2}a + ... + (-1)^{k-1}a^{k-1})(u^{-1})^k \\ &= (u^k + u^{k-1}a - u^{k-1}a + ... + (-1)^{k-2}ua^{k-1} + (-1)^{k-1}ua^{k-1} + (-1)^{k-1}a^k)(u^{-1})^k \\ &= (u^k + (-1)^{k-1}a^k)(u^{-1})^k \\ &= (u^k)(u^{-1})^k = 1.\end{aligned}$$

□

We will frequently consider polynomial rings over Galois rings, and so we need to study some properties. With proposition 9, we can prove a similar theorem regarding units in the polynomial rings over R. It is stated as follows.

Theorem 17. *Let $f(x) = a_0 + a_1x + ... + a_kx^k$ be a polynomial in $R[x]$. Then $f(x)$ is a unit in $R[x]$ if and only if $a_0 \notin M$ and $a_1, ..., a_k \in M$.*

Proof. Suppose $f(x)$ is a unit. Let ψ: $R[x] \to R[x]/M$ be the homomorphism taking the coefficients of polynomials to their corresponding cosets in R/M. Since R/M is a field, the units of $R[x]/M$ are exactly the constant polynomials $a+M$. For example, $(a+M)+(b+M)x$ is not a unit. Now if $f(x)$ is a unit in $R[x]$, then $\psi(f(x))$ is also a unit. This is because if $f(x)g(x) = 1$, then $1+M = \psi(1) = \psi(f(x)g(x)) = \psi(f(x))\psi(g(x))$. Using this fact, $\psi(f(x))$ being a unit in $R[x]$ means $\psi(f(x)) = \psi(a_0)+\psi(a_1)x+...+\psi(a_k)x^k+M = a_0+...+a_kx^k+M$ is constant and nonzero. Thus, $a_0 \notin M$ and the rest of the coefficients are in M.

We will now prove the other direction. Since $a_0 \notin M$, this means a_0 is a unit in R as well as in $R[x]$. The element $a_1x + ... + a_kx^k$ is clearly nilpotent in $R[x]$. Since $R[x]$ is a unitary ring, Proposition 9 implies $f(x) = a_0+a_1x+...+a_kx^k$ is a unit as well. This proves the theorem. □

In the next section, we will prove some more specific properties about polynomials over Galois rings that lead up to the Classification of Galois Rings.

4.3 Properties of Polynomials Over Galois Rings

Polynomials over Galois rings have some interesting properties that we will be able to exploit later when proving the Classification of Galois Rings. The theorem we wish to prove is stated as follows:

Theorem 18. *Let $f(x) = a_0 + a_1x + ... + a_nx^n$ be a polynomial in $R[x]$ where R is a Galois ring with maximal ideal M and characteristic p^k. Let $j \leq k$ be such that $a_j \notin M$ and $a_{j+1}, ..., a_k \in M$. Then $f(x)$ is an associate of a monic polynomial of degree j.*

Let's take a moment to consider what this theorem states. Here, a_jx^j is the highest degree term where $a_j \notin M$. Every term of degree higher than j has a nilpotent coefficient. The theorem says that for some unit polynomial $u(x)$, we have $f(x)u(x) = g(x)$ where $g(x)$ is monic and of degree j. In rings of polynomials defined over fields, unit polynomials are exactly the constant polynomials. However, since this polynomial ring is defined over a ring, it is possible to have unit polynomials which are not constants, i.e. degree greater than 0. Note also that Theorem 18 is a generalization of Theorem 17 in that if we take $j = 0$, we recover Theorem 17.

The general outline of the proof is by construction but is rather tedious. Defining some terms will make it easier to state and prove some lemmas.

Definition 17. For any polynomial in $R[x]$ define a *p-term* to be a nonzero term in that polynomial whose coefficient is divisible by p but not by p^2. Inductively, define a *p^k-term* to be a term whose coefficient is divisible by p^m but not p^{m+1}. A p^m-term of degree s is simply a term of the form rx^s where r is divisible by p^m but not p^{m+1}.

By the logic of the above definition, we could call a unit a 1-term, but there is no need to use a new word for a familiar concept. To gain intuition for these definitions, here is an example.

Example 15. Recall from Proposition 6 that $\mathbb{Z}_8$ is a Galois ring with maximal ideal (2). In the polynomial $1 + 6x + 4x^2 \in \mathbb{Z}_8[x]$, the term $6x$ is a 2-term of degree 1 and $4x^2$ is a 4-term of degree 2. There is no such thing as an 8-term in this case because such terms are zero.

Theorem 18 defines the polynomial $f(x)$ of which the highest degree of a term with coefficient not in M is degree j. Another term will be useful to define so that we do not have to always say "the term of highest degree with unit coefficient."

Definition 18. The *final unit term* will be the highest degree term of a polynomial with unit coefficient. The *final unit degree* will be the degree of the final unit term.

Now to move toward the proof, I will prove a series of lemmas.

Lemma 4. *Let $f(x)$ be defined as in Theorem 18 and let ℓ be the degree of the highest degree p-term. Then $f(x)$ is an associate of a polynomial $f_1(x) = b_0 + b_1x + ... + b_{n_1}x^{n_1}$, where $b_j \notin M$ and $b_{j+1}, ..., b_{n_1} \in M$, but the highest degree p-term has degree less than ℓ.*

Essentially this lemma says that we can constrain the highest degree p-term in $f(x)$ to a smaller degree in a way that preserves the final unit degree, in this case j.

Proof. From the statement of the lemma, $a_\ell x^\ell$ is the highest degree p-term. Consider the polynomial $f_1(x) = f(x)(\frac{a_\ell}{a_j}x^{\ell-j}) - f(x)$. Remember we can divide by a_j because $a_j \notin M$ so it is a unit. First we show that $b_j \notin M$. That is, we preserve the fact that the term of degree j has unit coefficient. If $\ell - j > j$, then $b_j = -a_j$ which is a unit. If $\ell - j \leq j$, then $b_j = \frac{a_\ell}{a_j}a_{2j-\ell} - a_j$ which by Proposition 9 is a unit (first term is nilpotent while a_j is a unit). Since b_j is a unit, $b_j \notin M$. We now show $b_m \in M$ for $m > j$. By the definition of j, for $m > j$ we have that b_m is divisible by p. Looking at $b_m = \frac{a_\ell}{a_j}a_{m-(\ell-j)} - a_m$ for $m > j$,we see that a_ℓ and a_m are both divisible by p. Since, $b_j \notin M$ and all terms of higher degree have coefficients in M, this proves that we have preserved the final unit degree.

We will now show that all p-terms in $f_1(x)$ are constrained to have degree strictly less than ℓ. First of all, the coefficient $b_\ell = (a_\ell/a_j)a_j - a_\ell = 0$ so the term of degree ℓ is not a p-term. Let m now be the degree of some nonzero term in $f_1(x)$. Then $b_m = \frac{a_\ell}{a_j}a_{m-(\ell-j)} - a_m$ where $a_m = 0$ if $m > k$ by convention. Consider the case that that $m - (\ell - j) > j$, or equivalently $m > \ell$. Then $a_{m-(\ell-j)}$ is certainly divisible by p, so $\frac{a_\ell}{a_j}a_{m-(\ell-j)}$ is divisible by p^2. In addition, since $m > \ell$ so we know that a_m is divisible by p^2. Thus b_m is divisible by p^2, so $b_m x^m$ is not a p-term . We have constrained our p-terms in $f_1(x)$ to have degree less than ℓ.

Finally we must show that $f(x)$ and $f_1(x)$ are associates. Rewriting $f_1(x)$ we have $f_1(x) = f(x)(\frac{a_\ell}{a_j}x^{\ell-j} - 1)$. By Proposition 17, $(\frac{a_\ell}{a_j}x^{\ell-j} - 1)$ is a unit. Hence $f_1(x)$ and $f(x)$ are associates. □

We can use this lemma to repeatedly constrain the degree of the highest degree p-term to smaller and smaller degrees until we reach j. That is, we are able to "get rid of" all p-terms with degree greater than the final unit degree all while preserving the final unit degree. This is the proof of the next lemma.

Lemma 5. *Let $f(x)$ be defined as in Theorem 18. Then $f(x)$ is an associate of a polynomial $g(x) = c_0 + c_1x + ... + c_tx^t$, where $c_j \notin M$ and $c_{j+1}, ..., c_t \in M$, but $g(x)$ has no p-terms of degree greater than j.*

Proof. By Lemma 4, $f(x)$ is the associate of a polynomial $f_1(x)$ with its highest degree p-term less than ℓ. Let $f_1(x) = f(x)u_1(x)$. Lemma 4 applies to $f_1(x)$ as well, so we can find an associate polynomial $f_2(x)$ that has its highest degree p-term at an even smaller degree with its term of degree j still having unit coefficient. Let $f_2(x) = f_1(x)u_2(x) = f(x)u_1(x)u_2(x)$. Of course, there are a finite number of degree between j and ℓ, so there is a finite sequence $\{u_1(x), u_2(x), ..., u_s(x)\}$ such that $g(x) = f_s(x) = f(x)u_1(x)u_2(x)...u_s(x)$. That is, we can constrain our p-terms to have lower and lower degree until we get rid of all p-terms of degree greater than j. □

It should be easy to see that we can generalize Lemmas 4 and 5 to higher powers of p. In the end, we would find the following lemma given without proof:

Lemma 6. *Let $f(x) = d_0 + d_1x + ... + d_rx^r$ be such that for $j \leq r$ we have $d_j \notin M$ but $d_{j+1}, ..., d_r \in M$. Suppose also that $f(x)$ has no p-terms, p^2-terms,..., or p^m-terms. Then $f(x)$ is an associate of a polynomial with final unit degree j and no p^{m+1}-terms of degree greater than j.*

We are essentially able to knock out higher and higher powers of p. Now we can prove Theorem 18.

Proof. Consider $f(x)$. By Lemma 5 (or equivalently by Lemma 6) it is the associate of a polynomial $g_1(x)$ with final unit degree j and which has no p-terms of degree greater than

j. Now by Lemma 6, $g_1(x)$ is the associate of a polynomial $g_2(x)$ with final unit degree j and no p^2 terms of degree greater than j. We may continue in this fashion. Now $f(x)$, $g_1(x)$, $g_2(x)$ and so on are defined over R which has characteristic p^k. Thus, $g_{k-1}(x)$ as defined in the above fashion will have no terms at all of degree greater than j which are divisible by p. And since j is still the final unit degree, we see that $g_{k-1}(x)$ has degree j with unit leading coefficient. Let e_j be this leading coefficient of $g_{k-1}(x)$. We finally, take $g(x) = e_j^{-1} g_{k-1}(x)$ to get a monic polynomial of degree j. We know that $f(x)$ and $g(x)$ are associates, so the theorem is proved. □

4.4 Lemmas for the Classification of Galois Rings

Just as we are able to classify all Galois fields by their order, we can also classify Galois rings by order and characteristic. This section will build up to a rigorous proof that this is the case. For the rest of the section, assume that R is a Galois ring with maximal ideal M and characteristic p^k.

We showed in the proof of Theorem 16 that as an Abelian group, $R \cong \mathbb{Z}_{p^{k_1}} \oplus \mathbb{Z}_{p^{k_2}} \oplus ... \oplus \mathbb{Z}_{p^{k_t}}$. We will now show that all k_i must be equal. That is, we will show that as an Abelian group under addition, R is isomorphic to the direct sum $\mathbb{Z}_{p^k} \oplus \mathbb{Z}_{p^k} \oplus ... \oplus \mathbb{Z}_{p^k}$. First, we have a couple of lemmas.

Lemma 7. *Suppose $y \in R$. Then there is a unique integer $j \in \{0, 1, ..., k\}$ such that y is an associate of p^j.*

Proof. If y is a unit, then from Proposition 7 we know that y is not divisible by p. Hence, $j = 0$ is the only exponent possible to have $y = yp^j$.

Now suppose $y = 0$. Then y is an associate of p^k because $up^k = 0$ for any unit u. There is no other value of j that works because up^j is nonzero for all units u and all j.

We may now suppose that y is not a unit and is nonzero. It is therefore divisible by p. We may thus write $y = u_1 p$. If u_1 is a unit, then we are done. If not, then u_1 is a nonzero nonunit like y and so $u_1 = u_2 p$. Plugging in for u_1 gives $y = u_2 p^2$. We find a sequence $\{u_\ell\}$ by continuing this process. If u_ℓ is never a unit for all $\ell \in \mathbb{N}$, then we find $y = u_k p^k = 0$. But we had assumed y was nonzero. We must therefore have some $j < k$ such that u_j is a

unit and $y = u_j p^j$.

To show that this j is unique, suppose $y = u_1 p^{j_1} = u_2 p^{j_2}$ for integers $j_1 < j_2$ and units u_1 and u_2. Then multiplying both sides by p^{k-j_2}, we have

$$u_1 p^{k+j_1-j_2} = u_2 p^{k-j_2+j_2} = u_2 p^k = 0.$$

But u_1 is a unit and $p^{k+j_1-j_2}$ is also certainly nonzero as $k + j_1 - j_2 < k$. Therefore we have a contradiction and it must be the case that $j_1 = j_2$.

□

From this lemma, lets define a notion of absolute value.

Definition 19. The absolute value of $y = up^j \in R$, where u a unit, is given by $|y| = j$ for $j < k$. If $j = k$, let $|y| = \infty$.

This definition makes sense because Lemma 7 gives a well-defined value of j for each element. Note that 0 is the only element with an absolute value of ∞. By convention we will say that adding ∞ to anything gives ∞.

Proposition 10. $|yz| = |y| + |z|$.

Proof. Let $|y| = j$ and $|z| = \ell$. Then $yz = (u_y)p^j(u_z)p^\ell = (u_y u_z)p^{j+\ell}$. Since $u_y u_z$ is a unit, we know yz is an associate of $p^{j+\ell}$. If $j + l < k$ then $|yz| = j + \ell$. If $j + \ell \geq k$ then $|yz| = \infty$. □

This definition of absolute value allows us to speak about ideals of Galois rings more easily. We can see this applied in the following lemma.

Lemma 8. *The chain of ideals $R = (1) \supseteq (p) \supseteq (p^2) \supseteq ... \supseteq (p^k) = \{0\}$ is a complete list of ideals of R.*

Proof. We will first show that any ideal in R can be written (p^j) where $j \in \{0, 1, ..., k\}$. If I is the trivial ideal, we know $\{0\} = (p^k)$. Now suppose I is a nontrivial ideal of R. This will guarantee that I contains an element of finite absolute value as 0 is the only element with infinite absolute value. Let y be an element of smallest absolute value in I, say $|y| = j$.

Since y is an associate of p^j, we have $(y) = (p^j)$. We know that $(p^j) = (y) \subseteq I$. All that remains is to show $I \subseteq (p^j)$. Suppose $a \in I$. By the definition of y, the absolute value of a is at least j, so $a = up^\ell$ where u is a unit and $\ell \geq j$. Thus, $a = (up^{\ell-j})p^j$ so $a \in (p^j)$. We have therefore shown that $I = (p^j)$. It is clear that $(p^j) \neq (p^{j+1})$ because $p^{j+1} \in (p^{j+1})$ but $p^j \notin (p^{j+1})$. This shows that the ideals listed are distinct. This completes the proof. $\square$

Note that this lemma also shows that Galois rings are Principal Ideal Rings .

Lemma 9. *As an Abelian group,* $R \cong \mathbb{Z}_{p^k} \oplus ... \oplus \mathbb{Z}_{p^k}$.

Proof. We begin by showing that for $j < k$, we have $R/(p) \cong (p^j)/(p^{j+1})$ as Abelian groups. This is a simple application of the First Isomorphism Theorem for Rings. Take the homomorphism $\phi\colon R \to (p^j)/(p^{j+1})$ given by $x \mapsto xp^j + (p^{j+1})$. It is clearly onto. To see that $\ker\phi = (p)$, note that $x \in \ker\phi$ if and only if $\phi(x) \in (p^{j+1})$ if and only if $xp^j + (p^{j+1}) = (p^{j+1})$ if and only if $xp^j \in (p^{j+1})$ if and only if $x \in (p)$. By the First Isomorphism Theorem for Rings, we have $R/(p) \cong (p^j)/(p^{j+1})$ as rings and therefore as Abelian groups under addition.

We showed in the proof of Theorem 16 that as an Abelian group, $R \cong \mathbb{Z}_{p^{k_1}} \oplus \mathbb{Z}_{p^{k_2}} \oplus ... \oplus \mathbb{Z}_{p^{k_t}}$. Let e be the minimum exponent of p in this direct sum and $f_1 \leq f_2 \leq ... \leq f_s$ be the rest of the powers not equal to e, though not necessarily distinct amongst themselves. Then

$$R \cong \mathbb{Z}_{p^e} \oplus \mathbb{Z}_{p^e} \oplus ... \oplus \mathbb{Z}_{p^e} \oplus \mathbb{Z}_{p^{f_1}} \oplus \mathbb{Z}_{p^{f_2}} .... \mathbb{Z}_{p^{f_s}}.$$

If the groups $\mathbb{Z}_{p^{f_i}}$ are indeed nontrivial, then it is clear that p^e is less than the characteristic of R. This is because we need only take the element that has 1 in the coordinate corresponding to $\mathbb{Z}_{p^{f_1}}$ and 0 elsewhere to find an element with additive order $p^{f_1} > p^e$. Now according to Lemma 8, we ought to have an isomorphism between $R/(p)$ and $(p^e)/(p^{e+1})$ given by $x + (p) \mapsto xp^e + (p^{e+1})$. However, if we take element $y = (1, 0, ..., 0) \in R$ then $y + (p)$ is certainly nonzero in $R/(p)$ because not all coordinates are divisible by p. Its image, however, is zero because

$$(1, 0, ..., 0) + (p) \mapsto (1, 0, ..., 0)p^e + (p^{e+1}) = (0, 0, ..., 0) + (p^{e+1})$$

Therefore, this so called isomorphism has a nontrivial kernel. By contradiction there can not be higher powers. So setting $k = e$, we have $R \cong \mathbb{Z}_{p^k} \oplus ... \oplus \mathbb{Z}_{p^k}$. □

Now that we know that a Galois ring of characteristic p^k is isomorphic to the direct sum $\mathbb{Z}_{p^k} \oplus \mathbb{Z}_{p^k} \oplus ... \oplus \mathbb{Z}_{p^k}$ as an Abelian group, we can see why the following theorem may become useful.

Proposition 11. *Let A be a finite Abelian group that is isomorphic to an external direct sum $\mathbb{Z}_{p^k} \oplus \mathbb{Z}_{p^k} \oplus ... \oplus \mathbb{Z}_{p^k}$ with j copies of $\mathbb{Z}_{p^k}$. Then A/pA is isomorphic to $B = \mathbb{Z}_p \oplus \mathbb{Z}_p \oplus ... \oplus \mathbb{Z}_p$. Furthermore, this is a vector space over $\mathbb{Z}_p$.*

Proof. A simple application of the Fundamental Theorem of Finite Abelian Groups. □

There is a tight connection with structure of A and that of its corresponding vector space A/pA. First lets recall the definition of an *internal direct product* [3].

Definition 20. Let H_1, H_2,...,H_n be a finite collection of normal subgroups of G. Then G is the *internal direct product* of the H_i (denoted $G = H_1 \times H_2 \times ... \times H_n$) if

1. $G = H_1H_2...H_n = \{h_1h_2...h_n | h_i \in H_i\}$,
2. $(H_1H_2...H_i) \cap H_{i+1} = \{e\}$ for $i = 1, 2, ..., n-1$.

Note that in the case for which the operation is addition, we have $G = H_1 + H_2 + ... + H_n$ as well as $(H_1 + H_2 + ... + H_i) \cap H_{i+1} = \{0\}$ for $i = 1, 2, ..., n-1$.

Theorem 19. *Let A be as defined in Proposition 11. If $a_1, ..., a_j \in A$ are such that $a_1 + pA, a_2 + pA, ..., a_j + pA$ form a basis for A/pA then there is an internal direct product $A = \langle a_1 \rangle \times ... \times \langle a_j \rangle$.*

Proof. We are trying to show that A is an internal direct product of $\langle a_1 \rangle$,$\langle a_2 \rangle$,...,and $\langle a_j \rangle$. The definition of the internal direct product requires that these subgroups be normal. Since A is Abelian, all $\langle a_i \rangle$ are normal subgroups, so this part is trivial.

We first show condition 1 that $A = \langle a_1 \rangle + ... + \langle a_j \rangle$. Let $B = \langle a_1 \rangle \oplus ... \oplus \langle a_j \rangle$ and consider the homomorphism ϕ: $B \to A$ given by $(b_1, b_2, ..., b_j) \mapsto b_1 + b_2 + ... + b_j$. If we can show that this homomorphism is onto, then we know we can write every element

as a sum of elements from all $\langle a_i \rangle$. We will first show that $A = pA + \phi(B)$. Suppose $a \in A$. Then $a + pA$ in the vector space A/pA can be written as a linear combination of the basis elements, say $n_1 a_1 + ... + n_j a_j + pA$. That is, $a + pa' = n_1 a_1 + ... + n_j a_j + pa''$ for some $a', a'' \in A$ and we have $a = n_1 a_1 + ... + n_j a_j + p(a'' - a')$. Note in addition that $n_1 a_1 + ... + n_j a_j = \phi(n_1 a_1, ..., n_j a_j) \in \phi(B)$. Hence we can write any $a \in A$ as a sum of an element from pA and from $\phi(B)$ telling us that $A \subseteq pA + \phi(A)$. Showing containment the other direction is trivial because $\phi(B) \subseteq A$ and $pA \subseteq A$. Since A is closed under addition, this implies $pA + \phi(B) \subseteq A$. Hence $A = pA + \phi(B)$.

We now show that $A = p^2 A + \phi(B)$. First note that $A = p(pA + \phi(B)) + \phi(B)$ is trivial as we replaced A in the expression we found in the last paragraph. That $p(pA + \phi(B)) = p^2 A + p\phi(B)$ follows from distribution over addition of the elements. So we now have that $A = p^2 A + p\phi(B) + \phi(B)$. To see that $p\phi(B) + \phi(B) = \phi(B)$, note that $p\phi(B) \subseteq \phi(B)$. And since $\phi(B)$ is closed under addition, we have $p\phi(B) + \phi(B) = \phi(B)$. This shows that $A = p^2 A + \phi(B)$.

We could easily generalize the last paragraph to higher powers of p, but we will spare the reader the details as it uses the same logic. Now repeating the process, we eventually we find,

$$A = pA + \phi(B) = p^2 A + \phi(B) + ... = p^k A + \phi(B) = 0 + \phi(B) = \phi(B),$$

where we used the fact that the characteristic of A is p^k. Thus, $A = \phi(B)$ proving that the homomorphism is onto. This completes the proof that $A = \langle a_1 \rangle \times ... \times \langle a_j \rangle$.

Note that since A and B have the same number of elements, namely p^{kj}, this also proves that they are isomorphic. That A has p^{kj} elements is clear from its definition. To show that B has this many elements, we must show that the order of each a_i in A is p^k. This is simple because $a_i + pA$ is nonzero in A/pA, so clearly $a_i = (a_{i,1}, a_{i,2}, ..., a_{i,j})$ where at least one of the $a_{i,s}$ is not divisible by p, so a_i must have order p^k.

It is now simple to prove condition 2. The subgroups $\langle a_1 \rangle \oplus \langle a_2 \rangle \oplus ... \oplus \langle a_i \rangle \oplus 0 \oplus ... \oplus 0$ and $0 \oplus 0 \oplus ... \oplus 0 \oplus \langle a_{i+1} \rangle \oplus ... \oplus 0$ of B have trivial intersection for all i. It follows that their images under ϕ also have trivial intersection. But these images are exactly $\langle a_1 \rangle + \langle a_2 \rangle + ... + \langle a_i \rangle$

and $\langle a_{i+1} \rangle$, so condition 2 is proved. $\square$

There is one more lemma that is rather fundamental to the study of polynomials. It is called Hensel's Lemma, and we include it here without proof for use in the Classification of Galois Rings. The proof given by Bini and Flamini in [4] is long and tedious. Note that if we have a polynomial in $R[x]$ and view this polynomial "modulo p^s", this will mean we are viewing the coefficients as being in $R/(p^s)$.

Lemma 10. *Let R be a Galois ring of characteristic p^k with maximal ideal (p). Suppose $s \leq k$. Suppose $u(x), f(x), g(x) \in R[x]$ are monic polynomials and that $f(x)$ and $g(x)$ are relatively prime modulo p. Suppose further that $u(x) \equiv f(x)g(x)$ modulo p^s. Then it is possible to uniquely determine two polynomials $f_1(x), g_1(x) \in R/(p^{s+1})[x]$. such that the following hold:*

1. $f(x) \equiv f_1(x) \pmod{p^s}$,
2. $g(x) \equiv g_1(x) \pmod{p^s}$,
3. $u(x) \equiv f_1(x)g_1(x) \pmod{p^{s+1}}$.

Essentially this theorem allows us to show the existence of factorizations of polynomials in $R[x]$. If we have a factorization of $u(x)$ over $R/(p^s)$, we are essentially able to lift that factorization to $R/(p^{s+1})$ and higher until we reach $R/(p^k)$ where p^k is the characteristic of R. Since $p^k = 0$, this is simply a quotient by the trivial ideal giving us a factorization over R.

4.5 Classification of Galois Rings

We are now ready to prove the Classification of Galois Rings.

Theorem 20. *Let R_1 and R_2 be Galois rings with the same characteristic p^k and the same order. Then R_1 and R_2 are isomorphic as rings.*

Proof. We will organize this long proof into sections.

1. Preliminary Definitions and General Outline

   We know R_1 and R_2 are isomorphic as Abelian groups because of Lemma 9 and

therefore they both contain $\langle 1 \rangle = \mathbb{Z}_{p^k}$. It is also clear that their residue fields $F_1 = R_1/(p)$ and $F_2 = R_2/(p)$ must also have the same number of elements and are therefore isomorphic because finite fields are unique up to order. Let $\phi\colon F_1 \to F_2$ be such an isomorphism.

The outline of the proof is to show that R_1 and R_2 are isomorphic by proving that $R_1 \cong \mathbb{Z}_{p^k}[x]/(u(x))$ for some monic, irreducible $u(x) \in \mathbb{Z}_{p^k}[x]$ and that $R_2 \cong \mathbb{Z}_{p^k}[x]/(u(x))$ as well.

2. Defining $u(x)$

Since $F_1 = R_1/(p)$ has characteristic p, we can construct F_1 as being isomorphic to an extension $\mathbb{Z}_p(\bar{\alpha})$ for some $\bar{\alpha} \in F_1$ [3]. Let $\alpha \in R_1$ be such that $\bar{\alpha} = \alpha + (p) \in F_1$. It can be found in [3] that $1 + pR_1, \alpha + pR_1, \alpha^2 + pR_1, ..., \alpha^{j-1} + pR_1$ form a basis for F_1 over $\mathbb{Z}_p$. Since this is the case, we know from Theorem 19 that $R_1 = \langle 1 \rangle \times \langle \alpha \rangle \times ... \times \langle \alpha^{j-1} \rangle$. We can subsequently write

$$\alpha^j = c_0 + c_1\alpha + ... + c_{j-1}\alpha^{j-1}, \qquad \text{where} \quad c_0, c_1, ..., c_{j-1} \in \mathbb{Z}_{p^k}.$$

Let $u(x) = x^j - c_{j-1}x^{j-1} - ... - c_1x - c_0 \in \mathbb{Z}_p[x]$ and note that $u(\alpha) = 0 \in R_1$.

In fact, $u(x)$ is the minimal polynomial for α over $\mathbb{Z}_p$. It is monic by definition, so we now need to show it has the minimum degree for a polynomial with α as a root. Suppose that there were some $v(x) = v_0 + v_1x + ... + x^t$ of degree less than $\deg(u)$ (i.e. $t < j$) for which $v(\alpha) = 0$. Then in the vector space $R_1/(p)$ we find $v(\alpha) + (p) = v_0 + v_1\alpha + ... + \alpha^t + pR_1 = 0 + (p)$. But since the powers of α form a basis and $t < j$, this implies that the coefficients are divisible by p contradicting the fact that $v(x)$ is monic. We find that $u(x)$ is the minimal polynomial for α by contradiction.

3. Showing $R_1 \cong \mathbb{Z}_{p^k}[x]/(u(x))$

We finish showing the isomorphism involving R_1 by using the First Isomorphism Theorem for Rings. Consider the homomorphism $\psi\colon \mathbb{Z}_{p^k}[x] \to R_1$ given by $f(x) \mapsto f(\alpha)$. This is onto because the R_1 is the internal direct product $\langle 1 \rangle \times \langle \alpha \rangle \times ... \times \langle \alpha^{j-1} \rangle$.

In addition, the kernel of this map is simply $(u(x))$. This follows directly from the fact that $u(x)$ is the minimal polynomial of α. It can be found in Gallian that if $f(\alpha) = 0$, then $u(x)|f(x)$ so that $f(x) \in (u(x))$ [3]. We are now able to apply the First Isomorphism Theorem for Rings and we find that $\mathbb{Z}_{p^k}[x]/(u(x)) \cong R_1$.

4. Passing to F_2 and R_2 Now we move on to showing that $R_2 \cong Z_{p^k}[x]/(u(x))$ for the same $u(x)$. First of all, recall that $u(\alpha) = 0 \in R_1[x]$. Let's we take the natural homomorphism $\psi_1 \colon R_1 \to F_1$ where $a \mapsto a + (p)$. Consider also, by abuse of notation, the homomorphism $\psi_1 \colon R_1[x] \to F_1[x]$ where the coefficients are sent to their image under the former definition of ψ_1. The argument of ψ_1 will determine which definition we use. Clearly in $F_1[x]$, $0 = \psi_1(0) = \psi_1(u(\alpha)) = \bar{u}(\psi_1(\alpha))$ where $\bar{u}(x)$ is the image of $u(x)$ under ψ_1. Now recall that $\langle 1 \rangle = \mathbb{Z}_{p^k} \subseteq R_2$ and that $u(x)$ is defined over $\mathbb{Z}_p$. We can therefore view $u(x)$ as also being a polynomial in $R_2[x]$. Let's consider the same natural homomorphism $\psi_2 : R_2 \to F_2$ that sends $a \mapsto a + (p)$ and its polynomial counterpart. Let $\bar{u}(x)$ be the image of $u(x)$ under ψ_2. Note that since $u(x)$ is defined over $\mathbb{Z}_p$, $\bar{u}(x)$ will be defined over $\mathbb{Z}_{p^k}/(p) \cong \mathbb{Z}_p$. Recall the isomorphism $\phi : F_1 \to F_2$ defined in part 1. Letting $\beta = \phi(\alpha)$, we have $\bar{u}(\beta) = 0$. Since $u(x)$ is monic, so is $\bar{u}(x)$, and so $\bar{u}(x)$ is the minimal polynomial of β over $\mathbb{Z}_p$.

5. Factorizing $\bar{u}(x)$

   Since β is a root of $\bar{u}(x)$, we have the following factorization over F_2:

$$\bar{u}(x) = (x - \beta)\bar{g}(x).$$

   Of course, since $\bar{u}(x)$ is the minimal polynomial over Z_p, then by Proposition 1 it has not repeated roots in F_2 and so $\bar{g}(x)$ does not have β as a root. Equivalently, $x - \beta$ does not divide $\bar{g}(x)$. Since $x - \beta$ and $\bar{g}(x)$ share no factors, they are relatively prime. We can now use Hensel's Lemma (Lemma 10) to lift this factorization from $F_2[x]$ to $R_2[x]$. That is,

$$u(x) = f(x)g(x) \in R_2[x],$$

where as in Hensel's Lemma, $f(x) \equiv \bar{f}(x) = x - \beta \in F_2[x]$ and $g(x) \equiv \bar{g}(x) \in F_2[x]$.

6. Showing $u(x)$ has a root and finishing

   Since $\bar{f}(x)$ has degree 1, this means that all terms of $f(x)$ of degree greater than 1 have coefficients in (p). Theorem 18 then implies that $f(x)$ is the associate of a monic polynomial of degree 1. That is $f(x)h(x) = x - \alpha'$ where $\alpha' \in R_2$ and $h(x)$ is a unit. It follows that,

$$u(x) = f(x)g(x) = (x - \alpha')(h(x))^{-1}g(x) \in R_2[x].$$

   So in $R_2[x]$ we have $u(\alpha') = 0$. We already know that $u(x)$ is irreducible over $\mathbb{Z}_{p^k}$, therefore we may use the First Isomorphism Theorem for Rings as we did with R_1 to show $R_2 \cong \mathbb{Z}_{p^k}[x]/(u(x))$ finally proving that $R_2 \cong R_1$.

□

Now the possible characteristics for Galois rings are p^k for any prime p and positive integer k. In addition, the possible orders are p^{kj} for some positive integer j. We see why this is from constructing R_1 and R_2 as isomorphic to quotients of polynomial rings. We know that up to isomorphism there can only be one Galois ring for each possible characteristic and order. Do Galois rings exist for all such characteristics and orders? The answer is yes. We can construct them from the finite fields. Let's take recall that finite fields of all possible characteristics exist (meaning all prime characteristics) and are isomorphic if they have the same order.-

Theorem 21. *Suppose $k \in \mathbb{N}$ and F is a finite field of characteristic p. Then there is a Galois ring R of characteristic p^k such that $R/(p) \cong F$.*

Proof. To construct R, let $f(x) \in \mathbb{Z}[x]$ be a monic polynomial such that $f(x)$ modulo p gives $F \cong Z_p[x]/(f(x))$. Now view $f(x)$ modulo p^k so that $f(x) \in \mathbb{Z}_{p^k}[x]$. That is, interpret $f(x)$ as being in $\mathbb{Z}_{p^k}[x]$ (similar to the way we did with the field F). The claim is that $R = \mathbb{Z}_{p^k}[x]/(f(x))$ is a Galois ring of characteristic p^k. That R is finite and commutative is obvious. It is clearly unitary having unity 1. Now we can determine the characteristic

because the characteristic of a unitary ring is simply the additive order of its unity. Since $1 \in \mathbb{Z}_{p^k}[x]$, we have that the additive order of 1 is simply p^k proving that the characteristic of R is p^k. To finish showing that R is Galois, we need only show that (p) is its unique maximal ideal.

To see that (p) is maximal, note that $(\mathbb{Z}_{p^k}[x]/(f(x)))/(p) \cong \mathbb{Z}_p[x]/(f(x))$ where $f(x)$ is viewed modulo p on the right hand side. We know this is the field F. Since an ideal is maximal if and only if its quotient with the ring is a field, we know that p is maximal.

Now to show unicity. Let M be any maximal proper ideal of R. We will show $(p) \subseteq M$. Suppose $s = pt \in (p)$. Then $s^k = p^k t^k = 0$. This means that in R/M we have $s^k + M = 0 + M$. But R/M is a field and fields have no zero-divisors, so $s + M = 0 + M$ implying that $s \in M$. This means $(p) \subseteq M$. Since (p) is a maximal ideal and M is proper, this means $(p) = M$. □

It is clear that for each characteristic p^k, there are infinitely many Galois rings with this characteristic. We simply construct them by taking quotients of $\mathbb{Z}_{p^k}[x]$ with larger degree polynomials. We could also ask the question of how many Galois rings there are for each order. Note that given an order for a Galois ring p^n, it must have characteristic p^k for some $k < n$. In particular, notice that since the order is p^{kj} for some j, it must be the case that $k|n$. Therefore, the number of Galois rings of order p^n is equal to the number of divisors of n.

5 Codes Over Galois Rings

Recall from Section 3 that the definition of a code was general enough to be defined over any finite set. We added structure to codes by considering them over a Galois field allowing us to use properties of vector spaces. We will now add structure to codes in a slightly more general setting by looking at codes over Galois rings. In most cases we will look more specifically the ring $\mathbb{Z}_{p^k}$. Before doing so, we must cover a concept that may be unfamiliar to the reader.

5.1 Modules

The concept of a module is not typically covered in an undergraduate course on abstract algebra. We will include a basic overview here to give some context. This concept becomes important when talking about codes over rings because it is the natural generalization of the idea of codes over fields. Here is the definition.

Definition 21. [5] A *module* M over a ring R is a set of objects which is an abelian group under addition and is closed under multiplication by scalars from R. The operation of scalar multiplication is distributive and associative. A *submodule* of M as a subset of M that is itself a module.

Basically, a module is a vector space over a ring rather than a field. In general, a module over a ring R is called an *R-module*. Technically since R is not necessarily commutative, we should have defined this in terms of left- and right- scalar multiplication, but in all practicality we will only focus on modules over Galois rings which, as we know, have commutative multiplication.

Example 16. An example of a module M is the set of m-tuples of a ring R. That is, $M = \{(r_1, r_2, ..., r_m) | r_i \in R\}$ where the addition is done component-wise and scalar multiplication distributes to each component as in the usual vector space scalar multiplication.

Example 17. Another interesting example of a module is any abelian group A over the integers $\mathbb{Z}$. Using the usual notation for the operation in an abelian group (i.e. a+b) we let "scalar multiplication" by $n \in \mathbb{Z}$ be the sum of an element n times. That is, $na = a+a+...+a$ where a appears n times.

5.2 Linear Codes Over Galois Rings

Denote by $R^{(m)}$ the set of m-tuples of a Galois ring R. By example 16 we know this is an R-module. A *code of length m over R* is a subset of $R^{(m)}$[4]. This is essentially an application of Definition 2. Now what is the natural generalization of linear codes? When speaking of linear codes over fields, we used the term "subspace". We now must use "submodules".

Definition 22. A code C over R is *linear* if it is a submodule of $R^{(m)}$.

In [4], Bini and Flamini only consider codes over the Galois rings $\mathbb{Z}_{p^n}$. They give the following definition.

Definition 23. A code C over $\mathbb{Z}_{p^n}$ is *linear* if it is a subgroup of $\mathbb{Z}_{p^n}^{(m)}$.

The difference is the use of "submodule" versus "subgroup". The more general definition is of course the former, but in the case of the module $\mathbb{Z}_{p^n}^{(m)}$, a subgroup and submodule are equivalent terms.

Example 18. The set $M = \mathbb{Z}_4^{(2)}$ is the set of all 2-tuples over $\mathbb{Z}_4$. That is, an element of M looks like (n_1, n_2) where $n_i \in \{0, 1, 2, 3\}$ and addition and scalar multiplication is done modulo 4. So $(1, 3) + (1, 2) = (2, 1)$ and $3(1, 3) = (3, 1)$. Note that M is a $\mathbb{Z}_4$-module. Now the subset $C = \{(0, 0), (1, 3), (2, 2), (3, 1)\}$ is a linear code over $\mathbb{Z}_4$ because its elements form a subgroup of $\mathbb{Z}_4^{(2)}$ under addition.

We can also extend the notion of distance. Codes over the Galois ring $\mathbb{Z}_{p^n}$ require a slightly different notion than the Hamming distance. With that in mind, we move the following definitions given by Bini and Flamini [4].

Definition 24. The *Lee weight* of an element $h \in \mathbb{Z}_{p^n}$ is

$$wt_L(h) := \min\{h, p^n - h\}$$

Definition 25. The *Lee weight* of an element $v = (v_1, v_2, ..., v_m) \in \mathbb{Z}_{p^n}^{(m)}$ is the sum of the Lee weights of its coordinates

$$wt_L(v) = \sum_{i=1}^{m} wt_L(v_i)$$

Example 19. Consider the element $(1, 6) \in \mathbb{Z}_8^{(2)}$. The Lee weight of 1 is $wt_L(1) = \min\{1, 7\} = 1$ and the Lee weight of 6 is $wt_L(6) = \min\{6, 2\} = 2$. This gives a Lee weight for the 2-tuple of $wt_L((1, 6)) = 1 + 2 = 3$.

Linear codes over finite fields have generator matrices. This notion extends to codes over $\mathbb{Z}_{p^n}$ as well [4]. Given a linear code C of length m over $\mathbb{Z}_{p^n}$, it is equivalent to a code

whose generator matrix is

$$G = \begin{pmatrix} I & A_{0,1} & A_{0,2} & A_{0,3} & ... & A_{0,n-1} & A_{0,n} \\ 0 & pI & pA_{1,2} & ... & ... & ... & pA_{1,n} \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & ... & p^{n-1}I & p^{n-1}A_{n-1,n} \end{pmatrix}. \tag{2}$$

Here, each I represents the identity matrix of some rank, though they may not be all of the same rank. Suppose the identity matrix in the first column has rank $k_0 \in \mathbb{N}$. Equivalently we could say that the first column actually represents k_0 columns. Next, we say the identity matrix in the second column has rank k_1 and so on up to k_n. Since we are looking at a code of length m, the constraint is that $\sum_{i=0}^{n} k_i = m$. This simply states that G must have m columns. The $A_{i,j}$ are all matrices of appropriate size. That is, the matrices in row i have k_i rows and the matrices in column j have k_j columns.

We will now motivate how this operates as a generator matrix. We could try to view this as we did before using linear combinations of row vectors, but there is a simpler way. Returning to the original definition of a generator matrix over Galois fields, suppose we have a code C of length n and dimension k over a field $GF(q)$. We could think of a vector in this code as some linear combination of the row vectors in a generator matrix G. One way to write this is that a vector in the code is given by $(a_1 a_2 ... a_k)G$, where each $a_i \in GF(q)$. For example, lets say we have a binary code of length 4 whose generator matrix is

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

Then if we wish to write the linear combination that sums the first two rows we would write

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

In the case of a code over $\mathbb{Z}_{p^n}$, with regards to our generator matrix G, an element of the

code is given by

$$\begin{pmatrix} v_0 & \dots & v_{n-1} \end{pmatrix} G. \tag{3}$$

Carrying out the multiplication, we would see that v_i left multiplies the matrices in row i of G. Thus v_i is a vector of length k_i and components in $\mathbb{Z}_{p^{n-i}}$. To confirm our understanding of this construction, here is a concrete example.

Example 20. Suppose a code C of length 5 over the ring $\mathbb{Z}_4$ has a generator matrix of the form

$$G' = \begin{pmatrix} I_0 & A_{0,1} & A_{0,2} \\ 0 & pI_1 & pA_{1,2} \end{pmatrix}.$$

Suppose $k_0 = 2$, $k_1 = 1$, and $k_2 = 2$. Note that $k_0 + k_1 + k_2 = 5$. These values mean that I_0 is the 2×2 identity matrix and I_1 is the 1×1 identity matrix. The matrix $A_{0,1}$ will have size 2×1, $A_{0,2}$ is 2×2, and $A_{1,2}$ is 1×2. Let's multiply by the vector $(v_0 v_1)$. We find

$$\begin{aligned} \begin{pmatrix} v_0 & v_1 \end{pmatrix} G' &= \begin{pmatrix} v_0 I_0 + v_1 0 & v_0 A_{0,1} + v_1 p I_1 & v_0 A_{0,2} + v_1 p A_{1,2} \end{pmatrix} \\ &= \begin{pmatrix} v_0 & v_0 A_{0,1} + p v_1 & v_0 A_{0,2} + v_1 p A_{1,2} \end{pmatrix}. \end{aligned}$$

We might ask whether the sums in the second and third entries are well-defined. That is, are these vectors the same length? Let's consider the third entry. First of all, the vectors v_0 and v_1 have the proper lengths for multiplication by these matrices. That is, they have lengths k_0 and k_1, respectively, while the matrices $A_{0,2}$ and $A_{1,2}$ have these as heights as well. Now for the result of their multiplication. The matrices $A_{0,2}$ and $A_{1,2}$ both have k_2 columns, so multiplying from the left by a vector gives a vector of length k_2 as well. Hence, the third entry is a sum of vectors of length k_2.

It is important to be able to calculate the order of a code C whose generator matrix is given by G from Equation 2. We know that any element of the code may be written in the form of Equation 3. Counting the number of codewords is therefore equivalent to counting the number of vectors of the form $(v_0 \dots v_{n-1})$. Since v_0 is of length k_0 and has components

in $\mathbb{Z}_{p^n}$, there are $(p^n)^{k_0} = p^{nk_0}$ possible values for v_0. Similarly there are $p^{(n-1)k_1}$ possible values for v_1. Overall $(v_0...v_{n-1})$ can take on $(p^{nk_0})(p^{(n-1)k_1})...(p^{k_{n-1}}) = p^\ell$ values where $\ell = \sum_{i=0}^{n-1}(n-i)k_i$. This is the order of the code C.

There is an analagous definition of the dual code.

Definition 26. The *dual code* of C is $C^\perp = \{\mathbf{x} \in \mathbb{Z}_{p^n}^{(m)} | \mathbf{x} \cdot \mathbf{y} = 0 \quad \forall \in C\}$ where $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{m} x_i y_i$.

Again, the *parity check matrix* is simply the generator matrix of the dual code. It turns out that if the generator matrix is given by G as above, then the parity check matrix H is

$$H = \begin{pmatrix} B_{0,n} & B_{0,n-1} & ... & B_{0,2} & B_{0,1} & I \\ pB_{1,n} & pB_{1,n-1} & ... & pB_{1,2} & pI & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ p^{n-1}B_{n-1,n} & p^{n-1}I & ... & 0 & 0 & 0 \end{pmatrix},$$

for some matrices $B_{i,j}$. This will satisfy the product $GH^T = 0$ as we saw from Theorem 5. The sizes of the matrices must match in the product. It does not take much work to see that this means the identity matrix in the last column must have size k_n and so on. So in general, the matrices in the ith row have k_{n-i} rows and the matrices in the jth column have k_{n-j} columns. Note here we are indexing the rows and columns by 0.

6 Conclusion

The mathematical definition of a code is rather general, but the alphabet over which the code is defined can imbue it with useful structure. Historically, the most fruitful alphabets in terms of study and application have been the Galois fields. Among these, the field of order 2 stands out due to its use in computers. The Galois fields allow for linear codes which brings about topics such as syndrome decoding, perfect codes, and cyclic codes. Galois rings generalize Galois fields and it is therefore natural to generalize the concepts associated with codes over Galois fields to include codes over Galois rings. We have seen that the concepts of linear codes, generator matrices, weight, distance, and dual codes all generalize. Further

work would have gone into generalizing the Hamming codes, cyclic codes, and other concepts we did not cover. A discussion of such concepts can be found in [4].

Acknowledgements

Thank you to Professor Patrick Keef for suggesting the topic of this paper and for providing guidance on Galois rings. Thank you to Marissa Childs for editing this paper, especially in keeping my explanations readable. Thank you also to Professor Barry Balof for instructing the Mathematics Senior Project course. Finally, thank you to my parents James and Donna Holdman for supporting me during my time a Whitman College and providing me the opportunity to explore my interests.

References

[1] Thompson, Thomas M. *From Error-Correcting Codes Through Sphere Packings to Simple Groups* The Mathematical Association of America, 1983.

[2] Pless, Vera. *Introduction to the Theory of Error-Correcting Codes.* Wiley: New York, 1982.

[3] Gallian, Joseph A. *Contemporary Abstract Algebra, Seventh Edition.* Brooks/Cole: Belmont, CA, 2010.

[4] Bini, Gilberto and Flaminio Flamini. *Finite Commutative Rings and Their Applications.* Kluwer Academic Publisher, Springer. 2015.

[5] Ash, Robert B. (2000). *Abstract Algebra: The Basic Graduate Year.* Retrieved from http://www.math.uiuc.edu/ r-ash/Algebra.html.

Index