Algebraic-Geometry Codes

Ian Blake, *Fellow, IEEE*, Chris Heegard, *Fellow, IEEE*, Tom Høholdt, *Senior Member, IEEE*, and Victor Wei, *Fellow, IEEE*

(Invited Paper)

***Abstract*— The theory of error-correcting codes derived from curves in an algebraic geometry was initiated by the work of Goppa as generalizations of Bose-Chaudhuri-Hocquenghem (BCH), Reed–Solomon (RS), and Goppa codes. The development of the theory has received intense consideration since that time and the purpose of the paper is to review this work. Elements of the theory of algebraic curves, at a level sufficient to understand the code constructions and decoding algorithms, are introduced. Code constructions from particular classes of curves, including the Klein quartic, elliptic, and hyperelliptic curves, and Hermitian curves, are presented. Decoding algorithms for these classes of codes, and others, are considered. The construction of classes of asymptotically good codes using modular curves is also discussed.**

***Index Terms*— Algebraic curves, algebraic-geometry codes, asymptotically good codes, decoding algorithms.**

I. Introduction

THE origins of the subject of error-correcting codes are found in the classical papers of Shannon [79]. The subject developed rapidly, both in engineering practice and as a mathematical discipline. The notions of Bose–Chaudhuri–Hocquenghem (BCH), Reed–Solomon (RS), and Goppa codes, in particular, achieved prominence with extensive research contributions over a period of almost four decades. Along with a developing mathematical theory of codes, went intense research on the most efficient algorithms to decode them, an effort that continues.

From a theoretical point of view, a significant research objective was to construct asymptotically good codes, codes whose parameters achieved the Varshamov–Gilbert lower bound, introduced in the next section. Although there was much interesting work on the problem [48], the goal remained elusive.

While the construction of asymptotically good codes proved difficult, the construction of many other interesting classes of codes proceeded swiftly. Prominent among these are the classes of BCH, RS, and Goppa codes, already mentioned, whose mathematical properties and decoding algorithms were widely studied. These classes of codes have codewords that can be viewed as either the evaluation of certain functions on a set of distinct elements in a finite field, or the evaluation of residues there, and these notions have proved to be important. While it was known [64] that there exists a sequence of Goppa codes that met the Varshamov–Gilbert bound, their actual construction proved more difficult. Goppa [33], [34] made the crucial observation in generalizing these notions by, in one instance, evaluating a set of rational functions at the points on an algebraic curve. In making this step, many of the tools needed to determine the important parameters of the code, or bounds on them, such as the code length, dimension, and minimum distance, already existed in the elegant theorems of algebraic geometry, notably the Hasse–Weil theorem and the Riemann–Roch theorem. Having evaluated the construction of codes in this manner, it quickly led Tsfasman, Vlăduţ, and Zink [91], [92], using modular curves, to show how asymptotically good codes could be constructed over alphabets of size $q \geq 49$, a truly remarkable achievement (see Fig. 1).

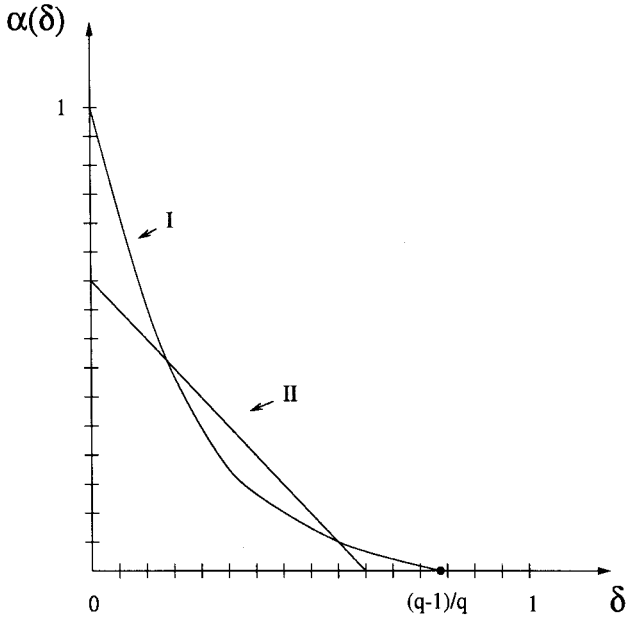

Fig. 1. I: Gilbert–Varshamov bound, II: Tsfasman–Vlăduţ–Zink bound, $q \geq 49$.

The theory of algebraic-geometry codes involves the relatively deep and fundamental results of algebraic geometry.

Manuscript received December 5, 1997; revised May 11, 1998. The work of C. Heegard was supported in part by the National Science Foundation under Grant NCR-9520981.

I. Blake is with Hewlett-Packard Laboratories, Palo Alto, CA 94304 USA (e-mail: ifblake@hpl.hp.com).

C. Heegard is with the School of Electrical Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: heegard@ee.cornell.edu) and with Alantro Communications, Santa Rosa, CA USA.

T. Høholdt is with the Mathematical Institute, Technical University of Denmark, DK 2800, Lyngby, Denmark (e-mail: tom@mat.dtu.dk).

V. Wei is with the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, New Territory, Hong Kong (e-mail: kwwei@ie.cuhk.edu.hk).

Publisher Item Identifier S 0018-9448(98)05712-5.

While there are now several books that attempt to give self-contained treatments of algebraic geometry and codes ([65], [83], [90]) it nonetheless requires effort on the part of the nonexpert to appreciate the significant developments of the area. The aim of this paper is not so much to give a survey of the rather large body of work that now exists in this area, but to trace the evolution of the subject over the past few decades from the earliest code constructions to the elegant and deep theory that exists today. In particular, an attempt is made to give some notion as to the role the properties of algebraic curves has played in the subject. While the review has been written for the nonexpert, some familiarity with the subject of error-correcting codes and algebra has been assumed. The aim has been to outline the construction of important classes of codes instrumental in the development. It is also intended to give a brief overview of those concepts from algebraic geometry needed to appreciate the development, in a relatively self-contained manner to allow such a nonexpert a glimpse into this development of the subject.

The next section reviews the constructions of certain basic classes of codes, RS, BCH, and Goppa, in such a manner that makes natural the critical step that was taken in extending these to constructions of codes from algebraic curves. The mathematical background needed to understand the application of algebraic geometry to coding is outlined in Section III. While no proofs are given, the theory is illustrated with examples and an informed reader should be able to appreciate the ideas involved. Section IV uses the ideas developed to outline the construction of codes that are derived from many of the more commonly used curves, including the Klein quartic, elliptic and hyperelliptic curves, and Hermitian curves. In addition, interesting constructions due to Feng and Rao ([22], [24]) are considered.

The study of decoding algorithms for codes from curves in an algebraic geometry has been intense over the last decade, meeting the challenge of extending the one-dimensional concepts of decoding BCH, RS, and Goppa codes, to two dimensions. This has involved consideration of the difficult problems encountered in extracting decoding information from the two-dimensional syndromes and the incorporation of the structure of the curves in the decoding process. Progress on this problem is covered in Section V.

Section VI outlines the use of modular curves in the construction of sequences of asymptotically good codes, a quest that started in the 1950's with the establishment of the Varshamov–Gilbert bound. The first step in this direction was taken with the interesting construction of Justesen [48]. The elegant and deep approach using the theory of modular curves holds promise for even greater insight into this challenging problem.

A few comments on the problems and challenges that might be of interest in the future are given in the final section of the paper. The introduction of algebraic geometry to the problem of constructing codes, and in particular, families of asymptotically good codes, has opened up fascinating possibilities of both a theoretical and practical nature for future research. It is hoped this paper might serve as a starting point from which these possibilities might be appreciated.

II. From Reed–Solomon Codes to Algebraic-Geometry Codes

A setting that has proved fruitful for coding theory is to view a code C as a subset of the vector space of n-tuples over the finite field of q elements, $\mathbb{F}_q$, which we denote as $\mathbb{F}_q^n$. The (Hamming) distance between any two vectors of the space, $\boldsymbol{a}$, $\boldsymbol{b} \in \mathbb{F}_q^n$ is then the minimum number of coordinate positions in which they differ, denoted by $d(\boldsymbol{a}, \boldsymbol{b})$. The Hamming weight of a vector $\boldsymbol{a} \in \mathbb{F}_q^n$, $w(\boldsymbol{a})$, is the number of its coordinate positions which are nonzero. The minimum distance of a code is then

$$d = \min_{\boldsymbol{a}, \boldsymbol{b} \in C, \boldsymbol{a} \neq \boldsymbol{b}} d(\boldsymbol{a}, \boldsymbol{b}).$$

If $|C| = M$ and C has minimum distance d, it is referred to as an $(n, M, d)_q$ code.

Defining the sphere of radius r with center $\boldsymbol{x}$ as

$$S(\boldsymbol{x}, r) = \{\boldsymbol{a} \in \mathbb{F}_q^n | d(\boldsymbol{a}, \boldsymbol{x}) \leq r\}$$

it is immediately seen that it is possible to surround the codewords of a code C with minimum distance d, with nonintersecting spheres of radius $t = \lfloor (d-1)/2 \rfloor$ where $\lfloor \cdot \rfloor$ is the floor function. Since each sphere contains

$$\sum_{i=0}^{t} \binom{n}{i} (q-1)^i$$

vectors it follows that

$$|C| \left\{ \sum_{i=0}^{t} \binom{n}{i} (q-1)^i \right\} \leq q^n$$

a result referred to as the Hamming bound for the code C. A code that achieves this bound with equality is called perfect and the existence of perfect codes is now a settled problem [61].

Designing codes that have a large minimum distance for a given code size, and alphabet size, without more structure is challenging. The addition of linearity to the code set, i.e., requiring that the codewords or vectors of C form a linear subspace of $\mathbb{F}_q^n$, allows considerably more to be said about the code properties. A linear $(n, k, d)_q$ code C is a k-dimensional subspace of $\mathbb{F}_q^n$ with the property that any two distinct codewords are at least distance d apart. Notice that the addition of two codewords is also a codeword, and so the minimum distance of the code is the weight of the smallest weight nonzero codeword, i.e.,

$$d = \min_{\boldsymbol{a} \neq \boldsymbol{b}} d(\boldsymbol{a}, \boldsymbol{b}) = \min_{\boldsymbol{a} \in C, \boldsymbol{a} \neq \boldsymbol{0}} w(\boldsymbol{a}).$$

The linear code C, as a k-dimensional subspace, can be generated by a set of k linearly independent codewords, $\boldsymbol{g}_1, \cdots, \boldsymbol{g}_k \in \mathbb{F}_q^n$. If the codeword $\boldsymbol{g}_i$ is viewed as the row of a $k \times n$ matrix G, the code C is the rowspace of G, and G is referred to as a generator matrix of C. A possible encoding procedure for C is then to encode the message vector $\boldsymbol{m} \in \mathbb{F}_q^k$ to $\boldsymbol{m}G$. Indeed,

$$C = \{\boldsymbol{c} | \boldsymbol{c} = \boldsymbol{m}G, \ \boldsymbol{m} \in \mathbb{F}_q^k\}.$$

Corresponding to the subspace C is the orthogonal subspace

$$C^{\perp} = \{\boldsymbol{x} \in \mathbb{F}_q^n | (\boldsymbol{x}, \boldsymbol{c}) = 0, \ \forall \boldsymbol{c} \in C\}$$

where $(\boldsymbol{x}, \boldsymbol{c})$ is the usual inner product on $\mathbb{F}_q^n$. Such a subspace will have a generator matrix H and, by definition

$$GH^t = 0$$

where 0 is the $k \times (n-k)$ matrix of zeros. Alternatively, we can express the code as

$$C = \{\boldsymbol{c} \in \mathbb{F}_q^n | \boldsymbol{c}H^t = \boldsymbol{0} \in \mathbb{F}_q^{n-k}\}.$$

Viewed in this manner, a codeword $\boldsymbol{c} \in C$ of weight w corresponds to a dependency relation among the w columns of the matrix H corresponding to nonzero coordinates of C. From this observation it follows immediately that the code C has minimum distance of at least d if and only if no subset of $d-1$ or fewer columns of H are linearly dependent over $\mathbb{F}_q$. Because the columns of H are $(n-k)$-tuples, and the maximum number of such independent columns is $n-k$, it follows that $d \leq n-k+1$. This is the Singleton bound for (n, k, d) linear codes. Codes which achieve equality are referred to as maximum-distance separable (MDS).

By similar reasoning, suppose it has been possible to construct a $(n-k) \times (n-1)$ matrix over $\mathbb{F}_q$ such that all sets of $(d-2)$ or fewer columns are linearly independent. In the "worst case" such sums give distinct $(n-k)$-tuples and hence if

$$q^{n-k} > \sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i$$

then it is possible to add a column to the matrix which is linearly independent to any set of $d-2$ other columns and hence achieve an $(n, k, d)_q$ code. This is referred to as the Varshamov–Gilbert bound. An asymptotic version of it will be used in a later section.

It will be useful to recall a few elementary properties of polynomials. By a fundamental theorem of algebra, a polynomial of degree n over a field $\mathbb{F}$ has at most n zeros in that field. The smallest extension of $\mathbb{F}$ containing all the zeros of the polynomial is called its splitting field. The polynomial $f(x) \in \mathbb{F}[x]$ has a zero of order m at β if $(x-\beta)^m$ divides $f(x)$ while $(x-\beta)^{m+1}$ does not. A zero of order one is called a simple zero.

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

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

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

Further, let $\{v_0, v_1, \cdots, v_{n-1}\}$ be a set of nonzero, not necessarily distinct elements from $\mathbb{F}_q$. The code

$$C' = \{(v_0 f(\alpha_0), v_1 f(\alpha_1), \cdots, v_{n-1} f(\alpha_{n-1})), \ f \in L\}$$

has the same parameters as the previous code and is referred to as a generalized RS (GRS) code with vector $\boldsymbol{v} = (v_0, v_1, \cdots, v_{n-1})$. This minor adjustment can be useful in some constructions.

The above code can be described in a slightly different manner which will provide a useful perspective for the subsequent transition to construction of codes from algebraic curves. Consider the set of pairs of elements (x_1, x_2), $x_i \in \mathbb{F}_q$. Pairs which are scalar multiples of each other are identified, i.e., the pairs $\beta(x_1, x_2) = (\beta x_1, \beta x_2)$ are identified for all $\beta \in \mathbb{F}_q^*$. Thus all pairs can be grouped into equivalence classes with representatives

$$(1, \alpha), \qquad \alpha \in \mathbb{F}_q \text{ and } (0, 1)$$

and such classes are identified as the projective line $\mathbb{P}^1$. The extension to higher dimensional projective spaces is immediate, constructing $\mathbb{P}^N$ from $N+1$-tuples over $\mathbb{F}_q$.

Consider the set of rational functions $a(x, y)/b(x, y)$ where $a(x, y)$ and $b(x, y)$ are homogeneous polynomials of the same degree. Define now L to be the vector space of all such rational functions over $\mathbb{F}_q$ with the additional property that they do not have poles on $\mathbb{P}^1$ except possibly at the point $(0, 1)$, a point we will subsequently refer to as the point at infinity. Furthermore, when the rational function does have a pole at the point at infinity, it is of order less than k. Clearly, a ratio of polynomials of the form $a(x, y)/x^l$, $l < k$ where $a(x, y)$ is homogeneous of degree l, has this property. The RS code can then be described as

$$C = \{(f(P_1), \cdots, f(P_n), \ f \in L\}$$

where the $P_1, \cdots, P_n$ are a subset of the projective points not at infinity. The process of evaluating rational functions at a sequence of points on a curve (so far only a line) will be of importance to our development.

The addition of the requirement that every cyclic shift of a codeword also be a codeword, has led to powerful techniques for the design of good linear codes. While cyclic codes will not be discussed in any detail here, the following construction of BCH codes will be of interest. Let α be a primitive nth root of unity in an extension field of $\mathbb{F}_q$, say $\mathbb{F}_{q^m}$, $n | q^m - 1$, and let $g(x) \in \mathbb{F}_q[x]$ be the polynomial of smallest degree with zeros $\{\alpha^i, i = 1, 2, \cdots, 2t\}$ for some integer $t \geq 1$. Let the degree of $g(x)$, referred to as the generator polynomial of the code, be $n-k$ and note that $n-k \leq 2tm$ since for general q, the maximum number of distinct cyclotomic cosets of these elements is $2t$, each containing at most m elements. Then

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

is a BCH code of length n, dimension $k \geq n-2tm$, and minimum distance $d \geq 2t+1$. The code can be viewed as the null space, over $\mathbb{F}_q$, of the rowspace of the parity-check

matrix $H = (\alpha^{ij})$, $i = 1, 2, \cdots, 2t$, $j = 1, 2, \cdots, n$. The bound on the minimum distance follows from the fact that any $2t$ or fewer columns are independent, from a van der Monde argument.

Notice that if we take the polynomial

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

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

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

i.e., the set of all codewords in C with all coordinates in the field $\mathbb{F}$. Such subfield subcodes have been of considerable interest in more general situations than the particular case of BCH codes described here, e.g., [83].

To prepare for a definition of Goppa codes, the definition of BCH codes is first recast. With the same notation as above, consider the computation

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

For $j = 0, 1, \cdots, d - 2$ the inner summation is zero, by definition. Thus

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

for some polynomial $f(x)$, i.e., the summation is divisible by x^{d-1}. Thus

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

Consequently, a word $(c_0, c_1, \cdots, c_{n-1})$, $c_i \in \mathbb{F}_q$, is a codeword iff it satisfies the above equation. The construction yields either an RS or BCH code depending on the field of definition. Notice that the polynomial x^{2t} has a zero of order $2t$ at $x = 0$.

The passage from the above definition to that of Goppa codes will involve nothing more than replacing the sequence of nth roots of unity with an arbitrary set of distinct elements and the polynomial x^{2t} with a more general polynomial $g(x)$. (Note that this is not the generator polynomial used in the BCH construction—it is conventional to use $g(x)$ in both cases.)

Definition 2.1: Let $L = \{\alpha_0, \alpha_1, \cdots, \alpha_{n-1}\}$ be a set of n distinct elements in $\mathbb{F}_{q^m}$ and $g(x) \in \mathbb{F}_{q^m}[x]$ be a monic polynomial such that $g(\alpha_i) \neq 0$, $i = 0, 1, \cdots, n-1$. Then the Goppa code $\Gamma(L, g)$ is the set of words $(c_0, c_1, \cdots, c_{n-1}) \in \mathbb{F}_q^n$ such that

$$\sum_{i=0}^{n-1}\frac{c_i}{x - \alpha_i} \equiv 0 \bmod g(x).$$

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

Comparing to the previous formulation, if $g(x) = x^{2t}$ and $L = \{\alpha^{-i}, 0 \leq i \leq n - 1\}$, α a primitive nth root of unity, then $\Gamma(L, g)$ is a BCH code with designed distance d, although it is noted that not all BCH codes are Goppa codes. By a simple manipulation of the definitions, it will be shown that $\Gamma(L, g)$ where g has degree t, has dimension at least $n - mt$ and minimum distance at least $t + 1$.

It is also noted that $\Gamma(L, g)$ is a subfield subcode of the dual of a generalized RS code. To see this, let $g(x) = \sum_{i=0}^{t} g_i x^i$. From the fact that

$$\frac{g(z) - g(x)}{z - x} = \sum_{k+j \leq t-1} g_{k+j+1}x^j z^k$$

it follows that, for any codeword $(c_0, \cdots, c_{n-1})$ we have

$$\sum_{i=0}^{n-1} c_i h_i \sum_{k+j \leq t-1} g_{k+j+1}(\alpha_i)^j z^k = 0$$

where $h_i = 1/g(\alpha_i)$. Since the coefficient of z^k must be zero for $0 \leq k \leq t - 1$, it follows that the inner product of the codeword with the rows of the following matrix must be zero:

$$\begin{bmatrix} h_0 g_t & \cdots & h_{n-1}g_t \\ h_0(g_{t-1} + g_t\alpha_0) & \cdots & h_{n-1}(g_{t-1} + g_t\alpha_{n-1}) \\ \vdots & \cdots & \vdots \\ h_0\sum_{i=1}^{t} g_i\alpha_0^{i-1} & \cdots & h_{n-1}\left(\sum_{i=1}^{t} g_i\alpha_{n-1}^{i-1}\right) \end{bmatrix}.$$

Using elementary row operations, this is easily reduced to a parity-check matrix for the code $\Gamma(L, g)$ of the form

$$\begin{bmatrix} h_0 & h_1 & \cdots & h_{n-1} \\ h_0\alpha_0 & h_1\alpha_1 & \cdots & h_{n-1}\alpha_{n-1} \\ \vdots & \vdots & \cdots & \vdots \\ h_0\alpha_0^{t-1} & h_1\alpha_1^{t-1} & \cdots & h_{n-1}\alpha_{n-1}^{t-1} \end{bmatrix}$$

from which the properties of the code noted above follow readily. Thus the Goppa code $\Gamma(L, g)$ is the dual of a GRS code with vector $\boldsymbol{v} = (h_0, h_1, \cdots, h_{n-1})$. As the rank of this matrix over $\mathbb{F}_{q^m}$ is exactly t, the rank over $\mathbb{F}_q$ is at most mt. Thus the dimension of $\Gamma(L, g)$ is at least $n - mt$ and the minimum distance is at least $t + 1$.

To put the transition to codes from algebraic curves in perspective, it will be of interest to recast the definition

of Goppa codes. Consider a polynomial corresponding to a codeword $(c_0, c_1, \cdots, c_{n-1})$

$$f(x) = \sum_{i=0}^{n-1} \frac{c_1}{(x-\alpha_i)} = \frac{\omega(x)}{\lambda(x)}$$
$$\lambda(x) = \prod_i (x-\alpha_i) \in \mathbb{F}_{q^m}[x]$$

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

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

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

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

and let

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

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

$$\mathrm{Res}_{\alpha_i}(f) = \left.\frac{\omega(x)(x-\alpha_i)}{\lambda(x)}\right|_{x=\alpha_i} = \frac{g(\alpha_i)}{\chi_i(\alpha_i)} q(\alpha_i)$$

which is zero only if $q(\alpha_i) = 0$ as $g(\alpha_i)$, $\chi_i(\alpha_i) \neq 0$ by definition. Thus much as was done for RS codes, define a vector space of rational functions $f(x)$, L, such that

i) $f(x)$ has zeros where $g(x)$ has zeros, with multiplicity at least those of $g(x)$;
ii) $f(x)$ has poles only contained in the set L and in that case only poles of order one.

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

$$C' = \{(\mathrm{Res}_{\alpha_0} f, \mathrm{Res}_{\alpha_1} f, \cdots, \mathrm{Res}_{\alpha_{n-1}} f),\ f \in L\}$$

where the residue of a rational function is defined in the usual manner. It is seen immediately that the Goppa code $\Gamma(L, g)$ is the subfield subcode of this set over $\mathbb{F}_q$.

The two important perspectives to be drawn from this section, perspectives that will survive the transition to codes from algebraic curves intact, are the notions of defining codewords in the first instance, as the evaluation of a rational function at a fixed set of distinct places, and in the second instance, as the set of residues of a rational function at a fixed set of places. In the setting of algebraic geometry, the fixed set of places will be drawn from the points on a curve in an algebraic geometry. The two code constructions, using evaluations and residues at this fixed set of places, will carry over. The determination of code parameters, however, will depend in crucial ways on the theory of algebraic curves. The next section will serve as an overview of this theory, in preparation for Section IV which considers classes of codes that use these notions for their construction.

III. Basic Theory of Algebraic Geometry

We introduce the basic notions of algebraic geometry, in order to extend the construction and properties of codes discussed in the previous section to algebraic-geometry codes, to be discussed in the next section. We will give no proofs but refer to the standard textbooks ([65], [83], [90]). The central concepts required are limited and the material is illustrated with examples. It attempts only to convey the central themes of what is required to appreciate their application to coding.

A. Affine and Projective Varieties

Definition 3.1: Let $\mathbb{F}_q$ be the finite field with q elements and $\overline{\mathbb{F}}_q$ its algebraic closure. The n-dimensional affine space $\mathbb{A}^n$ is the set $\mathbb{A}^n = \{(a_1, \cdots, a_n) | a_i \in \overline{\mathbb{F}}_q\}$.

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

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

$$(a_0, \cdots, a_n) \equiv (b_0, \cdots, b_n) \Leftrightarrow \exists \lambda \in \overline{\mathbb{F}}_q \setminus \{0\} \text{ s.t.}$$
$$b_i = \lambda a_i, \qquad i = 0, 1, \cdots, n.$$

The equivalence class of $(a_0, a_1, \cdots, a_n)$ is denoted $(a_0 : a_1 : \cdots : a_n)$.

Definition 3.2: The *n-dimensional projective space* $\mathbb{P}^n$ is the set of all equivalence classes $\{(a_0 : a_1 : \cdots : a_n) | a_i \in \overline{\mathbb{F}}_q$, not all $a_i = 0\}$. An element $P = (a_0 : a_1 : \cdots : a_n) \in \mathbb{P}^n$ is called a *point* and $(a_0 : a_1 : \cdots : a_n)$ are called homogeneous coordinates of P. If $\mathbb{G}$ is a subfield of $\overline{\mathbb{F}}_q$ which contains $\mathbb{F}_q$ and P is a point for which there exist homogeneous coordinates $a_0, \cdots, a_n \in \mathbb{G}$ is called a $\mathbb{G}$-*rational point* and the set of $\mathbb{G}$-rational points of $\mathbb{P}^n$ is denoted $\mathbb{P}^n(\mathbb{G})$.

The set $H = \{(0 : a_1 : \cdots : a_n) \in \mathbb{P}^n\}$ is called the *hyperplane at infinity* and the points $Q \in H$ are the *points at infinity*. The mapping φ: $\mathbb{A}^n \to \mathbb{P}^n \setminus H$ defined by $\varphi(a_1, a_2, \cdots, a_n) = (1 : a_1 \cdots, : a_n)$ embeds $\mathbb{A}^n$ in $\mathbb{P}^n$. As a matter of notation, Q will be reserved throughout to denote a point at infinity.

The one-dimensional projective space, also called the projective line, consists of the points $(1 : a_1)$, $a_1 \in \overline{\mathbb{F}}_q$ together with the point at infinity $(0 : 1)$ and this set has been used in the previous section for the construction of RS codes.

A polynomial $f \in \overline{\mathbb{F}}_q[x_1, x_2, \cdots, x_n]$ can be considered as a map f: $\mathbb{A}^n \to \overline{\mathbb{F}}_q$ defined by

$$f(P) = f(a_1, \cdots, a_n).$$

If $f(P) = 0$ we call P a *zero* of f.

More generally, with every $T \subseteq \overline{\mathbb{F}}_q[x_1, \cdots, x_n]$ we associate the *zero set of* T

$$Z(T) = \{P \in \mathbb{A}^n | f(P) = 0 \text{ for every } f \in T\}.$$

Definition 3.3: A subset V of $\mathbb{A}^n$ is called an *algebraic set* if there exists a $T \subset \overline{\mathbb{F}}_q[x_1, \cdots, x_n]$ such that

$$V = Z(T).$$

Definition 3.4: Let $V \subset \mathbb{A}^n$ be an algebraic set. The set

$$I(V) = \{f \in \overline{\mathbb{F}}_q[x_1, \cdots, x_n] | f(P) = 0 \text{ for every } P \in V\}$$

is called the ideal of V.

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

Lemma 3.5. Hilbert Nullstellensatz: Every maximal ideal of $\overline{\mathbb{F}}_q[x_1, \cdots, x_n]$ is of the form $\langle x_1 - a_1, \cdots, x_n - a_n \rangle$ with $a_i \in \overline{\mathbb{F}}_q$, $i = 1, \cdots, n$. For every element $P = (a_1, \cdots, a_n) \in \mathbb{A}^n$ the singleton $\{P\}$ is an algebraic set with ideal $I(P) = \langle x_1 - a_1, \cdots, x_n - a_n \rangle$.

Definition 3.6: An *affine variety* V in $\mathbb{A}^n$ is an algebraic set where $I(V)$ is a prime ideal. The set of $\mathbb{G}$-rational points of V is denoted $V(\mathbb{G})$. If $I(V)$ has a set of generators in $\mathbb{G}[x_1, \cdots, x_n]$ we say that V is *defined over* $\mathbb{G}$ and denote that $V/\mathbb{G}$. In this case we associate with the variety $V/\mathbb{G}$ the ideal

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

Definition 3.7: Let V be an affine variety. The quotient ring

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

is called the *coordinate ring* of V.

If V is defined over $\mathbb{G}$ the quotient ring

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

is called the *coordinate ring* of $V/\mathbb{G}$.

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

Since the ideal $I(V)$ of the variety V is a prime ideal the coordinate ring $\overline{\mathbb{F}}_q[V]$ is a domain. The following definition is therefore possible.

Definition 3.8: Let V be an affine variety. The field of fractions of $\overline{\mathbb{F}}_q[V]$, denoted $\overline{\mathbb{F}}_q(V)$ is called the *function field* of V. If V is defined over $\mathbb{G}$ we define the function field of $V/\mathbb{G}$, denoted $\mathbb{G}(V)$, as the field of fractions of $\mathbb{G}[V]$.

It follows from the definition of the function field $\overline{\mathbb{F}}_q(V)$ that it is a finitely generated extension of $\overline{\mathbb{F}}_q$, that is, there exists elements $x_1, \cdots, x_k \in \overline{\mathbb{F}}_q(V)$ such that $\overline{\mathbb{F}}_q(V) = \overline{\mathbb{F}}_q(x_1, x_2, \cdots, x_k)$.

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

Definition 3.9: An *affine curve* $\chi \subseteq \mathbb{A}^n$ is a variety of dimension 1.

As a matter of notation we will use χ to denote a curve in an algebraic geometry. When it is defined by a polynomial, we will denote the polynomial by F_χ or simply F when the curve is understood.

Example 3.10: Let $F \in \overline{\mathbb{F}}_q[x, y]$ be an irreducible polynomial and let us consider the variety

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

It is clear that the function field $\overline{\mathbb{F}}_q(\chi)$ has transcendence degree one, and therefore χ is an affine curve and since it is contained in $\mathbb{A}^2$ it is called an *affine plane curve*.

Example 3.11: In the affine plane, we consider the parabola V with equation $Y^2 = X$. Here the coordinate ring $\overline{\mathbb{F}}_q(V)$ consists of all the expressions of the form $A + By$, where A and B are in $\overline{\mathbb{F}}_q[x]$ and y satisfies $y^2 = x$. So, $\overline{\mathbb{F}}_q(V)$ is an algebraic extension of $\overline{\mathbb{F}}_q(x)$ by an element y, satisfying this equation of degree 2.

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

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

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

Example 3.12: The curve $F(x, y) = y^2 - x$ has a tangent line at the point $(x = 1, y = 1)$, $t_{(1,1)}(x, y) = -x + 2y - 1$ since $F_x(x, y) = -1$ and $F_y(x, y) = 2y$. On the other hand, a singular point occurs on the curve $F(x, y) = y^2 - x^3 + 2x^2 - x$ at the point $(x = 1, y = 0)$ since both derivatives $F_x(x, y) = -3x^2 + 4x - 1$ and $F_y(x, y) = 2y$ have a common zero at $(x = 1, y = 0)$ [1]. In this case, the curve has two distinct tangent lines at the singular point.

Definition 3.13: A curve χ is said to be *nonsingular* (or *smooth* or *regular*) if all the points on the curve are nonsingular, otherwise the curve is *singular*.

A more general definition of singularity will be given later in the section. In the example above, $y^2 - x$ is nonsingular while $y^2 - x^3 + 2x^2 - x$ is singular. A test for singularity of a curve χ is the existence, or not, of common zeros in the two partial derivatives.

Example 3.14: As an example over a finite field, consider the *Hermitian curve* from which an important class of codes will be considered in the next section. These curves will be used in a sequence of examples in this section. Consider the finite field $\mathbb{F}_q$ where $q = r^2 = p^{2m}$. The Hermitian curve is

described by the polynomial $F(x, y) = y^r + y - x^{r+1}$. The curve is nonsingular since the derivatives

$$F_x(x, y) = -(r+1)x^r = -x^r \quad (r = p = 0 \text{ in } \mathbb{F}_q)$$

and

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

have no roots in common (F_y has no roots).

A *monomial* of degree d is a polynomial

$$G \in \overline{\mathbb{F}}_q[x_0, \cdots, x_n]$$

of the form $G = a \cdot \prod x_i^{d_i}$ with $a \neq 0$ and $\sum_{i=0}^{n} d_i = d$, and a polynomial F is a *homogeneous polynomial* if F is the sum of monomials of the same degree.

A homogeneous polynomial $F \in \overline{\mathbb{F}}_q[x_0, \cdots, x_n]$ is said to have a *zero* at a point $P = (a_0 : a_1 : \cdots : a_n) \in \mathbb{P}^n/\overline{\mathbb{F}}_q$ if $F(a_0, a_1, \cdots, a_n) = 0$. This makes sense since

$$F(\lambda a_0, \cdots, \lambda a_n) = \lambda^d F(a_0, \cdots, a_n)$$

if F is homogeneous of degree d.

For a polynomial $f(x) \in \mathbb{F}_q[x]$ of degree d, the polynomial $y^d f(x/y)$ will be homogeneous of degree d. Conversely, one can reduce a homogeneous polynomial of degree d in n variables to a (nonhomogeneous) polynomial in $n-1$ variables.

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

$$Z(T) = \{P \in \mathbb{P}^n | f(P) = 0 \text{ for every } f \in T\}.$$

Definition 3.15: A subset V of $\mathbb{P}^n$ is called a projective algebraic set if there exists a set T of homogeneous polynomials such that

$$V = Z(T).$$

Definition 3.16: Let $V \subseteq \mathbb{P}^n$ be a projective algebraic set. The ideal in $\overline{\mathbb{F}}_q[x_0, \cdots, x_n]$ which is generated by all homogeneous polynomials F with $F(P) = 0$ for every $P \in V$ is called the ideal of V and is denoted $I(V)$.

Definition 3.17: A *projective variety* V in $\mathbb{P}^n$ is a projective algebraic set such that $I(V)$ is a prime ideal.

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

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

Definition 3.18: Let $V \subseteq \mathbb{P}^n$ be a nonempty projective variety. The quotient ring

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

is called the homogeneous coordinate ring of V. If V is defined over $\mathbb{G}$ then $\Gamma_h(V/\mathbb{G}) = \mathbb{G}[x_0, \cdots, x_n]/I(V/\mathbb{G})$.

An element $f \in \Gamma_h(V)$ is said to be a *form* of degree d if $f = F + I(V)$ where F is a homogeneous polynomial of degree d.

The function field of V is defined by

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

and

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

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

Definition 3.19: A *projective curve* $\chi \subseteq \mathbb{P}^n$ is a projective variety of dimension 1.

Example 3.20: Let $F \in \overline{\mathbb{F}}_q[X, Y, Z]$ be an irreducible homogeneous polynomial and let us consider the variety

$$\chi = \{F = 0\} = \{P \in \mathbb{P}^2 | F(P) = 0\}.$$

It is clear that this is a curve and since it is contained in $\mathbb{P}^2$ it is called the *projective curve*.

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

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

of degree d set

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

then F^* is a homogeneous polynomial of degree d in $n+1$ variables.

Consider now an affine variety $V \in \mathbb{A}^n$ and the corresponding ideal $I(V) \subset \overline{\mathbb{F}}_q[x_1, \cdots, x_n]$. Define the projective variety $\tilde{V} \subset \mathbb{P}^n$ as follows:

$$\tilde{V} = \{P \in \mathbb{P}^n | F^*(P) = 0 \text{ for all } F \in I(V)\}.$$

This variety is called the *projective closure* of V.

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

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

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

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

Then

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

is an affine variety and

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

and the projective closure of V is $\tilde{V}$.

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

Example 3.21: The projective closure of the Hermitian curve has the equation $y^r z + yz^r - x^{r+1} = 0$ and this curve has only one point at infinity, namely $(0 : 1 : 0)$.

B. The Local Ring at a Point

Let V be a variety and $P \in V$. If $f \in \overline{\mathbb{F}}_q(V)$ then $f = g/h$ with $g, h \in \overline{\mathbb{F}}_q[V]$ for an affine variety and $f = g/h$ with $g, h \in \Gamma_h(v)$ for a projective variety. If there exists a representative of $f = g/h$ and $h(P) \neq 0$, f is said to be *defined* at P.

The ring $O_P(V) = \{f \in \overline{\mathbb{F}}_q(V) | f \text{ is defined at } P\}$ is called the *local ring at* P.

The evaluation of an element $f \in O_P(V)$ is defined as $f(P) = g(P)/h(P)$ in the affine case and in the projective case let $g = G + I(V)$, $h = H + I(V) \in \Gamma_h(V)$ where G and H are homogeneous polynomials of degree d. Let $P = (a_0 : a_1 : \cdots : a_n)$. Since

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

we can put

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

if $H(P) \neq 0$.

$O_P(V)$ is indeed a local ring, its maximal ideal is

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

Definition 3.22: A *valuation ring* of the function field $\overline{\mathbb{F}_q}(V)$ is a ring O with the properties

- $\overline{\mathbb{F}}_q \subset O \subset \overline{\mathbb{F}}_q(V)$.
- For any $z \in \overline{\mathbb{F}}_q(V)$, $z \in O$, or $z^{-1} \in O$.

Theorem 3.23: Let O be a valuation ring of the function field $\overline{\mathbb{F}}_q(V)$. Then

- O is a local ring and has as unique maximal ideal $P = O \setminus O^*$ where $O^* = \{z \in O | \exists w \in O\colon zw = 1\}$.
- For $0 \neq x \in \overline{\mathbb{F}}_q(V)$, $x \in \mathcal{P} \Leftrightarrow x^{-1} \notin O$.
- $\mathcal{P}$ is a principal ideal.
- If $\mathcal{P} = tO$ then any $0 \neq z \in F$ has a unique representation of the form $z = t^n u$ for some $n \in \mathbb{Z}$, $u \in O^*$.
- O is a principal ideal domain. If $\mathcal{P}=tO$ and $\{0\}\neq I \subseteq O$ is an ideal then $I = t^n O$ for some $n \in \mathbb{N}$.

Definition 3.24: Let O be a valuation ring of $\overline{\mathbb{F}}_q(V)$ and $\mathcal{P}$ its unique maximal ideal with $\mathcal{P} = tO$. Then $z \in \overline{\mathbb{F}}_q(V)$ has a unique representation $z = t^n u$ with $u \in O^*$, $n \in \mathbb{Z}$. We define $\nu_{\mathcal{P}}(z) = n$ and $\nu_{\mathcal{P}}(0) = \infty$.

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

Theorem 3.25: The function $\nu\colon \overline{\mathbb{F}}_q(V) \to \mathbb{Z} \cup \{\infty\}$ satisfies

- $\nu_{\mathcal{P}}(x) = \infty \Leftrightarrow x = 0$
- $\nu_{\mathcal{P}}(xy) = \nu_{\mathcal{P}}(x) + \nu_{\mathcal{P}}(y)$
- $\nu_{\mathcal{P}}(x + y) \geq \min\{\nu_{\mathcal{P}}(x), \nu_{\mathcal{P}}(y)\}$ with equality if $\nu_{\mathcal{P}}(x) \neq \nu_{\mathcal{P}}(y)$
- $\exists z$ s.t. $\nu_{\mathcal{P}}(z) = 1$
- $\nu_{\mathcal{P}}(a) = 0$ for any $0 \neq a \in \overline{\mathbb{F}}_q$
- $\mathcal{P} = \{z \in \overline{\mathbb{F}}_q(V) | \nu_{\mathcal{P}}(z) > 0\}$
- $O = \{z \in \overline{\mathbb{F}}_q(V) | \nu_{\mathcal{P}}(z) \geq 0\}$
- $O^* = \{z \in \overline{\mathbb{F}}_q(V) | \nu_{\mathcal{P}}(z) = 0\}$.

A function satisfying the first five of these is called a *discrete valuation* and the ring O a *discrete valuation ring*.

We will now connect the points of a variety with discrete valuation rings of its function field. A more general definition of the singularity of a curve or variety than the one given earlier, follows:

Definition 3.26: Let V be a variety and $I(V) = \langle G_1, G_2, \cdots, G_s \rangle$. Let P be a point of V and consider the matrix

$$J_{V,P} = \{a_{ij}\}$$

where

$$a_{ij} = (\partial G_i / \partial x_j)(P)$$

for $i = 1, \cdots, s$, and $j = 1, \cdots, n$ (affine case) or $j = 0, 1, \cdots, n$ (projective case).

P is called *nonsingular* if

$$\operatorname{rank} J_{V,P} = n - \dim V$$

and *singular* otherwise. The variety V is called singular if it has at least one singular point and *regular* otherwise.

Theorem 3.27: Let χ be a curve (projective or affine) and P a point of χ. P is nonsingular if and only if $O_P(\chi)$ is a discrete valuation ring.

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

Let $\mathcal{P}_\chi$ denote the set of closed points of the curve χ.

Example 3.28: We will consider the projective plane curve χ with equation $zy^2 + yz^2 = x^3$ over the field $\overline{\mathbb{F}}_2$. In $Q = (0 : 1 : 1)$, we can take $t = x/z$ as a local parameter. Let $f = x/(y + z)$. We will determine $\nu_Q(f)$. We have

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

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

C. Divisors, the Vector Space $L(G)$, and the Theorem of Riemann–Roch

Let χ be a regular projective curve defined over $\mathbb{F}_q$. A *divisor* of χ is a formal sum

$$D = \sum_{P \in \chi} n_P P$$

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

$$\deg D = \sum_{P \in \chi} n_P \deg P.$$

The divisors of χ form an additive group $D(\chi)$, the *divisor group* of χ.

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

$$(f)_0 = \sum_{\nu_P(f)>0} \nu_P(f)P$$

and the *pole divisor* of f is

$$(f)_\infty = -\sum_{\nu_P(f)<0} \nu_P(f)P.$$

The degree of a principal divisor is zero which gives that

$$-\sum_{\nu_P(f)<0} \nu_P(f)P = \sum_{\nu_P(f)>0} \nu_P(f)P.$$

On $D(\chi)$ we define a partial order by

$$D_1 = \sum_{P\in\mathcal{P}_\chi} m_P P \leq D_2 = \sum_{P\in\mathcal{P}_\chi} n_P P \Leftrightarrow m_P \leq n_P, \qquad \text{for all } P \in \chi.$$

Definition 3.29: If $G \in D(\chi)$ let

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

be the set of rational functions with poles only at the zeros of the divisor G and have zeros at the poles of G.

Notice that the divisor of a product of two functions is the sum of the respective divisors, $(f \cdot h) = (f) + (h)$, and the divisor of the sum of two functions $(f + h)$ satisfies $(f + h) \geq \min\{(f), (h)\}$, i.e., the minimum coefficient is chosen, point by point. $L(G)$ is a finite-dimensional vector space over $\mathbb{F}_q$, its dimension is denoted $l(G)$. The *Theorem of Riemann* says that there exists a nonnegative integer m such that for every divisor G of χ

$$l(G) \geq \deg(G) + 1 - m$$

and the smallest nonnegative integer with this property is called the *genus* and is denoted by $g(\chi)$ or g.

In order to determine $l(G)$ one needs the so-called *differentials*. We can think of differentials as objects of the form fdh where f and h are rational functions, i.e., elements of $\mathbb{F}_q(\chi)$, such that the map which sends h to dh is a *derivation*. A derivation is $\mathbb{F}_q$-linear and the *Leibnitz rule* $d(h_1h_2) = h_1dh_2 + h_2dh_1$ holds. We denote the set of differentials on χ by Ω_χ. One can talk about zeros and poles of differentials. At every closed point P there exists a *local parameter* that is, a function u such that $\nu_P(u) = 1$, and for every differential ω there exists a function f such that $\omega = fdu$. The valuation $\nu_P(\omega)$ is now by definition $\nu_P(f)$, so we say that ω has a zero of order ρ if $\rho = \nu_P(f) > 0$ and ω has a pole of order ρ if $\rho = -\nu_P(f) > 0$. The divisor of ω is by definition $(\omega) = \sum \nu_P(u)P$. The divisor of a differential is called *canonical* and always has degree $2g - 2$.

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

$$\Omega(G) = \{\omega \in \Omega_\chi | w = 0 \text{ or } (\omega) \geq G\}.$$

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

Theorem 3.30. Riemann–Roch: For a divisor G of a curve of genus g

$$l(G) = \deg(G) + 1 - g + i(G).$$

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

Moreover it is a consequence of the Riemann–Roch theorem that

Theorem 3.31: For any divisor G with $\deg(G) > 2g - 2$,

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

Let ω be a differential. If P is a closed point of degree m and u is a local parameter at P, then there exists a rational function f such that $\omega = fdu$. This function f has a formal Laurent series $\sum_{i=\rho}^{\infty} a_i u^i$, where the coefficients $a_i \in \mathbb{F}_{q^m}$ and $\rho = \nu_P(\omega)$ and $a_P \neq 0$. The *residue of* ω at P is by definition $\mathrm{Tr}(a_{-1})$ and is denoted by $\mathrm{Res}_P(\omega)$, where Tr is the trace map from $\mathbb{F}_{q^m}$ to $\mathbb{F}_q$.

The *residue theorem* states that for $\omega \in \Omega_\chi$

$$\sum_{P\in\mathcal{P}_\chi} \mathrm{Res}_P(\omega) = 0.$$

Let P be a point of degree one. An integer $n \geq 0$ is called a *pole number* of P iff there exists an $f \in \overline{\mathbb{F}}_q(\chi)$ with $(f)_\infty = nP$. Otherwise, n is called a *gap number* of P. Clearly, n is a pole number if $l(nP) > l((n-1)P)$. Moreover, the set of pole numbers form an additive semigroup since if $(f_1)_\infty = n_1P$ and $(f_2)_\infty = n_2P$ then $(f_1 \cdot f_2)_\infty = (n_1 + n_2)P$.

Theorem 3.32: Suppose $g > 0$ and P is a closed point of degree one. Then there are exactly g gap numbers $i_1 < i_2 < \cdots < i_g$ of P and $i_1 = 1$ and $i_g \leq 2g - 1$.

An important case from the perspective of algebraic-geometry codes is when the curve χ is nonsingular and intersects the line at infinity in a single point, Q say. In this case the elements of $R = \bigcup_{m=0}^{\infty} L(mQ)$ has a simple description, since the rational functions $X = x/z$ and $Y = y/z$ represent a monomial generating set for R.

Example 3.33: Consider the Hermitian curve with equation $x^{r+1} = y^r + y$ over the field $\mathbb{F}_{r^2}$. Here $Q = (0 : 1 : 0)$ and $X = x/z$, $Y = y/z$ is a monomial generating set for

$$R = \bigcup_{m=0}^{\infty} L(mQ).$$

It is obvious that the sets $\{X^iY^j | 0 \leq i < r\}$ and $\{X^iY^j | 0 \leq i < r + 1\}$ each describes bases for R.

TABLE I
GAPS FOR THE HERMITIAN CURVE AT THE POINT Q

q $(= r^2)$	r	Number of points	genus g	gaps
4	2	9	1	$\{1\}$
9	3	28	3	$\{1, 2, 5\}$
16	4	65	6	$\{1, 2, 3, 6, 7, 11\}$
25	5	126	10	$\{1, 2, 3, 4, 7, 8, 9, 13, 14, 19\}$
64	8	513	23	$\{1, 2, 3, 4, 5, 6, 7, 10, 11, 12, 13, 14, 15, 19,$ $20, 21, 22, 23, 28, 29, 30, 31, 37, 38, 39, 46, 47, 55\}$

The notion of gaps and the genus of the curve are closely related in this situation. As before, let $X = x/z$ and $Y = y/z$ and let $o_X > 0$ and $o_Y > 0$ be the pole orders at Q of these two functions. The semigroup of gaps is then generated by o_X and o_Y, so the genus of the curve is the number of elements in $\mathbb{N}$ that are not of the form $io_X + jo_Y$, $i, j \in \mathbb{N}$. For example, if $o_X = 3$ and $o_Y = 4$ then the gaps are $\{1, 2, 5\}$, which in turn implies that there are no rational functions on the curve with these pole orders at Q.

Example 3.34: As an example, the Hermitian curve over $\mathbb{F}_q$, $q = r^2$ is regular and has genus $g = r(r-1)/2$. The order of X and Y is r and $r+1$, respectively. To see this first consider the function $X = (x/z)$. The equation $x = 0$ describes a line in the plane. The intersection with the Hermitian curve, described by $y^r z + yz^r - x^{r+1}$, are single points of the form $(0 : \beta : 1)$ where $\beta^r + \beta = 0$. There are exactly r solutions for $\beta \in \mathbb{F}_q$. These are the simple zeros of the function X over the curve. Thus we conclude that X has r zeros, and thus r poles at the point Q and so the degree of X is $o_X = r$. In the case of $Y = y/z$, the zeros in the plane correspond to the line $y = 0$, intersects the Hermitian curve at the single point $(0 : 0 : 1)$. However, the order of this single root is $r+1$. This implies that the pole order at Q, and thus the degree of Y on the curve, is equal to $o_Y = r+1$. Table I shows some of these results for small values of r. The discussion can be cast more algebraically by saying that at $Q = (0 : 1 : 0)$, the semigroup S of pole numbers are generated by the divisors rQ and $(r+1)Q$, that is, $S = \{ar + b(r+1) | a, b \in \mathbb{N}_0\}$, and it can be seen that $(x/z)_\infty = rQ$ and $(y/z)_\infty = (r+1)Q$ which implies that $L(mQ)$ has the functions $\{(x/z)^a (y/z)^b | ar + b(r+1) \le m\}$ as a basis. The above computation of the genus of the curve, noted above, follows from this basis.

Example 3.35: We can directly calculate the dimension of $L(aQ)$. We get

$$\dim(L(aQ)) = \begin{cases} 1, & a = 0 \\ 1 + a - m(a), & 0 < a < 2g \\ a - g + 1, & 2g \le a \end{cases}$$

where $m(a) = |\{i | i \text{ is a gap}, i \le a\}|$. Note that $m(a) = g$ for $a \ge 2g$.

D. Counting Points on Curves

Let χ be a regular curve defined over $\mathbb{F}_q$ and let N_m be the number of points on χ of degree one over $\mathbb{F}_{q^m}$.

Definition 3.36: The zeta function of χ is defined as

$$Z(t) = \exp \sum_{m=1}^{\infty} N_m \frac{t^m}{m}.$$

The zeta function contains information about the number of points in various extensions of $\mathbb{F}_q$. It has the following property.

Theorem 3.37. Hasse–Weil: Let g be the genus of χ. Then

$$Z(t) = \frac{P(t)}{(1-t)(1-qt)}$$

where

$$P(t) \in \mathbb{Z}[t], \qquad P(t) = \prod_{i=1}^{2g} (1 - \alpha_i t)$$

where

$$\alpha_i \in \mathbb{C}, \qquad |\alpha_i| = \sqrt{q},\ \alpha_i \alpha_{2g+1-i} = q$$

and

$$N_m = q^m + 1 - \sum_{i=1}^{2g} \alpha_i^r$$

and the α_i are complex algebraic integers.

The proof of $|\alpha_i| = \sqrt{q}$ is difficult. It is an analog of the Riemann hypothesis for curves over finite fields and was proved by Weil. It has as a consequence the Hasse–Weil bound.

Corollary 3.38. The Hasse–Weil Bound:

$$|N_m - (1 + q^m)| \le 2g\sqrt{q^m}.$$

Example 3.39: The Hermitian curve considered in Example 3.34 has $1 + r^2 + r(r-1)r = r^3 + 1$ points of degree one over $\mathbb{F}_q$ so $N_1 = 1 + q + 2g\sqrt{q}$ and is therefore optimal with respect to the Hasse–Weil bound. To calculate the number of points we first note that $(0 : 1 : 0)$ is the only point with $z = 0$. If $z = 1$ we have $x^{r+1} = y^r + y$. The right-hand side is the trace function from $\mathbb{F}_{r^2}$ to $\mathbb{F}_r$ so from each of the r values of y, where $y^r + y = 0$ we get one solution and from the $r^2 - r$ values of y where $y^r + y \ne 0$ we get $r+1$ x's. This gives $r + (r+1)(r^2 - r) + 1$ points, that is, $r^3 + 1$.

E. Algebraic-Geometry Codes

The two code constructions at the end of Section II, one consisting of evaluating rational functions at a sequence of points, such as the case for RS codes and polynomial functions, the other evaluating residues of rational functions at a sequence of points, such as for Goppa codes, will be emulated for the case when the sequence of points is obtained from curves in an algebraic geometry.

In the first instance, let χ be a nonsingular projective curve over $\mathbb{F}_q$ of genus g and let $P_1, P_2, \cdots, P_n$ be rational points on χ and $D = P_1 + P_2 + \cdots + P_n$, a divisor. Let G be a divisor with support disjoint from D as noted, and assume that $2g - 2 < \deg(G) < n$.

Define the linear code $C(D, G)$ over $\mathbb{F}_q$ as the image of the linear map

$$\begin{aligned} \alpha\colon L(G) &\longrightarrow \mathbb{F}_q^n \\ f &\longmapsto (f(P_1), f(P_2), \cdots, f(P_n)) \end{aligned}$$

where

$$L(G) = \{f \in \overline{\mathbb{F}}_q(C) | (f) + G \geq 0\} \cup \{0\}$$

where $\overline{\mathbb{F}}_q(\chi)$ is the function field of the curve χ and

$$(f) = \sum_P \operatorname{ord}_P(f) P.$$

The parameters of the code are established by using the properties discussed in the previous section. The kernel of the map is the set $L(G - D)$ and

$$\begin{aligned} k &= \dim C(D, G) = \dim L(G) - \dim L(G - D) \\ &= \deg(G) - g + 1 \end{aligned}$$

since $\dim(L(G - D)) = 0$ if $\deg G \leq n$. The minimum distance follows from the following theorem.

Theorem 3.40: The minimum distance d of the code $C(D, G)$ satisfies

$$d \geq d^* = n - \deg(G).$$

Proof: $F \in L(G)$ has at most $\deg(G)$ zeros. □

Thus the designed minimum distance of $C(D, G)$ is within g of the Singleton bound.

To emulate the residue construction of the classical Goppa codes, choose G and D as in the previous construction and recall that for a divisor $D \in D(\chi)$ of the curve χ

$$\Omega(D) = \{\omega \in \Omega | \operatorname{div}(\omega) \geq D\} \cup 0$$

where Ω is the set of differentials. Define the map

$$\begin{aligned} \alpha^*\colon \Omega(G - D) &\longrightarrow \mathbb{F}_q^n \\ \omega &\longmapsto (\operatorname{Res}_{P_1}(\omega), \operatorname{Res}_{P_2}(\omega), \cdots, \operatorname{Res}_{P_n}(\omega)). \end{aligned}$$

The code $C^*(D, G)$ is defined as the image under α^*. Again, from the properties developed in the previous section, in particular as a consequence of the Riemann–Roch theorem, it is straightforward to establish that

$$\begin{aligned} \dim(C^*(D, G)) &= n - \deg(G) + g - 1 \\ d^* &\geq \deg(G) - 2g + 2 \end{aligned}$$

again within g of the Singleton bound.

It follows from the Residue theorem that the codes $C(D, G)$ and $C^*(D, G)$ are duals of each other. Furthermore, it is possible to show [97] that there exists a rational differential form ω with simple poles and with residue 1 at the points P_i, $i = 1, 2, \cdots, n$ so that

$$C^*(D, G) = C(D, K + D - G)$$

with K the divisor of ω. This implies that the residue construction gives exactly the same class of codes as the first construction. It is nonetheless useful to retain the two approaches to code construction.

The next section considers some particular classes of curves and constructions of codes by the methods given here.

IV. Classes of Algebraic-Geometry Codes and Their Properties

The previous section has established constructions of codes from algebraic curves as a natural evolution from RS and Goppa codes. Some classes of codes of particular interest that arise from these constructions applied to specific classes of curves are considered here. As a matter of notation, let $N_q(g)$ be the maximum number of points possible on a curve of genus g over $\mathbb{F}_q$. As in the previous section, for a specific curve we will denote the number of rational points of the curve over $\mathbb{F}_{q^m}$ by N_m, where the genus and field size q are understood.

A. Codes from the Klein Quartic

The homogeneous curve χ

$$x^3y + y^3z + z^3x = 0$$

is referred to as the Klein quartic [83] which can be considered over any field. Interest in this curve will often be for fields of characteristic two. Since the curve is nonsingular of degree $d = 4$ over fields of characteristic not equal to 7 (and the curve is singular in that case), the genus of the curve is, by the Plücker formula

$$g = \frac{1}{2}(d-1)(d-2) = 3.$$

Consider the number of points on such a curve over a field of characteristic 2. It can be seen from the zeta function in Section III, that to determine the number of points on the curve over any extension field, it is sufficient to determine the number of points over $\mathbb{F}_{2^r}$, $r = 1, 2, 3$.

Over $\mathbb{F}_2$ the homogeneous equation has the three solutions $P_0 = (1 : 0 : 0)$, $P_1 = (0 : 1 : 0)$, and $P_2 = (0 : 0 : 1)$. To determine the number of points over $\mathbb{F}_{2^2}$ argue as follows. For $z \neq 0$ convert the equation to projective coordinates and define $u = x/z$, $v = y/z$ to give

$$u^3v + v^3 + u = 0.$$

Consider solutions of the form $(\beta, \gamma, 1)$, $\beta, \beta \in \mathbb{F}_4^*$, $\mathbb{F}_4 = \{0, 1, \alpha, \alpha^2\}$ where $\alpha^2 + \alpha + 1 = 0$. For a fixed $\beta \in \mathbb{F}_4^*$, the equation reduces to

$$v^3 + \beta^3 v + \beta = 0.$$

If $\beta = 1$, the polynomial is irreducible over $\mathbb{F}_4$ and there are no solutions. For $\beta = \alpha$, there is one solution $(\alpha, \alpha^2, 1)$ and for $\beta = \alpha^2$ the single solution is $(\alpha^2, \alpha, 1)$ giving $N_2 = 5$.

To obtain the points over $\mathbb{F}_{2^3} = \{0, 1, \alpha, \alpha^2, \cdots, \alpha^6\}$ where $\alpha^3 + \alpha + 1 = 0$, α primitive, it is readily checked that

$$\sigma_1 \colon (x, y, z) \longmapsto (\alpha x, \alpha^4 y, \alpha^2 z)$$

and

$$\sigma_2 \colon (x, y, z) \longmapsto (z, x, y)$$

are automorphisms of the set of points, σ_1 of order 7 and σ_2 of order 3. The point $P_0 = (1, \alpha^2, \alpha^4)$ is a solution as are

$$P_{ij} = \sigma_1^i \sigma_2^j P_0, \qquad i = 0, 1, \cdots, 6, \;\; j = 0, 1, 2.$$

These 21 points plus the original three points yield $N_3 = 24$. The numerator of the zeta function for the Klein quartic is then obtained as

$$P_\chi(t) = 1 + 5t^3 + 8t^6$$

and the zeta function is

$$Z(t) = \frac{1 + 5t^3 + 8t^6}{(1-t)(1-2t)}.$$

The number of points on the curve over $\mathbb{F}_{2^i}$, N_i, is the coefficient of t^i in the series expansion of $Z(t)$.

Codes of differing lengths can be defined with the Klein quartic. Following the work of Hansen [39], define a set of codes of length 21 over $\mathbb{F}_8$. Using the evaluation construction of the previous section, define $D = \sum P_{ij}$ and the divisor with disjoint support

$$G = m(P_0 + P_1 + P_2), \qquad 2 \le m \le 6.$$

The code is defined by the mapping

$$\alpha \colon L(G) \longrightarrow \mathbb{F}_8^{21}$$
$$f \longmapsto (f(P_{ij})), \qquad i = 0, 1, 2, \cdots, 6, \;\; j = 0, 1, 2.$$

Using the results of previous sections it can be shown the dimension of $L(G)$ over $\mathbb{F}_8$ is $3m - 2$ and the minimum distance of the code is $\ge 21 - 3m$. The codes have the parameters

$$(21, 3m - 2, \ge 21 - 3m), \qquad 2 \le m \le 6$$

and in fact the lower bound on the minimum distance is achieved for all values of m in the range shown.

In a similar fashion, define a code of length 23 by choosing $G = 10P_0$ and D the sum of the other 23 points. In this case, the code has the parameters $n = 23$, $k = 8$, $d = 13$.

B. Codes from Elliptic and Hyperelliptic Curves

An elliptic curve in homogeneous coordinates over a field K (more formally taken to be the algebraic closure $\overline{K}$) is irreducible and of the form

$$u^2 w + a_1 uvw + a_3 vw^2 = u^3 + a_2 u^2 w + a_4 uw^2 + a_6 w^3, \qquad a_i \in \overline{K}.$$

All such curves are of genus 1. Making the transformation $x = u/w$, $y = v/w$ yields

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6, \qquad a_i \in \overline{K}. \quad (1)$$

In homogeneous coordinates, the point at infinity is $Q = (0 : 1 : 0)$. If the coordinates are in K, the elliptic curve $E(K)$ is said to be over K.

For any such curve, it is possible to define an addition of points by observing that the straight line through any two points P_1, P_2 intersects the curve in a unique third point P_3. The "addition" of P_1 and P_2 is then defined as $Q \oplus P_3$ where $\oplus$ represents addition of points on the curve. Such an addition of points is easy to appreciate over the reals and all the geometric notions involved have straightforward analogs over a finite field. Thus the points of the elliptic curve over $\mathbb{F}_q$, $E(\mathbb{F}_q)$, have the structure of an Abelian group under this operation of point addition. This structure is the basis of elliptic curve cryptosystems [53], [62] which has generated further interest in the subject, with its deep results and important connections to number theory and other areas of mathematics [81].

In terms of the coefficients in (1) [62, p. 19], two fundamental quantities for the curve can be defined, the discriminant Δ, and the j-invariant $j(E)$. The curve is nonsingular iff $\Delta \neq 0$ and, if two curves are isomorphic they have the same j-invariant. Elliptic curves are well classified in terms of their isomorphism classes and their group structure [81] but such information is more than is required for our objectives.

Let $P_1, P_2, \cdots, P_n$ be rational points of the elliptic curve and Q the point at infinity. Choose the divisor D as $D = P_1 + P_2 + \cdots, P_n$ and as $G = mQ$ for some positive integer m. The code $C(D, G)$ over $\mathbb{F}_q$ obtained has the parameters $(n, m, d \ge n - m)$ and

$$|n - (q+1)| \le \lfloor 2\sqrt{q} \rfloor.$$

Almost all of these codes will have a minimum distance of $d = n - m$ falling short of the Singleton bound by one. Van der Geer [94, p. 32] gives an example, credited to R. Pellikaan, where a $(6, 3, 4)$ code over $\mathbb{F}_4$, obtained from an elliptic curve, is actually an MDS code.

In the case when $G = mQ$, it is known [14] that $L(mQ)$ has a basis of the polynomials

$$\{1, x, x^2, \cdots, x^\delta, y, yx, \cdots, yx^{\hat{\delta}}\}$$
$$\delta = \left\lfloor \frac{m}{2} \right\rfloor \qquad \hat{\delta} = \left\lfloor \frac{m-2}{2} \right\rfloor$$

to give a code dimension of $\delta + \hat{\delta} + 2 = \lfloor m/2 \rfloor + \lfloor (m-2)/2 \rfloor + 2$ which is $m+1$ if m is even and m if m is odd. The minimum distance of the code is $n - m$.

To determine the number of points on the curve over $\mathbb{F}_{q^k}$, it is sufficient to determine the number of points over $\mathbb{F}_q$, since the curves are of genus 1. Thus if $t = q + 1 - |E(\mathbb{F}_q)|$ then

$$|E(\mathbb{F}_{q^k})| = q^k + 1 - \alpha^k - \beta^k$$

where α and β are solutions of the equation $x^2 - tx + q = 0$.

The weight enumerator of MDS codes is uniquely specified by the code parameters. Since for (most) codes from elliptic curves, $d = n - k$, it is not surprising that some information is available for such codes. The subject has been considered by Katsman and Tsfasman, [50] (and in the book [90]). Define

the weight enumerator of such a code C by

$$W_C(x) = \sum_{i=0}^{n} A_{n-i}x^i$$

or

$$W_C(x, y) = x^n + y^d \sum_{i=0}^{n-d} A_{d+i}y^i x^{n-d-i}$$

and, by the MacWilliams identities,

$$W_{C^\perp}(x, y) = q^{-k}W_C(x + (q-1)y, x - y).$$

By applying these relationships to the case of an $(n, k, d = n-k)$ code, it is found that [90, p. 302]

$$W_C(x) = x^n + \sum_{i=0}^{k-1} \binom{n}{i}(q^{k-i} - 1)(x-1)^i + B_k(x-1)^k.$$

It is known that $B'_{n-k} = B_k$, where the prime indicates weights in the dual code. For $n = 2k$, $W_C(x) = W_{C^\perp}(x)$ and the codes are formally self-dual. If $B_k = 0$ the code is MDS. In general, it is possible to show that

$$0 < B_k = B'_{n-k} \le \binom{n}{k}(q-1), \qquad d = n-k$$

and the determination of B_k uniquely specifies the weight enumerator of the code and hence of its dual. It is possible to characterize B_k as $(q-1)M$ where M is an integer with a combinatorial interpretation in terms of curve parameters [90, p. 303] but this is beyond our interests.

Recent authors ([9], [21]) have defined the defect of an $(n, k, d = n-k+1-\delta)$ code to be $\delta > 0$, i.e., the amount by which the minimum distance of the code fails to meet the Singleton bound. Codes with defect $\delta = 1$ which includes most of the codes derived from elliptic curves, are referred to as quasi- or almost-MDS codes. Further structural properties of such curves can be established.

Clearly, curves with small genus and large numbers of points are of interest in constructing codes. The quantity $N_q(g)$, the largest number of points over $\mathbb{F}_q$ for any curve of genus g, has been studied by several authors (e.g., [16], [58]) and the results find implications in coding theory. These refinements of the Hasse–Weil bound (and the Serre improvement) are briefly discussed for curves of small genus.

For curves of genus one, let $q = p^e$, $p = \mathrm{char}\,(\mathbb{F}_q)$. The maximum number of points of a genus one curve then is

$$N_q(1) = q + 1 + \lfloor 2\sqrt{q} \rfloor$$

unless $p | \lfloor 2\sqrt{q} \rfloor$ and e is an odd integer, in which case

$$N_q(1) = q + \lfloor 2\sqrt{q} \rfloor.$$

The case of genus 2 is a little more complicated but, interestingly, a complete answer is still possible [78]. If q is an even power of a prime, $q \neq 4, 9$, then

$$N_q(2) = q + 1 + 4\sqrt{q}$$

and $N_4(2) = 10$, $N_9(2) = 20$. If q is an odd power of a prime p call q special if $p | \lfloor 2\sqrt{q} \rfloor$ or if there is an integer l such that $q = l^2 + l + 1$ or $q = l^2 + l + 2$. If q is not special then

$$N_q(2) = q + 1 + 2\lfloor 2\sqrt{q} \rfloor$$

and if q is special then

$$N_q(2) = \begin{cases} q + 2\lfloor 2\sqrt{q} \rfloor, & \text{if } \{2\sqrt{q}\} > (\sqrt{5}-1)/2 \\ q + 2\lfloor 2\sqrt{q} \rfloor - 1, & \text{if } \{2\sqrt{q}\} < (\sqrt{5}-1)/2 \end{cases}$$

where $\{2\sqrt{q}\} = 2\sqrt{q} - \lfloor 2\sqrt{q} \rfloor$, the fractional part.

For $N_q(3)$ specific results are available. Thus [30] for $3 < q \le 19$, $N_q(3) = 2q + 6$ except for $q = 8, 9$ where $N_q(3) = 4q - 8$. Tables for such functions for $q \le 25$ are given ([2], [90]).

As a generalization of elliptic curves, consider the hyperelliptic curves, defined by an equation of the form

$$\mathcal{X}\colon y^2 + h(x)y = k(x), \qquad h(x), k(x) \in K[x]$$

where $h(x)$ is a polynomial of degree at most g and $k(x)$ is monic of degree exactly $2g + 1$. We require the curve to be nonsingular, i.e., have no singular points $(x, y) \in \overline{K}^2$ where both of the partial derivatives $2y + h(x) = 0$ and $h'y - k'(x) = 0$ vanish. In such a case the genus of the curve is g. Notice that for elliptic curves $h(x)$ is at most a linear polynomial and $k(x)$ a monic cubic, and the curve is of genus 1.

If $\mathrm{char}\,(K) \neq 2$ then the change of variables

$$x \mapsto x, \; y \mapsto y - (h(x)/2)$$

transforms the equation to

$$y^2 = f(x) \qquad \deg f(x) = 2g + 1.$$

In this case, if $P = (x, y)$ is a point on the curve, then $(x, -y - h(x))$ is also a point on the curve, the sum (addition of points on the curve, $\oplus$) of P and the point at infinity. If $\mathrm{char}\,(K) = 2$ and (x, y) is on the curve, then so also is $(x, y + h(x))$. In homogeneous coordinates the point at infinity is $(0 : 1 : 0)$ and has multiplicity $2g - 1$. The number of rational points on the curve N is bounded by

$$|N - (q+1)| \le g\lfloor 2\sqrt{q} \rfloor.$$

The following example [53], [94, p. 61], is instructive. Consider the curve

$$y^2 + y = x^5 + 1$$

of genus 2 over $\mathbb{F}_2$. It can be shown using the techniques established previously that the number of points on this curve over $\mathbb{F}_{2^k}$ is $2^k + 1$ unless $k = 4m$ in which case it is

$$2^{4m} + 1 + (-1)^{m+1}2^{2m+2}.$$

One can use either of the code constructions for hyperelliptic curves to obtain codes. Continuing the previous example [94, p. 61], over $\mathbb{F}_{2^4}$ the curve has 33 points. Let D be the sum of the points not at infinity and $G = mQ$, Q the point at infinity. For the code $C^*(D, G)$ one obtains a sequence of codes $(32, k = 33 - m, d_m)$ over $\mathbb{F}_{16}$ with $d_m = m - 2$ for

$3 \le m \le 31$, $m \ne 3, 5$, $d_3 = 2$, $d_5 = 4$. The general question of the maximum length possible for a code over $\mathbb{F}_q$ from a curve of genus either one or two is examined in [6] and [66].

C. Codes from Hermitian Curves

The Hermitian curve in homogeneous coordinates is given by

$$u^{q+1} + v^{q+1} + w^{q+1} = 0$$

over $\mathbb{F}_{q^2}$, a special case of the Fermat curve

$$u^m + v^m + w^m = 0$$

for $(m, q) = 1$. The Hermitian curve is nonsingular and its genus is given by $g = q(q-1)/2$. It will be shown constructively that the number of rational points on the curve is given by $q^3 + 1$ and since, by the Hasse–Weil theorem

$$N_q \le q^2 + 1 + 2g\sqrt{q^2} = q^3 + 1$$

all such curves are maximal. In projective coordinates the curve is written

$$u^{q+1} + v^{q+1} = 1.$$

Choose [30, p. 1558] $\gamma, \delta \in \mathbb{F}_{q^2}$ so that

$$\gamma^q + \gamma = \delta^{q+1} = -1$$

which is always possible (note that the left side is the trace over $\mathbb{F}_q$ and $\delta^{q+1} \in \mathbb{F}_q$). Make the transformations

$$x = \frac{1}{v - \delta u} \text{ and } y = \delta u x - \gamma$$

to yield the equation

$$y^q + y = x^{q+1}.$$

The common pole of x and y is the point at infinity.

To describe the q^3 rational points on the curve [83, p. 203] in slightly more detail than previously (Example 3.39), note that for each $\alpha \in \mathbb{F}_{q^2}$ there exists q distinct elements $\beta \in \mathbb{F}_{q^2}$ such that

$$\mathrm{Tr}\,(\beta) = \beta^q + \beta = \alpha^{q+1} \in \mathbb{F}_q$$

and the q^3 solutions are (α, β).

To form a code of length $n = q^3$ over $\mathbb{F}_{q^2}$, take D as the sum of the q^3 rational points and $G = mQ$ for a positive integer m. It can be shown [83] that the elements

$$\{x^i y^j,\ i \ge 0,\ 0 \le j \le q-1 \text{ and } iq + j(q+1) \le m\}$$

forms a basis of $L(mQ)$. The above monomials can be used to construct a generator matrix of the code $C(D, mQ) = C_m$.

To determine the code parameters, define

$$A(m) = \{0 \le l \le m | i,\ j \in Z,\ l = iq + j(q+1),\ i \ge 0,\ 0 \le j \le q-1\}$$

and let $\nu(m) = |A(m)|$.

To determine the dimension of the code C_m, note that for $m > 0$, C_m is empty, and for $m > q^3 + q^2 - q - 2$, $\dim(C_m) = n = q^3$. Thus the interesting range for m is $0 \le m \le q^3 + q^2 - q - 2$.

It is first noted that the dual code to C_m is

$$C_m^{\perp} = C_{q^3+q^2-q-2-m}$$

and hence C_m is self-orthogonal if $2m \le q^3 + q^2 - q - 2$ and self-dual iff $m = (q^3 + q^2 - q - 2)/2$, a case that is only possible if $q = 2^k$ for some positive integer k. The dimension of C_m is given by the cases

$$\dim(C_m) = \begin{cases} m + 1 - (q^2 - q)/2, & \text{if } q^2 - q - 2 < m < q^3 \\ \nu(m), & \text{if } 0 \le m \le q^2 - q - 2 \\ q^3 - \nu(q^3 + q^2 - q - 2 - m), & \text{if } m > q^3. \end{cases}$$

The minimum distance of C_m satisfies

$$d \ge q^3 - m.$$

It can be shown that $d = q^3 - m$ when

$$q^2 - q \le r \le q^3 - q^2 + q.$$

Finally, it is noted [83] that the automorphism group of C_m is quite large. As before, let α, β be such that

$$\beta^q + \beta = \alpha^{q+1}, \qquad \alpha, \beta \in \mathbb{F}_{q^2}.$$

For each of the q^2 values of α there are exactly q values of β satisfying the equation. For $\epsilon \in \mathbb{F}_{q^2} \setminus \{0\}$ it is verified that if (x, y) is a point on the curve then so is $(\sigma(x), \sigma(y))$ where

$$\sigma(x) = \epsilon x + \delta \qquad \sigma(y) = \epsilon^{q+1} y + \epsilon \delta^q x + \mu.$$

Thus the automorphism group of the code contains a subgroup of size $q^3(q^2 - 1)$. More recently it has been shown by Xing [107] that, for $q + 1 \le m \le q^3 + q^2 - 2q - 3$, this is in fact, precisely the automorphism group. For either $0 \le m \le q - 1$ or $m \ge q^3 + q^2 - 2q - 1$

$$\mathrm{Aut}\,(C_m) \cong S_{q^3}$$

the symmetric group on q^3 letters. For the two cases of $m = q$ or $m = q^3 + q^2 - 2q - 2$ the group is slightly more complex, of the form

$$\mathrm{Aut}\,(C_m) \cong \mathrm{AGL}\,(2, q^2) \otimes S_q^{q^2}$$

where $\mathrm{AGL}\,(2, q^2)$ is the affine linear transformation of a line over $\mathbb{F}_{q^2}$ and $S_q^{q^2}$ is a copy of the q^2 symmetric group S_q. Notice that in the case $m = q$, $\dim(C_m) = 2$ and C_m is generated by the all-ones vector and

$$(x_1, \cdots, x_1, x_2, \cdots, x_2, \cdots, x_{q^2}, \cdots, x_{q^2})$$

from which the result follows. The case for $m = q^3 + q^2 - 2q - 2$ follows from duality.

The Hermitian codes clearly have interesting structure. Given their monomial basis of the form $x^i y^j$ with restrictions on the size of i and j, it is not surprising they can be expressed as catenated versions of generalized RS codes [105].

D. Other Constructions

Feng and Rao [22], [24] introduced two classes of codes, in some sense a generalization of previous classes, but in terms of methods used to establish their properties, quite novel. A brief introduction to their definition is given here. The two classes are given by

$$\text{Type I: } x^a + y^b + G(x, y) = 0$$

where $\deg(G) < b < a$, $\gcd(a, b) = 1$, and

$$\text{Type II: } x^a y^c + y^{b+c} + G(x, y) = 0$$

where $\deg(G) < \min(a + c, b + c)$, $\gcd(a, b) = 1$. This last type of curve is a generalization of the Klein curves with equation: $x^m y + y^m z + z^m x = 0$. The curves of Type I are always irreducible and have exactly one point at infinity which is regular iff $a = b+1$. The properties of the evaluation codes derived from these curves depends very much on the form of the polynomials $G(x, y)$ and the method used for the code construction. Feng and Rao use the notion of a well-behaving sequence of monomials for the code construction. Høholdt *et al.* [44] analyze the properties of the codes using the notion of an order function. This approach is briefly described here. Let $\prec$ be an admissible order function on monomials and $f_i \prec f_{i+1}$. Every polynomial $f \in R$, R an $\mathbb{F}$-algebra, can be written in a unique manner as

$$f = \sum_{i=1}^{j} \alpha_i f_i, \qquad \alpha_i \in \mathbb{F},\ \alpha_j \neq 0.$$

Define the mapping $\rho(.)$ from $\mathbb{F}[x_1, \cdots, x_m]$ to $N_0 \cup \{-\infty\}$ by $\rho(0) = -\infty$ and $\rho(f) = j - 1$ if j is the smallest positive integer for which f can be written as a linear combination of the first j monomials. The function $\rho(.)$ satisfies the following conditions:

i) $\rho(f) = -\infty$ iff $f = 0$;
ii) $\rho(\lambda f) = \rho(f)$ for all nonzero $\lambda \in \mathbb{F}$;
iii) $\rho(f + g) \leq \max\{\rho(f), \rho(g)\}$ with equality when $\rho(f) < \rho(g)$;
iv) if $\rho(f) < \rho(g)$ and $h \neq 0$ then $\rho(fh) < \rho(gh)$;
v) if $\rho(f) = \rho(g)$, then there exists $\lambda \in \mathbb{F}^*$ such that $\rho(f - \lambda g) < \rho(g)$

and any function satisfying these conditions is called an *order function* on R. A weight function is an order function satisfying the additional condition that $\rho(fg) = \rho(f) + \rho(g)$. It is not difficult to show that if there exists an order function on R then R is an integral domain.

If I is an ideal generated by the Type I polynomial $x^a + y^b + G(x, y)$, and $R = \mathbb{F}[x, y]/I$ then there exists a weight function ρ on R, R is an integral domain, and I is a prime ideal. Furthermore the set

$$\{x^\alpha y^\beta | 0 \leq \alpha < \beta\}$$

is a basis for R and $\rho(x) = a$ and $\rho(y) = b$.

For I an ideal generated by a Type II equation

$$x^a y^c + y^{b+c} + G(x, y)$$

then $\mathbb{F}[x, y]/I$ has a basis

$$\{x^\alpha y^\beta | 0 \leq \alpha < a, \text{ or } \beta < c\}.$$

In this case, an order function may not exist but when it does it must have the property that $\rho(x^a) = \rho(y^b)$.

Using the notion of an order function, codes of Types I and II, as evaluation codes, can be defined and bounds on their minimum distance established. Readers are referred to [44] for details on this approach.

V. Decoding of Algebraic-Geometry Codes

The decoding problem can be formulated as follows:

Definition 5.1: Let C be an (n, k)-code over the field $\mathbb{F}_q$ and $0 \leq \tau < n$. The function dec_τ: $\mathbb{F}_q^n \to \mathcal{P}(C)$, $\mathcal{P}(C)$ the power set on C, where for any $\boldsymbol{y} \in \mathbb{F}_q^n$

$$\text{dec}_\tau(\boldsymbol{y}) = S(\boldsymbol{y}, \tau) \cap C$$

is called a *decoder* with capability of correcting τ errors. Here $S(\boldsymbol{y}, \tau)$ is the sphere with radius τ centered at $\boldsymbol{y}$.

With this definition the following lemma is immediate.

Lemma 5.2: Let C be an (n, k)-code over the field $\mathbb{F}_q$ with minimum distance d and let $= \lfloor (d-1)/2 \rfloor$. Then for any $\boldsymbol{y} \in \mathbb{F}_q^n$, $|\text{dec}_t(\boldsymbol{y})| \leq 1$.

Proof: Suppose $\boldsymbol{c}_1, \boldsymbol{c}_2 \in S(\boldsymbol{y}, t)$. Then

$$d(\boldsymbol{c}_1, \boldsymbol{c}_2) \leq d(\boldsymbol{c}, \boldsymbol{y}) + d(\boldsymbol{c}_2, \boldsymbol{y}) \leq t + t \leq d - 1$$

so $\boldsymbol{c}_1 = \boldsymbol{c}_2$. □

Most of the work in the constructions of decoders for algebraic-geometry codes has been focused on designing dec_t decoders with $t = \lfloor (d^* - 1)/2 \rfloor$ where d^* is some lower bound, e.g., the ones presented in Section III-E, on the minimum distance of the code, but recently Shokrollahi and Wassermann [80] have constructed dec_τ-decoders for larger τ by extending previous work of Sudan [88], who constructed such decoders for Reed–Solomon codes. For a survey on the decoding of algebraic-geometry codes and the history see Høholdt and Pellikaan [45].

In this section we will describe a decoding algorithm for algebraic-geometry codes of the form

$$C(m) = C(D, mG)^\perp = C^*(D, mG)$$

where $D = P_1 + \cdots + P_n$; $P_1, \cdots, P_n, G$ distinct $\mathbb{F}_q$-rational points on a nonsingular absolutely irreducible curve χ of genus g defined over $\mathbb{F}_q$.

The decoder corrects up to $\lfloor (d_{\text{FR}} - 1)/2 \rfloor$ errors where d_{FR} is the Feng–Rao distance to be defined later. One has $d_{\text{FR}} \geq m - 2g + 2$ with equality if $m \geq 4g - 2$. The code $C(m)$ has length n and for any $\boldsymbol{y} \in \mathbb{F}_q^n$ we have

$$\boldsymbol{y} \in C(m) \Leftrightarrow \sum_{j=1}^{n} y_j f(P_j) = 0, \qquad \text{for all } f \in L(mQ).$$

When $2g - 2 < m < n$ the code $C(m)$ has dimension $k = n - (m - g + 1)$, since $L(mG)$ has dimension $m - g + 1$ from the Riemann–Roch theorem because $m > 2g - 2$ and the

evaluation map $L(mG) \to \mathbb{F}_q^n$ which maps $f \in L(mG)$ into $(f(P_1), \cdots, f(P_n))$ is injective since $m < n$.

Recall that a number ρ_i is a nongap for G if $L(\rho_i G) \neq L((\rho_i - 1)G)$. In this case there exists a function

$$\varphi_i \in L(\rho_i G) \setminus L((\rho_{i-1})G)$$

which means that φ_i has a pole of order ρ_i at G and no other poles.

The nongaps satisfy

$$0 = \rho_1 < \rho_2 < \cdots < \rho_g < \rho_{g+1} = 2g$$

and $\rho_i = i + g - 1$ for $i \geq g + 1$, and the functions φ_i, $i = 1, 2, \cdots, m - g + 1$ provide a basis for $L(mG)$.

Let R denote the ring of all rational functions on X with poles only at G, that is,

$$R = \bigcup_{a=0}^{\infty} L(aG)$$

and let for $f \in R$, $\rho(f)$ denote the pole order of f at G, that is the smallest number b, such that $f \in L(bQ)$. If $f \in R$ and $\boldsymbol{y} \in \mathbb{F}_q^n$ we define the *syndrome* $S_{\boldsymbol{y}}(f)$ to be

$$S_{\boldsymbol{y}}(f) = \sum_{j=1}^{n} y_j f(P_j)$$

so we have

$$\boldsymbol{y} \in C(m) \Leftrightarrow S_{\boldsymbol{y}}(f) = 0, \qquad \text{for all } f \in R \text{ with } \rho(f) \leq m.$$

In the decoding situation we receive a vector $\boldsymbol{y}$ which is the sum of a codeword $\boldsymbol{c}$ and an error vector $\boldsymbol{e}$. We have $S_{\boldsymbol{e}}(f) = S_{\boldsymbol{y}}(f)$ if $\rho(f) \leq m$, so the syndromes $S_{\boldsymbol{e}}(f)$ can be calculated directly from the received word if $\rho(f) \leq m$. The standard decoding procedure for Reed–Solomon codes has the following five steps.

1) Syndrome calculation.
2) Obtaining a polynomial, called the error-locator polynomial, which has the error positions among its roots.
3) Obtaining error positions.
4) Calculating error magnitudes.
5) Recovering the codeword and the information symbols.

We will not discuss Steps 1) and 5) in detail since they are fairly easy once we have a basis for $L(mG)$, but we will demonstrate how Steps 2)–4) are generalized to the codes $C(m)$.

Let the error positions be $P_{i_1}, P_{i_2}, \cdots, P_{i_\tau}$. An element $h \in R$ satisfying $h(P_{i_1}) = h(P_{i_2}) = \cdots = h(P_{i_\tau}) = 0$ is called an *error locator*. We remark that it follows from the theorem of Riemann that the space $L((\tau + g)G)$ indeed contains a nonzero error locator since

$$\deg((\tau + g)Q - (P_{i_1} + \cdots + P_{i_\tau})) = g$$

so

$$\text{dimension } L((\tau + g)G - (P_{i_1} + \cdots + P_{i_\tau})) \geq g - g + 1 = 1.$$

If h is an error locator we have $S_{\boldsymbol{e}}(fh) = 0$ for all $f \in R$ since

$$S_{\boldsymbol{e}}(fh) = \sum_{i \in I = \{i_1, \cdots, i_\tau\}} e_i f(P_i) h(P_i) = 0.$$

On the other hand, we can prove the following.

Theorem 5.3: Let $h \in L((\tau + g)G)$ satisfy $S_{\boldsymbol{e}}(fh) = 0$ for all $f \in R$ with $\rho(f) \leq \tau + 2g - 1$ then h is an error locator.

Proof: The condition implies that the vector $\boldsymbol{u}$ with coordinates $e_i h(P_i)$, $i = 1, \cdots, n$ is a codeword in $C(\tau + 2g - 1)$. But $\tau + 2g - 1 > 2g - 2$ so this code has minimum weight at least $\tau + 2g - 1 - 2g + 2 = \tau + 1$ which is greater than τ and, therefore, $\boldsymbol{u} = \underline{0}$ and hence $h(P_{i_j}) = 0$, $j = 1, \cdots, \tau$. □

By combining the remark and the theorem one gets the idea of obtaining the error locator in the following way. Consider the following system of equations in the unknowns $\lambda_1, \lambda_2, \cdots, \lambda_{\tau+1}$:

$$\sum_{j=1}^{n} e_j \varphi_l(P_j) \sum_{i=1}^{\tau+1} \lambda_i \varphi_i(P_j) = 0, \qquad l = 1, 2, \cdots, \tau + g$$

or, equivalently,

$$\sum_{i=1}^{\tau+1} S_{\boldsymbol{e}}(\varphi_i \varphi_l) \lambda_i = 0, \qquad l = 1, 2, \cdots, \tau + g.$$

It then follows from the discussion above that this system indeed has a nontrivial solution and that $h = \sum_{i=1}^{\tau+1} \lambda_i \varphi_i$ is an error locator.

The problem is that we only know $S_{\boldsymbol{e}}(\varphi_i \varphi_j)$ if $\rho(\varphi_i \varphi_j) \leq m$, so in order to solve it we must have

$$(\tau + 1 + g - 1) + (\tau + g + g - 1) \leq m$$

that is, $2\tau + 3g - 1 \leq m$, with $d^* = m - 2g + 2$ we get $\tau \leq (d^* - g - 1)/2$.

i	1	2	3	4	5	6	7	8	9	10	11	12	13
φ_i	1	X	Y	X^2	XY	Y^2	X^3	X^2Y	XY^2	Y^3	X^4	X^3Y	X^2Y^2
ρ_i	0	4	5	8	9	10	12	13	14	15	16	17	18
$S(\varphi_i)$	α^9	α^{14}	0	α^5	α^9	α^9	α^7	α^{14}	α^{11}	α^6	α^2	α^{12}	0
i	14	15	16	17	18	19	20	21	22	23	24	25	26
φ_i	XY^3	Y^4	X^4Y	X^3Y^2	X^2Y^3	XY^4	Y^5	X^4Y^2	X^3Y^3	X^2Y^4	XY^5	Y^6	X^4Y^3
ρ_i	19	20	21	22	23	24	25	26	27	28	29	30	31
$S(\varphi_i)$	α^4	α^5	α^5	α^{12}	α^7	α^7	α^6	α^6	α^3	α^6	α^4	α^{11}	α^{10}

Example 5.4: The Hermitian curve over $\mathbb{F}_{16}$ has equation $x^5+y^4+y=0$. It has genus $g=6$, 64 $\mathbb{F}_q$-rational points in the affine part, and one point Q at infinity. As in Example 3.34 we have $\rho(X)=4$, $\rho(Y)=5$, and the functions X^aY^b, $0\le a<5$, $0\le b$, $4a+5b\le m$ gives a basis for $L(mQ)$. Let $m=31$ then we get a $(64, 38, 21)$ code over $\mathbb{F}_{16}$, so the algorithm described above corrects $(21-1-6)/2=7$ errors. Let α be a primitive element of $\mathbb{F}_{16}$ satisfying $\alpha^4+\alpha+1=0$. We consider a seven-error pattern where the errors are located at the points $P_1=(1,\alpha)$, $P_2=(\alpha^8,\alpha^3)$, $P_3=(\alpha,\alpha^7)$, $P_4=(\alpha^2,\alpha^3)$, $P_5=(\alpha^{11},\alpha^3)$, $P_6=(\alpha^5,\alpha^3)$, $P_7=(\alpha^{14},\alpha^3)$, and the corresponding error values $e_1=\alpha^6$, $e_2=\alpha^8$, $e_3=\alpha^7$, $e_4=\alpha$, $e_5=1$, $e_6=\alpha^6$, $e_7=\alpha^{10}$. We get the table at the bottom of the preceding page. We want a locator of the form

$$\lambda_1+\lambda_2x+\lambda_3y+\lambda_4x^2+\lambda_5xy+\lambda_6y^2+\lambda_7x^3+\lambda_8x^2y$$

where the coefficients λ_i satisfy the equation

$$\begin{bmatrix}
\alpha^9 & \alpha^{14} & 0 & \alpha^5 & \alpha^9 & \alpha^9 & \alpha^7 & \alpha^{14}\\
\alpha^{14} & \alpha^5 & \alpha^9 & \alpha^7 & \alpha^{14} & \alpha^{11} & \alpha^2 & \alpha^{12}\\
0 & \alpha^9 & \alpha^9 & \alpha^{14} & \alpha^{11} & \alpha^6 & \alpha^{12} & 0\\
\alpha^5 & \alpha^7 & \alpha^{14} & \alpha^2 & \alpha^{12} & 0 & \alpha^5 & \alpha^5\\
\alpha^9 & \alpha^{14} & \alpha^{11} & \alpha^{12} & 0 & \alpha^4 & \alpha^5 & \alpha^{12}\\
\alpha^9 & \alpha^{11} & \alpha^6 & 0 & \alpha^4 & \alpha^5 & \alpha^{12} & \alpha^7\\
\alpha^7 & \alpha^2 & \alpha^{12} & \alpha^5 & \alpha^5 & \alpha^{12} & 1 & \alpha^5\\
\alpha^{14} & \alpha^{12} & 0 & \alpha^5 & \alpha^{12} & \alpha^7 & \alpha^5 & \alpha^6\\
\alpha^{11} & 0 & \alpha^4 & \alpha^{12} & \alpha^7 & \alpha^7 & \alpha^6 & \alpha^3\\
\alpha^6 & \alpha^4 & \alpha^5 & \alpha^7 & \alpha^7 & \alpha^6 & \alpha^3 & \alpha^6\\
\alpha^2 & \alpha^5 & \alpha^5 & 1 & \alpha^5 & \alpha^6 & \alpha^8 & \alpha^{13}\\
\alpha^{12} & \alpha^5 & \alpha^{12} & \alpha^5 & \alpha^6 & \alpha^3 & \alpha^{13} & \alpha\\
0 & \alpha^{12} & \alpha^7 & \alpha^6 & \alpha^3 & \alpha^6 & \alpha & \alpha^{10}
\end{bmatrix}
\begin{bmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\\ \lambda_4\\ \lambda_5\\ \lambda_6\\ \lambda_7\\ \lambda_8\end{bmatrix}=\underline{0}.$$

Here we have used that $S(X^5)=S(Y^4+Y)=\alpha^5+0=\alpha^5$ and the corresponding expressions for $S(X^6)$, $S(X^5Y)$, $S(X^7)$, $S(X^6Y)$, and $S(X^5Y^2)$.

It can be seen that

$$(\lambda_1,\lambda_2,\cdots,\lambda_8)=(\alpha^{11},\alpha^{13},\alpha^{13},0,\alpha^{10},1,0,0)$$

is a solution so

$$h=\alpha^{11}+\alpha^{13}X+\alpha^{13}Y+\alpha^{10}XY+\alpha^2X^2.$$

The zeros of this polynomial are $P_1,\cdots,P_7$ and $(\alpha^{11},\alpha^{14})$; $(\alpha^{14},\alpha^{12})$; (α^4,α^6).

Let $d^*=m-2g+2$, then we get

$$2\tau+3g-1\le d^*+2g-2$$

so

$$\tau\le(d^*-1-g)/2.$$

This was the original approach of Justesen, Larsen, Havemose, Elbrønd Jensen, and Høholdt [47], and of Skorobogatov and Vlăduţ [82].

The calculation above makes it natural to look at the matrix of syndromes

$$\boldsymbol{S_e}=(S_{ij}(\boldsymbol{e}))=(S_{\boldsymbol{e}}(\varphi_i\varphi_j)),\qquad 1\le i,\ j\le N$$

where n is the smallest number such that $C(N)=\underline{0}$. We will first prove

Lemma 5.5: $\operatorname{rank}(\boldsymbol{S_e})=\tau$.

Proof: Decompose $\boldsymbol{S}$ as a product of the three matrices $\boldsymbol{A}$ with elements $a_{ij}=\varphi_i(P_j)$, $i=1,\cdots,N$, $j=1,\cdots,n$, $\boldsymbol{B}$ a diagonal $n\times k$ matrix with $e_1,\cdots,e_n$ in the diagonal, and $\boldsymbol{A}^T$.

Then we have $\boldsymbol{S_e}=\boldsymbol{ABA}^T$ and

$$\operatorname{rank}(\boldsymbol{A})=n=\operatorname{rank}(\boldsymbol{A}^T)$$

so

$$\operatorname{rank}(\boldsymbol{S})=\operatorname{rank}(\boldsymbol{B})=\operatorname{weight}(\boldsymbol{e})=\tau.\qquad\square$$

Definition 5.6: For $l\in\mathbb{N}_0$ let

$$N_l=\{(i,j)\in\mathbb{N}_0^2|\rho_i+\rho_j=l+1\}$$

and let ν_l be the number of elements in N_l.

From the definition of the codes $C(m)$ we have that if $\boldsymbol{c}\in C(m)$ and $\rho_i+\rho_j\le m$ then $S_{\boldsymbol{c}}(\varphi_i\varphi_j)=0$ but if $\boldsymbol{c}\in C(m)\backslash C(m+1)$ and $\rho_i+\rho_j=m+1$ then $S_{\boldsymbol{c}}(\varphi_i\varphi_j)\ne 0$ but this implies that

Lemma 5.7: If $\boldsymbol{c}\in C(m)\backslash C(m+1)$ then $\operatorname{weight}(\boldsymbol{c})\ge\nu_m$.

Proof: We can repeat the decomposition of the syndrome matrix $\boldsymbol{S_c}$ so this has $\operatorname{rank}=\operatorname{weight}(\boldsymbol{c})$, but the nonzero elements appears in different rows and columns with zeros above, so this rank is at least ν_m. $\square$

Definition 5.8: For the code $C(m)$ we define

$$d_{\mathrm{FR}}=\min_{l\ge m}\{\nu_l\}.$$

Theorem 5.9: The minimum distance d of $C(m)$ satisfies

$$d\ge d_{\mathrm{FR}}.$$

Proof: This follows directly from the lemma. $\square$

Theorem 5.10: If $m\ge 4g-2$ then $d_{\mathrm{FR}}=m-2g+2$.

Proof: If $m\ge 4g-2$, $l\ge m$, and $\rho_i+\rho_j=l+1$, we see that if $i\ge g+1$ and $j\ge g+1$ we get $l-4g+2$ solutions and if $i\le g$ or $j\le g$ we get $2g$ solutions so $\nu_l=l-2g+2$ from which the result follows. $\square$

We will now describe a procedure that, based on the known syndromes $S_{\boldsymbol{e}}(\varphi_i\varphi_j)$, $\rho_i+\rho_j\le m$, determines the syndromes $S_{\boldsymbol{e}}(\varphi_i\varphi_j)$, $\rho_i+\rho_j\le 2\tau+3g-1$ when $\tau\le\lfloor(d_{\mathrm{FR}}-1)/2\rfloor$. Combined with Theorem 5.3 this then gives a method to find an error locator. This is the brilliant idea of Feng and Rao [23], that was made precise by Duursma [17].

We first note that in the syndrome matrix the first unknown entries correspond to the indices $(i,j)\in N_m$ but as soon as we know one s_{ij} with $(i,j)\in N_m$ we know all $s_{i'j'}$ with $(i',j')\in N_m$ since $\rho(\varphi_i\varphi_j)=\rho(\varphi_{i'}\varphi_{j'})$ so

$$\varphi_i\varphi_j=\lambda\varphi_{i'}\varphi_{j'}+g\qquad(2)$$

where $\lambda\in\mathbb{F}_q^*$ and $\rho(f)\le m$ and this relation is independent of the error vector.

Consider the matrix

$$\boldsymbol{S}(i, j) = \{s_{i'j'} | 1 \le i' \le i,\ 1 \le j' \le j\}.$$

If $\rho_i + \rho_j = m + 1$, then all entries of this matrix except s_{ij} are known

$$\begin{pmatrix} s_{1,1} & \cdots & s_{1,j-1} & s_{1,j} \\ \vdots & & \vdots & \vdots \\ s_{i-1,1} & \cdots & s_{i-1,j-1} & s_{i-1,j} \\ s_{i,1} & \cdots & s_{i,j-1} & ? \end{pmatrix}.$$

Definition 5.11: If $(i, j) \in N_m$, that is, $\rho_i + \rho_j = m+1$ and the three matrices $\boldsymbol{S}(i-1, j-1)$, $\boldsymbol{S}(i-1, j)$, and $\boldsymbol{S}(i, j-1)$ have equal rank, then (i, j) is called a *candidate* with respect to the code $C(m)$. If (i, j) is a candidate, then there is a unique value s'_{ij} to assign to the unknown entry s_{ij} such that the matrices $\boldsymbol{S}(i, j)$ and $\boldsymbol{S}(i-1, j-1)$ have equal rank. The element s'_{ij} is called the *candidate value* of the unknown syndrome s_{ij}. If $s'_{ij} = s_{ij}$ the candidate is called *true* and *false* otherwise.

Using the relation (2) every $(i', j') \in N_l$ gives a candidate value for s_{ij}. Denote the number of true candidates T and the number of false candidates F. An entry (i, j) is called a *discrepancy* if it is a candidate but the matrices $\boldsymbol{S}(i, j)$ and $\boldsymbol{S}(i-1, j-1)$ have different ranks.

Suppose now $\text{weight}(\boldsymbol{e}) \le \lfloor (d_{\text{FR}} - 1)/2 \rfloor$. Denote the number of discrepancies in the known part of the matrix by K. A candidate is incorrect if and only if it is a discrepancy, so

$$K + F \le \text{total number of discrepancies} \le \text{weight}(\boldsymbol{e}).$$

If entry (r, s) is a known discrepancy, then all entries (r, s') is the rth row with $s' > s$ and all entries (r', s) the sth column with $r' > r$ are noncandidates. If $(i, j) \in N_m$ is not a candidate, then there is at least one known discrepancy in the same row or column. Thus the number of pairs $(i, j) \in N_m$ which are noncandidates is at most $2K$. The number of candidates is $T + F$, so

$$\nu_m = \#\text{candidates} + \#\text{noncandidates} \le T + F + 2K.$$

Since $\text{weight}(\boldsymbol{e}) \le \lfloor (d_{\text{FR}} - 1)/2 \rfloor \le (\nu_m - 1)/2$, we therefore get $K + F \le (\nu_m - 1)/2 \le (T + F + 2K - 1)/2$ and hence

$$F < T.$$

But this means that all true candidates give the same correct value for s_{ij}, so we have proven the following theorem.

Theorem 5.12: If the number of errors in a received word with respect to the code $C(m)$ is at most $(d_{\text{FR}} - 1)/2$, then the majority of the candidates vote for the correct value of $S(\varphi_i \varphi_j)$, $\rho_i + \rho_j = m + 1$.

Hence we can use this theorem until we have all the syndromes and then find the error locator. This completes Step 2). In this way we get an error locator h in $L((\tau + g)Q)$, so h has at most $\tau + g$ zeros. If $m \ge 4g - 2$ we have $\tau + g \le d_{\text{FR}} - 1$. If $m < 4g - 2$, it is possible to extend the algorithm to get more error locators, with the property that the number of common zeros are $\le d_{\text{FR}} - 1$. We can therefore find a set of at most $\tau + g$ points that contains the error positions by evaluating the function h at all the points $P_1, P_2, \cdots, P_n$, so this almost completes Step 3).

Example 5.13: We consider the Hermitian curve of Example 5.4 with $m = 25$, so we get a $(64, 44, 15)$ code over $\mathbb{F}_{16}$. We will correct the same seven-error pattern as in Example 5.4. The syndrome matrix is (in powers of α, with $*$ indicating a zero) as shown at the bottom of this page. Here $\times$ denotes the

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

syndrome corresponding to x^4y^2 with $\rho(x^4y^2) = 26$, which is the first unknown.

Using row operations one gets the matrix at the bottom of this page. From this it is seen that the positions $(6, 11)$, $(8, 8)$, and $(11, 6)$ are candidate positions and by keeping track of the performed row operations one sees also that the candidate values in all three cases are α^6, so this is the correct syndrome $S(f_{21})$. In the same manner one finds $S(f_{22}) = \alpha^3$, $S(f_{23}) = \alpha^6$, $S(f_{24}) = \alpha^4$, $S(f_{25}) = \alpha^{11}$, $S(f_{26}) = \alpha^{10}$, which corresponds to the results of Example 5.4 and we can now apply the basic algorithm for the code C_{26} as before.

However, since we now know at most $d_{\mathrm{FR}} - 1$ positions $J \subseteq \{1, \cdots, n\}$ that include the error positions we can solve the system of linear equations given by

$$\boldsymbol{H}\boldsymbol{x}^T = \boldsymbol{H}\boldsymbol{y}^T,\ x_j = 0, \qquad \text{for all } j \notin J$$

where $\boldsymbol{H}$ is the parity-check matrix of the code. It is clear that the error vector $\boldsymbol{e}$ is a solution and another solution $\boldsymbol{x}$ would give $\boldsymbol{H}(\boldsymbol{x}-\boldsymbol{e})^T = \underline{0}$ so $\boldsymbol{x}-\boldsymbol{e}$ would be a codeword of weight at most $d_{\mathrm{FR}} - 1$ and hence must be zero. Therefore, $\boldsymbol{x} = \boldsymbol{e}$.

In this way we have found the error vector and Steps 3) and 4) are completed.

The way we have presented the solution to the decoding problem has the complexity of solving systems of linear equations, and is similar to the Peterson algorithm in using error-locator polynomials for decoding Reed–Solomon codes. It is possible to get lower complexity by using Sakata's generalization of the Berlekamp–Massey algorithm, as was done by Sakata, Elbrønd Jensen, and Høholdt in [77], O'Sullivan in [67], and Saints and Heegard in [71]. It is also possible to find the error values by using a generalization of the Forney formula, this was done by Leonard [56], Hansen, Jensen, and Kötter [40], and by O'Sullivan [68]. One could use the voting procedure described above to find all syndromes and then use a discrete Fourier-like transform to get the error values. This is the approach of Sakata, Jensen, and Høholdt [77].

VI. Asymptotics of Algebraic-Geometry Codes

One reason for the interest in algebraic-geometry codes is the fact that those codes can be used to give an asymptotically good sequence of codes with parameters better than the Varshamov–Gilbert bound in a certain range of the rate and for large enough alphabets. In this section we will review the construction.

Recall that a code C is called an (n, M, d)-code over $\mathbb{F}_q$ if C is a subset of $\mathbb{F}_q^n$ with minimum distance d and $|C| = M$.

Let

$$A_q(n, d) = \max\{M | \text{there exists an } (n, M, d) \text{ code over } \mathbb{F}_q\}$$

and

$$\alpha(\delta) = \limsup_{n\to\infty} \frac{\log_q A_q(n, \delta_n)}{n}, \qquad \text{for } 0 \le \delta \le 1. \tag{3}$$

It is not hard to see that

$$\alpha(\delta) = 0, \qquad \text{for } 1 - \frac{1}{q} \le \delta \le 1 \tag{4}$$

and the Varshamov–Gilbert bound is the fact that

$$\alpha(\delta) \ge 1 - H_q(\delta), \qquad \text{for } 0 \le \delta \le 1 - \frac{1}{q} \tag{5}$$

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

where $H_q(x)$ is the q-ary entropy function defined by

$$H_q(0) = 0$$
$$H_q(x) = x \log_q(q-1) - x \log_q x - (1-x)\log_q(1-x), \quad 0 < x \le 1 - \frac{1}{q}.$$

In [91], it is shown that by using algebraic-geometry codes it is possible to prove that

$$\alpha(\delta) + \delta \ge 1 - \frac{1}{\sqrt{q}-1} \tag{6}$$

if q is a square.

It turns out that (6) gives an improvement of (5) if $q \ge 49$. The inequality (6) is the Tsfasman–Vlăduţ–Zink bound. Fig. 1 shows the two bounds for $q = 256$.

Let $R = k/n$ be the rate and $\delta = d/n$ the relative minimum distance of an algebraic-geometry code as defined in Section III. It then follows from the results in that section that

$$R + \delta \ge 1 - \frac{g-1}{n} \tag{7}$$

where g is the genus of the curve involved in the construction.

In order to construct a sequence of good codes we therefore need curves with low genera and many $\mathbb{F}_q$-rational points. For a curve over $\mathbb{F}_q$ of genus g with N $\mathbb{F}_q$-rational points we get from the Hasse–Weil bound (3.38) that

$$N \le q + 1 + 2g\sqrt{q}. \tag{8}$$

Let

$$A(q) = \limsup_{g \to \infty} \frac{N(g)}{g}$$

where $N(g)$ is the maximal number of $\mathbb{F}_q$-rational points on a curve of genus g over $\mathbb{F}_q$. The Hasse–Weil bound implies that

$$A(q) \le 2\sqrt{q}. \tag{9}$$

In 1983, Vlăduţ and Drinfeld [16] improved on (9) by showing that

$$A(q) \le \sqrt{q} - 1. \tag{10}$$

When q is a square, Ihara in [46] and Tsfasman, Vlăduţ, and Zink [92] showed that

$$A(q) = \sqrt{q} - 1 \tag{11}$$

by studying the so-called *modular curves* over finite fields.

This in turn means that there exists a sequence of codes satisfying

$$R + \delta \ge 1 - \frac{1}{\sqrt{q}-1} \text{ when } q \text{ is a square} \tag{12}$$

and, therefore, (6) follows.

The construction using modular curves is difficult. It is possible to do this with polynomial complexity but the actual construction of generator or parity-check matrices is intractable, so many researchers have tried to find a more simple construction. In [22], Feng and Rao suggested that one could get asymptotically good codes by using the so-called generalized Klein curves, which are defined by the equations

$$x_{i+1}^3 x_i + x_i^3 + x_{i+1} = 0, \qquad i = 1, 2, \cdots, m-1$$

over $\mathbb{F}_8$.

Pellikaan tried to determine whether this claim was correct (the curves are asymptotically bad as recently proved by Garcia and Stichtenoth), and suggested using the curves with equations

$$x_{i+1}^2 x_i + x_i^2 + x_{i+1} = 0, \qquad i = 1, 2, \cdots, m-1$$

over $\mathbb{F}_4$. This led Garcia and Stichtenoth in [28] to study the affine variety χ_m over $\mathbb{F}_q$, $q = r^2$ given by the equations

$$x_i^{r-1} x_{i+1}^r + x_{i+1} = x_i^r, \qquad i = 1, 2, \cdots, m-1 \tag{13}$$

and they showed that χ_m is indeed a curve and

$$\lim_{m \to \infty} \frac{N(\chi_m)}{g(\chi_m)} = r - 1 \tag{14}$$

so in this way one can obtain an asymptotically good sequence of codes meeting the Tsfasman–Vlăduţ–Zink bound. Notice that the equations are of the following type:

$$F(x_i, x_{i+1}) = 0, \qquad \text{for } i = 1, \cdots, m-1$$

where

$$F(x, y) = x^{r-1} y^r + y - x^r.$$

The affine plane curve with equation $F(x, y) = 0$ has the property that for every nonzero element $x \in \mathbb{F}_q$ there are exactly r nonzero solutions in $\mathbb{F}_q$ of the equation $F(x, y) = 0$. This is seen by multiplying the equation with x and replacing xy with z. Then we get the equation $z^r + z = x^{r+1}$, which is an equation of the Hermitian curve over $\mathbb{F}_q$. For every given x in $\mathbb{F}_q$ the element x^{r+1} is in $\mathbb{F}_r$ and since the left side is the trace map from $\mathbb{F}_q$ to $\mathbb{F}_r$ we get r distinct z's such that $z^r + z = x^{r+1}$. If, furthermore, x is not zero, then $y = z/x$ is defined and is also nonzero. Therefore, the curve χ_2 has $(q-1)r$ points with nonzero coordinates in $\mathbb{F}_q$. Consider the map

$$\pi_m \colon \chi_m \to \chi_{m-1}$$

defined as

$$\pi_m(x_1, \cdots, x_{m-1}, x_m) = (x_1, \cdots, x_{m-1}).$$

If $(x_1, \cdots, x_{m-1})$ is a given $\mathbb{F}_q$-rational point of χ_{m-1} and $x_{m-1} \neq 0$, then there are exactly r possible nonzero values for $x_m \in \mathbb{F}_q$ such that $(x_1, \cdots, x_{m-1}, x_m)$ is a point of χ_m. Therefore, by induction, it is shown that

$$N(\chi_m) \ge (q-1)r^{m-1}.$$

The genus of the curve χ_m is more difficult to calculate. It is done by induction using the Hurwitz–Zeuthen formula [83]

to the covering π_m: $\chi_m \to \chi_{m-1}$, which in this case is an Artin–Schreier covering. The result [30] is

$$g(\chi_m) = \begin{cases} r^m + r^{m-1} - r^{(m+1)/2} \\ \quad -2r^{(m-1)/2} + 1, & \text{if } m \equiv 1 \bmod 2 \\ r^m + r^{m-1} - \frac{1}{2}r^{m/2+1} \\ \quad -\frac{3}{2}r^{m/2} - r^{m/2-1} + 1, & \text{if } m \equiv 1 \bmod 2 \end{cases} \tag{15}$$

from which (14) follows.

In order to make the codes really constructive one needs to find the right divisor G and the bases for the vector spaces $L(G)$. This seems to be very difficult. For the codes coming from χ_3 it has been done by Voß and Høholdt in [101].

Garcia and Stichtenoth [28] also presented another asymptotically good sequence of curves. Here the defining equations are

$$x_{i+1}^q + x_{i+1} = \frac{x_i^q}{x_i^{q-1} + 1}, \qquad i = 1, \cdots, m-1 \tag{16}$$

over $\mathbb{F}_{q^2}$.

Here one also has

$$\lim_{n\to\infty} \frac{N(\chi_m)}{g(\chi_m)} = q - 1 \tag{17}$$

and, moreover, quite recently Pellikaan, Stichtenoth, and Torres in [69] succeed in calculating recursively the nongap sequence of Q, the point at infinity.

Let S_m denote the semigroup of nongaps at Q in χ_m. For $m \geq 1$ let

$$A_m = \begin{cases} q^m - q^{m/2}, & \text{if } m \text{ is even} \\ q^m - q^{(m+1)/2}, & \text{if } m \text{ is odd.} \end{cases} \tag{18}$$

Then $S_1 = \mathbb{N}_0$ and for $m \geq 1$

$$S_{m+1} = qS_m \cup \{x \in \mathbb{N}_0 | x \geq A_{m+1}\}. \tag{19}$$

One could hope that this will lead to a determination of the basis of $L(rQ)$ for χ_m.

The two sequences of curves given by (13) and (16) have recently been shown to be specific examples of modular curves by Elkies [20].

VII. Conclusions

The past fifteen years has seen extraordinary developments in the application of the ideas of algebraic geometry to the construction of codes and their decoding algorithms. While the highlight of these developments has been the construction of asymptotically good codes, very significant advances have been achieved in other directions. For all of the developments achieved to date, it is clear that many interesting challenges remain and other avenues yet to explore. These might include the following problems.

It is perhaps true that codes from geometries have yet to have an impact in practice. The development of classes of codes with the simplicity and efficiency of both encoding and decoding algorithms that rival those of Reed–Solomon codes, for example, will be required to break through current practice. While the development of asymptotically good codes over fields of size at least 49 is an impressive achievement, the development of asymptotically good binary codes, using ideas from algebraic geometry, remains an elusive and challenging goal.

The combinatorial structure of linear codes has been an interesting chapter in coding theory. Designs with excellent parameters often result from codes with exceptional structure, such as the Golay and quadratic residue codes and extremal self-dual codes [10]. With the superior properties of code classes developed from algebraic geometry, one might expect an investigation of their combinatorial properties would show promise. The investigation of such structure that utilizes the properties of the curves from which the codes are obtained, might prove interesting.

The intimate connections between codes and lattices, and more generally, sphere packings in Euclidean spaces, is now well established and a very active area of research [7]. The lattices and sphere packings derivable from codes in algebraic geometries where the resulting properties can be related to the properties of the curves used, might prove interesting. For example, lattices resulting from certain elliptic curves [18], [19], yield the best known packing densities for their dimensions. Perhaps further investigations in these directions will yield results of interest.

There is little doubt that future investigations of the ideas of algebraic geometry applied to these, and other, areas will reveal new and exciting results and directions. One cannot help but feel that the mathematical elegance of the ideas of algebraic geometry has yet to be fully exploited. It is hoped that this brief review has provided a look at where the subject stands today, as a platform for further work.

Acknowledgment

The authors would like to express their appreciation to Prof. R. E. Blahut and Prof. H. Stichtenoth, and to the anonymous reviewers, for their many helpful comments on the original draft of this paper. Their efforts are greatly appreciated.

References

[1] S. S. Abhyankar, *Algebraic Geometry for Scientists and Engineers*. Providence, RI: Amer. Math. Soc., 1990.
[2] A. M. Barg, S. L. Katsman, and M. A. Tsfasman, "Algebraic geometric codes from curves of small genus," *Probl. Inform. Transm.*, vol. 23, pp. 34–38, 1987.
[3] E. R. Berlekamp, "On decoding binary Bose–Chaudhuri–Hocquenghem codes," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 577–580, Oct. 1965.
[4] R. C. Bose and D. K. Ray-Chaudhuri, "On a class of error correcting binary group codes," *Inform. Contr.*, vol. 3, pp. 68–79, Mar. 1960.
[5] B. Buchberger, "Multidimensional systems theory: Progress, directions and open problems in multidimensional systems," in *An Algorithmic Method in Polynomial Ideal Theory*, N. K. Bose, Ed. Dordrecht, The Netherlands: Reidel, 1985, ch. "Gröbner Bases."
[6] H. Chen and S.-T. Yau, "Contribution to Munuera's problem on the main conjecture of geometric hyperelliptic MDS codes," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1349–1354, July 1997.
[7] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. Berlin, Germany: Springer-Verlag, 1993.
[8] D. Cox, J. Little, and D. O'Shea, *Ideals, Varieties, and Algorithms*. New York: Springer-Verlag, 1992.
[9] M. A. de Boer, "Almost MDS codes," *Desi., Codes Cryptogr.*, vol. 9, pp. 143–155, 1996.
[10] P. Delsarte, "An algebraic approach to the association schemes of coding theory," *Philips Res. Repts. Suppl.*, no. 10, 1973.
[11] ——, "On subfield subcodes of modified Reed–Solomon codes," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 575–576, 1975.

[12] Y. Driencourt and J. F. Michon, "Elliptic codes over fields of characteristic 2," *J. Pure and Appl. Alg.*, vol. 45, pp. 15–39, 1987.

[13] Y. Driencourt, "Some properties of elliptic codes over a field of characteristic 2," in *Lecture Notes in Computer Science*, vol. 229. Berlin: Springer-Verlag, 1985.

[14] Y. Driencourt and J. F. Michon, "Remarques sur les codes géométriques," *Compt. Rend.*, vol. 301, pp. 15–17, 1985.

[15] Y. Driencourt and H. Stichtenoth, "A criterion for self-duality of geometric codes," *Comm. in Alg.*, vol. 17, pp. 885–898, 1989.

[16] V. G. Drinfeld and S. G. Vlăduţ, "Number of points of an algebraic curve," *Func. Anal.*, vol. 17, pp. 53–54, 1993.

[17] I. M. Duursma, "Decoding codes from curves and cyclic codes," Ph.D. dissertation, Eindhoven Univ. Technol., Eindhoven, The Netherlands, Aug. 1993.

[18] N. D. Elkies, "Mordell–Weil lattices in characteristic 2 I: Construction and first properties," preprint.

[19] ———, "Mordell–Weil lattices in characteristic 2 II: The Leech lattice as a Mordell–Weil lattice," preprint.

[20] ———, "Beyond Coppa codes," in *Proc. 35th Allerton Conf. Communication, Control, and Computing* (Allerton House, Monticello, IL, 1997).

[21] A. Faldum and W. Willems, "Codes of small defect," *Des., Codes Cryptogr.*, vol. 10, pp. 341–350, 1997.

[22] G.-L. Feng and T. R. N. Rao, "Improved geometric Goppa codes, Part I: Basic theory," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1678–1693, Nov. 1995.

[23] ———, "Decoding algebraic–geometric codes up to the designed minimum distance," *IEEE Trans. Inform. Theory*, vol. 39, pp. 37–45, Jan. 1993.

[24] ———, "A simple approach for construction of algebraic–geometric codes from affine plane curves," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1003–1012, 1994.

[25] G.-L. Feng, V. K. Wei, T. R. N. Rao, and K. K. Tzeng, "Simplified understanding and efficient decoding of a class of algebraic–geometric codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 981–1002, 1994.

[26] P. Fitzpatrick, "A new derivation of an algorithm for solving the key equation," manuscript, 1993.

[27] W. Fulton, *Algebraic Curves.* New York: Benjamin, 1969.

[28] A. Garcia and H. Stichtenoth, "A tower of Artin–Schreier extensions of function fields attaining the Drinfeld–Vlăduţ bound," *Invent. Math.*, vol. 121, pp. 211–222, 1995.

[29] ———, "On the asymptotic behaviour of some towers of function fields over finite fields," *J. Number Theory*, vol. 61, pp. 248–273, 1996.

[30] ———, "Algebraic function fields over finite fields with many rational points," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1548–1563, 1995.

[31] M. J. E. Golay, "Notes on digital coding," *Proc. IRE*, vol. 37, p. 657, June 1949.

[32] V. D. Goppa, "A new class of linear error-correcting codes," *Probl. Inform. Transm.*, vol. 6, pp. 207–212, Sept. 1970.

[33] ———, "Codes on algebraic curves," *Sov. Math.–Dokl.*, vol. 24, pp. 170–172, 1981, translation from *Dokl. Akad. Nauk S.S.S.R*, vol. 259, pp. 1289–1290, 1981.

[34] ———, "Codes associated with divisors," *Probl. Inform. Transm.*, vol. 13, pp. 22–27, 1977.

[35] ———, "Algebraico–geometric codes," *Math. USSR Izv.*, vol. 21, pp. 75–91, 1983.

[36] ———, *Geometry and Codes.* Dordrecht, The Netherlands: Kluwer, 1988.

[37] ———, "Rational representation of codes and (L, g)-codes," *Probl. Inform. Transm.*, vol. 7, pp. 223–229, 1971.

[38] R. W. Hamming, "Error detecting and error correcting codes," *Bell Syst. Tech. J.*, vol. 29, pp. 147–160, Apr. 1950.

[39] J. P. Hansen, "Codes on the Klein quartic, ideals and decoding," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 923–925, 1987.

[40] J. P. Hansen, H. E. Jensen, and R. Kötter, "Determination of error values for AG-codes and the Forney formula," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1263–1269, July 1996.

[41] C. Heegard, J. H. Little, and K. Saints, "Systematic encoding via Gröbner bases for a class of algebraic geometric Goppa codes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1752–1761, Nov. 1995.

[42] J. W. P. Hirschfeld, M. A. Tsfasman, and S. G. Vlăduţ, "The weight hierarchy of higher dimensional Hermitian codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 275–278, 1994.

[43] A. Hocquenghem, "Codes correcteurs d'erreurs," *Chiffres*, vol. 2, pp. 147–156, 1959.

[44] T. Høholdt, J. H. van Lint, and R. Pellikaan, "Algebraic geometry codes," to be published.

[45] T. Høholdt and R. Pellikaan, "On the decoding of algebraic–geometric codes," *IEEE-Trans. Inform. Theory*, vol. 41, pp. 1589–1614, Nov. 1995.

[46] Y. Ihara, "Some remarks on the number of rational points of algebraic curves over finite fields," *J. Fac. Sci. Univ. Tokyo Japan*, vol. 28, pp. 721–724, 1981.

[47] J. Justesen, K. J. Larsen, H. E. Jensen, A. Havemose, and T. Høholdt, "Construction and decoding of a class of algebraic geometry codes," *IEEE Trans. Inform. Theory*, vol. 35, pp. 811–821, 1989.

[48] J. Justesen, "A class of constructive, asymptotically good algebraic codes," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 652–656, 1972.

[49] T. Kasami, S. Lin, and W. W. Petersen, "Some results on weight distributions of BCH codes," *IEEE Trans. Inform. Theory*, vol. IT-12, p. 274, Apr. 1966.

[50] G. L. Katsman and M. A. Tsfasman, "Spectra of algebraic geometry codes," *Probl. Inform. Transm.*, vol. 23, pp. 262–275, 1987.

[51] G. L. Katsman, M. A. Tsfasman, and S. G. Vlăduţ, "Modular curves and codes with a polynomial construction," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 353–355, 1984.

[52] C. Kirfel and R. Pelikaan, "The minimum distance of codes in an array coming from telescopic semigroups," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1720–1732, 1995.

[53] N. Koblitz, "A course in number theory and cryptography," in *Graduate Texts in Mathematics*, vol. 114. Berlin, Germany: Springer-Verlag, 1994.

[54] G. Lachaud, "Les codes géométriques de Goppa," *Séminaire Bouraki*, no. 641, pp. 189–207, 1985.

[55] D. le Brigand, "Decoding of codes on hyperelliptic curves," in *Lecture Notes in Computer Science*, G. Cohen and P. Charpin Eds., vol. 514. Berlin: Springer-Verlag, 1990, pp. 126–134.

[56] ———, "A generalized Forney formula for AG-codes," *IEEE-Trans. Inform. Theory*, vol. 42, pp. 1263–1269, July 1996.

[57] R. Lidl and H. Niederreiter, "Finite fields," in *Encyclopedia of Mathematics and Its Applications.* Reading, MA: Addison-Wesley, 1983, vol. 20.

[58] Y. I. Manin, "What is the maximum number of points on a curve over F_2," *J. Fac. Sci. Univ. Tokyo*, vol. 28, pp. 715–720, 1981.

[59] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes.* Amsterdam, The Netherlands: North-Holland, 1977.

[60] R. J. McEliece, "The theory of information and coding," in *Encyclopedia of Mathematics and Its Applications.* Reading, MA: Addison-Wesley, 1977, vol. 3.

[61] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes.* Amsterdam, The Netherlands: North-Holland, 1997.

[62] A. J. Menezes, *Elliptic Curve Public Key Cryptosystems.* Dordrecht, The Netherlands: Kluwer, 1993.

[63] J. F. Michon, "Codes and curves," in *Lecture Notes in Mathematics*, H. Stichtenoth and M. A. Tsfasman, Eds., vol. 1518. Berlin, Germany: Springer-Verlag, 1988, pp. 22–30.

[64] C. Moreno, "Algebraic curves over finite fields," in *Cambridge Tracts in Mathematics 97.* Cambridge, U.K.: Cambridge Univ. Press, 1991.

[65] ———, "Algebraic curves over finite fields," in *Cambridge Tracts in Mathematics 97.* Cambridge, U.K.: Cambridge Univ. Press, 1993.

[66] C. Munuera, "On the main conjecture on geometric MDS codes," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1573–1577, 1992.

[67] M. E. O'Sullivan, "Decoding of codes defined by a single point on a curve," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1709–1719, 1995.

[68] ———, *The Key Equation for One-Point Codes and Efficient Error Evaluation*, Univ. Puerto Rico, Apr. 1998, preprint.

[69] R. Pellikaan, H. Sticthenoth, and F. Torres, "Weierstrass semigroups in an asymptotically good tower of function fields," *Finite Fields and Its Applications*, to be published.

[70] M. Perret, "Families of codes exceeding the Varshamov–Gilbert bound," in *Lecture Notes in Computer Science*, G. Cohen and P. Charpin, Eds., vol. 388. Berlin, Germany: Springer-Verlag, 1989, pp. 28–36.

[71] K. Saints and C. Heegard, "Algebraic geometric codes and multidimensional cyclic codes: A unified theory using Grøbner bases," *IEEE-Trans. Inform. Theory*, vol. 41, pp. 1733–1751, Nov. 1995.

[72] S. C. Porter, B.-Z. Shen, and R. Pellikaan, "Decoding geometric Goppa codes using an extra place," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1663–1676, Nov. 1992.

[73] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *J. Soc. Ind. Appl. Math.*, vol. 9, pp. 300–304, June 1960.

[74] D. Rotillon and J. A. Thongly, "Decoding of codes on the Klein quartic," in *Lecture Notes in Computer Science*, G. Cohen and P. Charpin, Eds., vol. 514. Berlin, Germay: Springer-Verlag, 1990, pp. 135–149.

[75] K. Saints and C. Heegard, "Algebraic-geometric codes and multidimensional cyclic codes: A unified theory and algorithms for decoding using Grøbner bases," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1733–1751, Nov. 1995.

[76] S. Sakata, "Extension of the Berlekamp–Massey algorithm to n dimensions," *Inform. Comp.*, vol. 84, pp. 207–239, 1989.

[77] S. Sakata, H. Elbrønd Jensen, and T. Høholdt, "Generalized Berlekamp–Massey decoding of algebraic–geometric codes up to half the Feng–Rao bound," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1762–1768, Nov. 1995.
[78] J. P. Serre, "Sur les nombres des points rationnels d'une courbe algebrique sur un corps fini," *Compt. Rend. Acad. Sci. Paris*, vol. 297, sèr. I, pp. 397–401, 1983.
[79] C. E. Shannon, "A texinputematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, 1948.
[80] M. A. Shokrollahi and H. Wassermann, "Decoding Algebraic–Geometric Codes Beyond the Error-Correction Bound," Berkeley, CA, 1997, preprint.
[81] J. Silverman, "The arithmetic of elliptic curves," in *Graduate Texts in Mathematics*, vol. 106. Berlin, Germany: Springer-Verlag, 1986.
[82] A. N. Skorobogatov and S. G. Vlăduţ, "On the decoding of algebraic–geometric codes," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1461–1463, 1990.
[83] H. Stichtenoth, "Algebraic function fields and codes," in *Universitext.* Berlin, Germay: Springer-Verlag, 1993.
[84] ———, "A note on Hermitian codes over GF (q^2)," *IEE Trans. Inform. Theory*, vol. 34, pp. 1345–1348, 1988.
[85] H. Stichtenoth and C. Voß, "Generalized Hamming weights of trace codes," *IEEE Trans. Inform. Theory*, vol. 40, pp. 554–558, 1994.
[86] H. Stichtenoth, "On the dimension of subfield subcodes," *IEEE Trans. Inform. Theory*, vol. 36, pp. 90–93, 1990.
[87] ———, "Self-dual Goppa codes," *J. Pure Appl. Alg.*, vol. 55, pp. 199–211, 1988.
[88] M. Sudan, "Decoding of Reed–Solomon codes beyond the error-correction bound," *J. Complexity*, vol. 13, pp. 180–193, 1997.
[89] H. J. Tiersma, "Remarks on codes from Hermitian curves," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 605–609, 1987.
[90] M. A. Tsfasman and S. G. Vlăduţ, *Algebraic–Geometric Codes.* Dordrecht, The Netherlands: Kluwer, 1991.
[91] M. A. Tsfasman, S. G. Vlăduţ, and T. Zink, "On Goppa codes which are better than the Varshamov–Gilbert bound," *Math. Nachr.*, vol. 109, pp. 21–28, 1982.
[92] ———, "Modular curves, Shimura curves and Coppa codes, better than the Varshamov–Gilbert bound," *Math. Nachr.*, vol. 109, pp. 21–28, 1982.
[93] M. A. Tsfasman, "Goppa codes that are better than the Varshamov–Gilbert bound," *Probl. Inform. Transm.*, vol. 18, pp. 163–166, 1982.
[94] G. van der Geer, "Codes and elliptic curves," in *Effective Methods in Algebraic Geometry*, T. Mora and C. Traverso, Eds. Base, Switzerland: Birkhäuser, 1991, pp. 160–168.
[95] J. H. van Lint, "Introduction to coding theory," in *Gradate Texts in Mathematics*, vol. 86. Berlin, Germany: Springer-Verlag, 1982.
[96] J. H. van Lint and T. A. Springer, "Generalized Reed–Solomon codes from algebraic geometry," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 305–309, 1987.
[97] J. H. van Lint and G. van der Geer, *Introduction to Coding Theory and Algebraic Geometry.* Basel, Switzerland: Birkhäuser, 1988.
[98] J. H. van Lint, "Algebraic geometric codes," in *Coding Theory and Design Theory*, D. Ray-Chaudhuri, Ed. New York: Springer, 1990, pp. 137–162.
[99] S. G. Vlăduţ and Y. I. Manin, "Linear codes and modular curves," *J. Sov, Math.*, vol. 30, pp. 2611–2643, 1985.
[100] S. G. Vlăduţ, "An exhaustion bound for algebraic–geomeric 'modular' codes," *Probl. Inform. Transm.*, vol. 23, pp. 22–34, 1985.
[101] C. Voß and T. Høholdt, "An explicit construction of a sequence of codes attaining the Tsfasman–Vlăduţ–Zink bound. The first steps," *IEEE Trans. Inform. Theory*, vol. 43, pp. 128–135, Jan. 1997.
[102] S. B. Wicker and V. K. Bhargava, Eds., *Reed–Solomon Codes and Their Applications.* Piscataway, NJ: IEEE Press, 1994.
[103] J. K. Wolf, "Adding two information symbols to certian nonbinary BCH codes and some applications," *Bell Syst. Tech. J.*, vol. 49, no. 2, pp. 2405–2424, 1969.
[104] K. Yang and P. V. Kumar, "On the true minimum distance of Hermitian codes," in *Lecture Notes in Mathematics*, H. Stichtenoth and M. A. Tsfasman, Eds., vol. 1518. Berlin, Germany: Springer-Verlag, 1988, pp. 99–107.
[105] T. Yaghoobian and I. F. Blake, "Hermitian codes as generalized Reed–Solomon codes," *Des., Codes Cryptogr.*, vol. 2, pp. 5–17, 1992.
[106] ———, "Codes from hyperelliptic curves," in *Allerton Conf. Communication, Control and Computing*, 1992.
[107] C. Xing, "On automorphism groups of the Hermitian codes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1629–1635, 1995.