Preliminary version
February 28, 2000

II Università degli
Studi di Roma

TOR VERGATA

Algebraic curves and coding theory. Abuja 1990

René Schoof

Dipartimento di Matematica
2^{a} Università di Roma "Tor Vergata"
I-00133 Roma ITALY
Email: schoof@fwi.uva.nl

Abstract. These notes are an extension of lectures given at the National Mathematical Centre in Abuja, Nigeria during the summer of 1990. They contain brief introductions to algebraic coding theory, the geometry of algebraic curves and class field theory of algebraic curves. Goppa's method to contruct codes by means of linear systems on algebraic curves over finite fields is discussed. Bombieri's proof of the Riemann hypothesis for curves is given. In the last sections some recent results on the number of points on curves over finite fields are discussed. These include work by Drinfeld, Ihara, Oesterlé, Serre and Vlădut. Finally we prove a new lower bound for the number of rational points a curve of high genus over $\mathbf{F}_2$ can have.

Introduction.

Codes are used to correct errors when transmitting messages through noisy channels such as telegraph wires and telephone cables. They have been used transmitting photographs from the planet Mars and they are present in compact discs. Their purpose is to detect and correct errors that occurred during the transmission of information. The basic idea to do this, is to add, in a structured way, redundant symbols to the messages. The receiver can then check whether the structure has been preserved after transmission. If it hasn't, he knows that an error has occurred. If the added structure has been sufficiently rich, the receiver can, as we will see, even *correct* the errors.

In an appendix to these notes I've included an illustration of how codes can correct errors. The reader will find there three versions of a text taken from pages 6 and 7 of [34]. The first is the original (including some unintentional typing errors). The second is the text after a simulated transmission of the text trough a very noisy channel: using a random number generator 4% of the approximately 21 000 bits were flipped during transmission. We see that as a result of this the text became more or less unreadable. The third text is the text after transmission through the same noisy channel, this time using a code. We see that the text is readable again ! Only a few errors were not corrected. The code used was the so-called *binary Golay code* . We will describe it in more detail in section 2.

In this introduction we will only mention a few simple codes. A well-known example is the so-called "parity-check-bit": Suppose, by way of example, that one transmits characters using 7 bits to represent them in binary. In this way one can represent $2^7 = 128$ different characters. One now *encodes* the 7 bits $\epsilon_1 \ldots \epsilon_7$ by adding an 8th bit ϵ_8, the parity-check-bit, such that the sum

$\sum_{i=1}^{8} \epsilon_i$ is congruent to 0 (mod 2). Now one transmits messages using common 8-bit bytes. The receiver checks whether the relation $\sum_{i=1}^{8} \epsilon_i \equiv 0 \pmod 2$ is satisfied. In this way he can detect 1-bit errors in a byte but there is not have enough information to correct them.

Another example is the so-called "repetition code": instead of sending a bit once, just send it many times, say, five times. The receiver assumes that only a few errors have been made during transmission. He will now *decode* the received messages as follows: when he receives more ones than zeroes he decides that a one has been sent and vice versa. Using this code, the receiver can correct errors of at most two bits in one quintuple. There is a price to pay, however: the sender has to transmit 5 times as many bits.

These two examples illustrate the basic conflict in coding theory: on the one hand one wishes to be able to correct as many errors as possible, on the other hand, one wishes to transmit as few extra symbols as necessary. Our first example is extreme in one sense. One adds only very little redundancy, but one can merely detect certain errors. The second example is extreme in the other sense. One can correct more errors, but one has to add very many redundant bits.

Example. The ISBN-code.
Many books have nowadays an International Standard Book Number. It is a 9 digit number followed by either a digit or an X. The famous *Theory of Error-Correcting Codes* by MacWilliams and Sloane [16] has the number 0-444-85193-3. The book *Introduction to Coding Theory and Algebraic Geometry* by Van Lint and Van der Geer [33] has 3-7643-2230-6. The text *Algebraic Number Theory* by Stewart and Tall [28] has 0-412-29690-X. The last symbol is a "parity-check" symbol. It can be computed as follows: if the first 9 digits are $a_1 a_2 \ldots a_9$ then the last symbol a_{10} is defined by

$$\sum_{i=1}^{9} i a_i \equiv a_{10} \pmod{11}.$$

Here a_{10} is taken to be X whenever $a_{10} \equiv 10 \pmod{11}$.

C.E. Shannon showed in 1948 that good codes exist: no matter how noisy the channel is, for every $\epsilon > 0$ there exist codes, with reasonable redundancy rate, that will correct each error with probability $> 1 - \epsilon$. These codes are very long i.e. the redundant bits will not be computed from 7 bits as in the parity-check-code above but will depend on many, many more bits. One of the goals of coding theory is to contruct such good codes explicitly. In 1982 M.A. Tsfasman, S.G. Vlădut and Th. Zink [29] contructed a sequence of very good codes. The codes, so-called Goppa codes, violated the conjectural Gilbert-Varshamov bounds and their existence was a big surprise. They were contructed by means of certain linear systems on algebraic curves over finite fields.

In this course we discuss codes in sections 1 and 2 and algebraic curves in section 3. Both these sections are very brief and sketchy. They should not be considered as a serious introduction to the respective topics. In particular, no attention will be paid to encoding and decoding algorithms and we will use, but not prove the Riemann-Roch theorem for algebraic curves. We will discuss the Goppa-codes that can be constructed by means of algebraic curves over finite fields in section 4. It appears that curves over a finite field $\mathbf{F}_q$ that have many points over $\mathbf{F}_q$ with respect to their genus, give rise to good codes. Unfortunately, an algebraic curve cannot have too many points with respect to its genus. This issue is the subject matter of the remaining sections.

In section 5 we study ζ-functions of algebraic curves over finite fields. In section 6 we prove André Weil's famous theorem [35]: the Riemann hypothesis for curves over finite fields. It implies certain bounds for the number of rational points on a curve. Weil's proof appeared in 1948. We present S.A. Stepanov's proof given by E. Bombieri [3] in Séminaire Bourbaki in 1973. Stimulated by the relations to coding theory, the estimates for the number of $\mathbf{F}_q$-rational points implied by Weil's theorem were reconsidered by several people in the early 1980's. It appeared, rather unexpectedly, that the estimates were not best possible. In section 7 we will prove the theorem of V.G. Drinfeld and S.G. Vlădut [7] which, roughly speaking, says that the number of $\mathbf{F}_q$-rational points on a curve of genus g over $\mathbf{F}_q$ has to be rather small with respect to g when g is very large. This indicates the limitations of the methods to construct codes with algebraic curves. Nevertheless, it is now interesting to search for curves that have many $\mathbf{F}_q$-rational points with respect to their genus. Using class field theory, explained briefly and without proofs in section 8, we study this problem in section 9. The discussion is based on J.-P. Serre's Collège de France course in 1983–1984. See [22,23,24,25]. Section 9 contains the only new result in these notes: the existence of algebraic curves X of arbitrary high genus g over $\mathbf{F}_2$, for which the ratio $\#X(\mathbf{F}_2)/g$ is at least $2/9$; see [19].

1. Coding Theory: The Hamming code.

In this section and the next we will briefly outline the basic principles of coding theory. This should not be considered as a thorough introduction to the subject. We will only mention a few codes as examples or because of their relations to certain Goppa codes that will be studied in section 4. The reader who wishes to obtain a more complete knowledge of coding theory should consult more extensive texts like the book by MacWilliams and Sloane [16], the one by Van Lint [32] or the book by Tsfasman and Vlădut [30], which will be translated into English soon. For an informal historical introduction to coding theory see [6].

We will discuss one special code in detail: the [7,4,3]-Hamming code. We will explain in detail how to use this code: how to encode and how to decode. We begin by introducing some basic notions in coding theory.

The messages will be transmitted using an *alphabet.* In practice, this alphabet will often consist of the two bits 0 and 1. We will, more generally, consider alphabets that are finite fields. We write $\mathbf{F}_q$ for a field with q elements. The alphabet $\{0, 1\}$ corresponds to the field $\mathbf{F}_2$. Codes over this field are called *binary codes* . When transmitting, the entire message will be split into blocks of fixed length and each block will be encoded i.e. each block will be provided with certain redundant letters of the alphabet. In other words, encoding is an injective map E

$$\mathbf{F}_q^k \xrightarrow{E} \mathbf{F}_q^n$$

where k is the length of the blocks and $n \geq k$ is the length of an encoded block. We will call the image of E the *code* C. Its elements are called *code words* or simply *words* . The integer n is called the length of the code. The code words are the ones being transmitted.

When no errors occur during transmission, the receiver will receive a code word and, knowing the map E, he can recover the k bits of the original block. If however, the receiver encounters a word $v \notin C$, then an error has occurred and he will assume that the error made is small i.e. he will look for a word $\tilde{v} \in C$ which resembles v in the sense that $\tilde{v}$ and v are different in only a few coordinates. The whole process is very similar to the way we ourselves detect and correct printing errors in a text: the word *immediaxely* is not an English word, but it is very close to the English word *immediately.* It differs in one place only. When we encounter the word *immediaxely* in a text, we "correct" the error by assuming that a small error has been made and that *immediately* was meant.

This leads to the notion of distance in $\mathbf{F}_q^n$: For $v \in \mathbf{F}_q^n$ we define the *weight* $|v|$ of a word v by

$$|v| = \text{ the number of non-zero coordinates of } v.$$

The *Hamming distance* between two vectors v and w in $\mathbf{F}_q^n$ is simply $|v - w|$. A very important invariant of a code is the *minimal distance* d: it is the minimal Hamming distance between distinct words in the code C. Since the Hamming distance satisfies the triangle inequality, every "ball" of radius less than $d/2$ contains at most one code word. Therefore one can, in principle, correct upto $[(d-1)/2]$ errors in each word.

Clearly, the map E should be injective. We will only consider codes where the map E is also linear. This gives rise to the class of so-called *linear* codes. The linearity of C implies that

$$d = \min_{v \neq 0}\{|v| : v \in C\}.$$

Recapitulating we have

Definition (1.1). *A linear code C over an alphabet $\mathbf{F}_q$ is an $\mathbf{F}_q$-linear subspace of $\mathbf{F}_q^n$. The dimension k of C is the dimension of C as an $\mathbf{F}_q$-vector space and the minimal distance d of C is $d = \min_{v \neq 0}\{|v| : v \in C\}$.*

A code C with the parameters n, k and d, is called a $[n, k, d]$-code. There are certain restrictions on the possible values of k and d. We mention a very easy one.

Proposition (1.2). *(The Singleton bound) For every linear $[n, k, d]$-code one has the inequality $k + d \leq n + 1$.*

Proof. Since the code has dimension k, there exists a codeword with at least $k-1$ coordinates equal to 0. This word has weight at most $n-(k-1)$. Therefore $d \leq n+1-k$ as required.

Example (1.3). The parity-check code is just $\{v = (v_1, \ldots, v_n) : \sum_i v_i = 0\} \subset \mathbf{F}_q^n$. It is a $[n, n-1, 2]$-code. The repetition code is the subspace of $\mathbf{F}_q$-multiples of the vector $(1, 1, \ldots, 1) \in \mathbf{F}_q^n$. It is a $[n, 1, n]$-code.

Example (1.4). The code that we used to encode and decode the text in the appendix was the binary Golay code. We will describe this code in more detail in section 2. It is a $[23, 12, 7]$-code. We encoded two letters at the time by taking the $2 \times 6 = 12$ bits of their ASCII-codes as the $k = 12$ information symbols. Using the Golay code, one can correct upto 3 errors in each 23-bit word.

One could view the ISBN-code from the introduction as a subset of $\mathbf{F}_{11}^{10}$. It is not a linear code, however, since the first 9 coordinates of the code words are only in $\{0, 1, \ldots, 9\}$ which is not a full set of representatives for $\mathbf{F}_{11}$ in $\mathbf{Z}$. One could say that it is a $[10, 9, 2]$-code over $\mathbf{F}_{11}$ not all of which code words are used.

The rest of this section will be devoted to the $[7, 4, 3]$-Hamming code H. It is the binary code in $\mathbf{F}_2^7$ generated by the four vectors

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

Equivalently, it is the 4-dimensional subspace of $\mathbf{F}_2^7$ given by the linear equations

$$\begin{pmatrix}0&0&0&1&1&1&1\\0&1&1&0&0&1&1\\1&0&1&0&1&0&1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\\x_4\\x_5\\x_6\\x_7\end{pmatrix} = 0.$$

Here is a complete list of all 16 code words:

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

When using this code, we let the first 4 coordinates contain the information bits. We'll transmit the unique vector $v \in H$ which has its first 4 coordinates $x_1, \ldots, x_4$ equal to the 4 information bits. It is easy to find v by means of a matrix multiplication:

$$v = \begin{pmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 0 & 1 & 1 & 1\\ 1 & 0 & 1 & 1\\ 1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1\\ x_2\\ x_3\\ x_4 \end{pmatrix} = \begin{pmatrix} x_1\\ x_2\\ x_3\\ x_4\\ x_2+x_3+x_4\\ x_1+x_3+x_4\\ x_1+x_2+x_4 \end{pmatrix}.$$

Now v will be transmitted and a, possibly erronous, vector $\tilde{v}$ is received. We must decode $\tilde{v}$ by looking for a vector $w \in C$ nearest to $\tilde{v}$ and "hope" that w is indeed the message sent. In general this is a very difficult problam, but here it can be done as follows: the vector $\tilde{v}$ is in some coset C' of the code H. We look for a vector $v' \in C'$ of minimal weight. The vector $w = \tilde{v} - v'$ is now in C and as close as possible to $\tilde{v}$. We will assume that w is the original message.

How to find vectors of small weight in the cosets of C? For our small Hamming-code H this is very easy. By inspecting the 16 vectors in C given above one sees that the minimal distance d is equal to 3. This implies at once that the 7 basis vectors $e_1, \ldots e_7$ are each in a different non-trivial coset of C. It so happens that there are precisely $2^3 - 1 = 7$ non-trivial cosets of H. So each basis vector e_i is contained in precisely one coset C_i of H. To decode an erronous $\tilde{v}$ we must decide in which coset C_i it is. This can be done by means of the following elegant trick: Consider the matrix

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

it is the matrix of equations for the code H given above. So $C = \ker(A)$ and A is constant on cosets of C. Observe now that the column vectors of A can be viewed as the numbers "1" upto "7" written in binary: we see that $Ae_i =$"i". Now the decoding recipee is clear: apply A to $\tilde{v}$. You will find a vector "i" where $0 \leq i \leq 7$. If $i = 0$ the vector $\tilde{v}$ is in C and probably no error has been made. If $i \neq 0$ then $\tilde{v}$ is in the same coset as e_i and you should take, in the above notation, $v' = e_i$. In other words: an error has been detected in the i-th coordinate. Change the bit there.

Exercises.

(1.A) Find the parameters n, k, and d of the code $\{(0,0,0,0),(0,0,1,1),(1,1,0,0),(1,1,1,1)\} \subset \mathbf{F}_2^4$.

(1.B) Let $< v, w >$ denote the usual scalar product on the vector space $\mathbf{F}_q^n$: For $v = (v_i)_i$ and $w = (w_i)_i$ one has $< v, w >= \sum_{i=1}^n v_i w_i$. By $V^\perp$ we denote the *dual* of a linear subspace V of $\mathbf{F}_q^n$: that is $V^\perp = \{v \in \mathbf{F}_q^n : < v, w >= 0 \text{ for all } w \text{ in } V\}$. Show that $(V^\perp)^\perp = V$. Show that the dual of a code is again a code. What can you say about its parameters ? Show that the parity-check code and the repetition codes are dual to one another.

(1.C) Suppose that $\mathbf{F}_q$ is a subfield of $\mathbf{F}_r$ and suppose that $C \subset \mathbf{F}_r^n$ is a code over $\mathbf{F}_r$. Show that $C' = \mathbf{F}_q^n \cap C$ is a code over $\mathbf{F}_q$. The code C' is called the *restriction* of C to $\mathbf{F}_q$. What can you say about its parameters?

(1.D) Show that the weights of the code words of a *self-dual* binary code C are even. Suppose that this code C is generated by $v_1, \ldots, v_m$. Show that if $|v_i| \equiv 0 \pmod 4$ for $1 \le i \le m$ then $|x| \equiv 0 \pmod 4$ for every code word x.

1.E) Use the Hamming code and encode the the vectors $(0,1,1,0)$ and $(1,1,1,0)$. Decode the following words: $(0,0,1,1,1,1,1)$ and $(1,0,1,0,0,0,0)$. Which are the weights that occur in the Hamming code H? Which weights and with which frequencies occur in the cosets of H? What are the parameters of the even-weight subcode of H?

(1.F) Construct an *extended* Hamming code H' by adding a parity-check bit to the words of H. What are the parameters of the new code ? Show that it is self-dual. Determine the weight distribution of H'.

(1.G) The Hamming ball $B(v,r)$ with center v and radius r is the collection of vectors $w \in \mathbf{F}_q^n$ with $|w-v| \le r$. What is the cardinality of $B(v,r)$? A code $C \subset \mathbf{F}_q^n$ is called *perfect* if there is an integer r such that $\mathbf{F}_q^n$is the disjoint union of the Hamming balls $B(c,r)$ with $c \in C$. Show that the Hamming code is perfect. Are the parity-check codes or the repetition codes perfect?

(1.H) Let $V = \mathbf{F}_2^{23}$. Calculate the cardinalities of the Hamming balls $B(0,k)$ for $1 \le k \le 23$. Show that the binary $[23,12,7]$-Golay code is perfect.

2. Coding Theory: Cyclic Codes.

In this section we will discuss the important class of cyclic codes. We will study generalized Hamming codes, BCH-codes and quadratic resisue codes.

Definition (2.1). *A cyclic code is an ideal in $\mathbf{F}_q[X]/(X^n-1)$. The $\mathbf{F}_q$-basis for this ring is formed by the monomials $1, X, \ldots, X^{n-1}$.*

Examples. The ideal generated by $X-1$ consists of the poynomials f with $f(1)=0$. We see that this is the parity-check code. The other extreme, the ideal generated by $(X^n-1)/(X-1)$ consists of the scalar multiples of the polynomial $1 + X + X^2 + \ldots + X^{n-1}$. This is clearly the repetition code.

Cyclic codes are called cyclic because whenever $f(X)$ is in the code, so is $Xf(X)$. Looking at the coefficients of f, we see that this means that whenever a vector is in a cyclic code, so are all its cyclic shifts. Next we introduce binary Hamming codes. For non-binary Hamming codes see Exercise 2.A.

For $f \ge 1$ let α denote a generator of the cyclic group $\mathbf{F}_{2^f}^*$ and let $\phi(X) \in \mathbf{F}_2[X]$ denote its minimal polynomial. The ideal generated by $\phi(X)$ in the ring $\mathbf{F}_2[X]/(X^{2^f-1}-1)$ is called a *(binary) Hamming code* .

Let H be a Hamming code. Since α has order 2^f-1 in the multiplicative group, there are no non-zero words of weight 2 or less in a Hamming code. We conclude that the minimal distance d is at least 3. On the other hand, the Hamming balls (see Ex.1.F) of radius 1 and center in H are disjoint, they each contain 2^f vectors and there are 2^{2^f-f} of them. By counting we see that these balls cover the vector space $\mathbf{F}_2[X]/(X^{2^f-1}-1)$. We conclude that the minimal distance is actually equal to 3. The Hamming codes are $[2^f-1, 2^f-f-1, 3]$-codes. They are perfect (see Ex.1.F). One recovers the Hamming code of section 1 as the ideal generated by the polynomial X^3+X+1 in $\mathbf{F}_2[X]/(X^7-1)$. It is the case with $f=3$. To see that the codes coincide it suffices to permute the first, second and fourth coordinates of the code of section 1. A vector $(a_1, a_2, \ldots, a_7)$ then corresponds to the polynomial $a_1X^6 + a_2X^5 + \ldots + a_7$.

Next we study BCH-codes over arbitrary base fields $\mathbf{F}_q$. BCH-codes were invented by Bose, Ray-Chaudhuri and Hocquenghem in 1959. These codes come with a designed distance t: Let n be a positive integer and let $\alpha \in \overline{\mathbf{F}}_q$ have order n. The ideal $I \subset \mathbf{F}_q[X]/(X^n - 1)$ consisting of the polynomials $f(X)$ for which $f(\alpha^b) = f(\alpha^{b+1}) = \ldots = f(\alpha^{b+t-2}) = 0$ for some b is a called a *BCH-code of designed distance t.*

The ideal I is, of course, generated by the lowest common multiple of the minimum polynomials of $\alpha^b, \alpha^{b+1}, \ldots, \alpha^{b+t-2}$.

Proposition (2.2). *A BCH-code with designed distance t has minimal distance $d \geq t$.*

Proof. Suppose $f(X)$ is a code word of weight less than t. Suppose $a_{i_1}, a_{i_2}, \ldots a_{i_{t-1}}$ are its only possibly non-zero-coefficients. Substituting $\alpha^b, \alpha^{b+1}, \ldots, \alpha^{b+t-2}$ into f we find a system of linear equations:

$$\begin{pmatrix} \alpha^{i_1 b} & \ldots & \alpha^{i_{t-1} b} \\ \alpha^{i_1 (b+1)} & \ldots & \alpha^{i_{t-1}(b+1)} \\ . & & . \\ . & & . \\ . & & . \\ \alpha^{i_1(b+t-1)} & \ldots & \alpha^{i_{t-1}(b+t-1)} \end{pmatrix} \begin{pmatrix} a_{i_1} \\ a_{i_2} \\ . \\ . \\ . \\ a_{i_{t-1}} \end{pmatrix} = 0.$$

It is easily seen that this matrix is equal to $\alpha^{(i_1+i_2+\ldots+i_{t-1})b}$ times an invertible Vandermonde matrix. This implies that the matrix is non-singular and we conclude that $a_{i_1} = a_{i_2} = \ldots = a_{i_{t-1}} = 0$ and hence that $f = 0$ as required.

Since every $f(X) \in \mathbf{F}_2[X]$ for which $f(\alpha) = 0$ also satisfies $f(\alpha^2) = 0$ we see that binary BCH-codes automatically have minimum distance at least 3. Binary Hamming codes are precisely the binary BCH-codes with designed distance 3.

Example (2.3). Let $\alpha \in \mathbf{F}_{16}^*$ be a generator of the multiplicative group. It has order 15. Let I be the ideal in $\mathbf{F}_2[X]/(X^{15} - 1)$ generated by the product $h(X)$ of the minimal polynomials of α and α^3. This is a (binary) BCH-code. Clearly $h(\alpha) = h(\alpha^2) = h(\alpha^3) = h(\alpha^4) = 0$ and we see that its designed distance is 5. The dimension of the code I is $7 = 15 - 2 \times 4$. This is a $[15, 7, \geq 5]$-code. Using it, one can correct upto 2 errors in each code word.

Example (2.4). Consider the polynomial $X^{23} - 1 \in \mathbf{F}_2[X]$. Since 2 has order 11 modulo 23, this polynomial factors as a product of $X - 1$ and two irreducible polynomials of degree 11. The *binary Golay code* is defined to be the ideal generated by one of these 11-th degree polynomials.

Example (2.5). *Quadratic residue codes.* Let p be a prime and let ℓ be a prime which is a square modulo p. Let α be a primitive root mod p and let Q denote the subgroups of squares in $\mathbf{F}_p^*$

$$q(X) = \prod_{x \in Q} (X - \alpha^x) \quad \text{and} \quad n(X) = \prod_{x \notin Q} (X - \alpha^x).$$

We have that and that $X^p - 1 = (X - 1)q(X)n(X)$ in $\mathbf{F}_\ell$. The ideals in $\mathbf{F}_l[X]/(X^p - 1)$ generated by $q(X)$, $(X - 1)q(X)$, $n(X)$ or $(X - 1)n(X)$ are the quadratic residue codes over $\mathbf{F}_l$ of length p. The binary Hamming code and the binary Golay code are examples of quadratic residue codes. Quadratic residue codes have minimum distance at least $\sqrt{p}$ (see Exercise 2.E).

As an illustration we show that the binary Golay code, that was used in the example mentioned in the introduction, is perfect.

Proposition (2.6). *The binary Golay code is a perfect $[23, 11, 7]$-code.*

Proof. Let G denote the binary Golay code. In the notation of Example 2.4 it is the ideal generated by $q(X)$ in $\mathbf{F}_2[X]/(X^{23}-1)$. There are two possibilities for $q(X)$ depending on the choice of a primitive 23-th root of unity in $\overline{\mathbf{F}}_2$. Since the sum of all non-trivial 23-th roots of unity is equal to 1, we see that one factor of $(X^{23}-1)/(X-1)$ has a X^{10}-term, while the other has not. We will take $q(X)$ to be the factor without this term.

Let $\overline{G}$ denote the code obtained from G by adding a parity check bit. This code has length 24 and dimension 12. It follows easily from Exercise 2.D below that $\overline{G}$ is self-dual. One can verify that the words $(\sum_{i\in(\mathbf{F}_{23}^*)^2} X^i, 1)$ and $(X^{22}+\ldots+X+1,1)$ form a basis for $\overline{G}$. Since there are precisely 11 non-zero squares modulo 23, these words all have weights divisible by 4. By Exercise 1.D we conclude that every word in $\overline{G}$ has its weight divisible by 4. Therefore it follows from Exercise 2.E that the minimum distance of $\overline{G}$ is at least 8 and we see that G itself has minimum distance at least 7. The result now follows from Exercise 1.H.

Finally we mention the so-called *Reed-Solomon codes.* These are BCH-codes of length $q-1$ over $\mathbf{F}_q$. So, they are ideals generated by a polynomial of the form $\prod_{i=1}^{d}(X-\alpha^i)$ where $\alpha\in\mathbf{F}_q^*$ is some primitive root. They have length $n-d+1$ and designed distance d. By the Singleton-bound (Prop.1.2), the minimal distance is actually equal to d. Therefore the Reed-Solomon codes are $[n, n-d+1, d]$-codes. Of course one should have $q>2$ in order to obtain a somewhat interesting code. Reed-Solomon codes are used in compact discs.

Exercises.

(2.A) Show that the BCH-codes of length $(q^f-1)/(q-1)$ over $\mathbf{F}_q$ and designed distance 2 do, in fact, have minimum distance 3 and are perfect. These codes are the so-called *q-ary Hamming codes.*

(2.B) Show that in Example 2.3 one can take $h(X)=(X^4+X+1)(X^4+X^3+X^2+X+1)$.

(2.C) Let α be a zero of the polynomial that generates the binary Golay code. Express the other zeroes in terms of α. Show that Proposition 2.2 implies that the minimum distance is at least 5.

(2.D) Let C be a cyclic code with generator polynomial g i.e. $C=(g)\subset\mathbf{F}_q[X]/(X^n-1)$. Let $h(X)=(X^n-1)/g(X)$. Show that $C^\perp$ has generator polynomial $X^{\deg h(X)}h(X^{-1})$.

(2.E) Prove that the minimum distance d of a quadratic residue code of length p over $\mathbf{F}_l$ is at least $\sqrt{p}$ (Consider "both" codes at once). Prove that, if $p\equiv 3 \pmod 4$, then one even has that $d^2-d+1\geq p$. Show that the binary Golay code (Example 2.4) has minimum distance at least 6.

3. Algebraic Curves.

In this section we will briefly review some of the fundamental results in the theory of algebraic curves over a field. In the next section we will apply these results to curves over finite fields and construct the Goppa codes. It is therefore important to consider curves over fields that are not necessarily algebraically closed. Although in the literature many of the basic theorems on curves are formulated over an algebraically closed field, they usually hold over perfect fields as well, see [27,Lemma II.5.8.1]. We do not pretend to give a self-contained introduction to the theory of curves. This would take us too far. Very often proofs will not be given. If the reader wishes to study the theory more thoroughly, she should study other, more extensive texts, like Hartshorne's book [12], or older books like [5,8,20,26]. There is also a proof of the Riemann-Roch theorem in [31].

A curve X over a field k is a smooth irreducible projective algebraic variety of dimension 1. Every curve can be embedded into some projective space $\mathbf{P}^n$ over k as a closed subvariety. This means that we can describe a curve as the common zero locus of a finite collection of homogeneous polynomials. By the irreducibility of X, the ideal generated by these polynomials is a prime ideal. The function field i.e. the field of rational functions on X over k will be denoted by $k(X)$. Note that these "functions" need not be defined everywhere. They may have poles. When $k=\mathbf{C}$ they are precisely the meromorphic functions on X.

We will henceforth assume that k is perfect and we fix an algebraic closure $\bar{k}$ of k. For any field F with $k \subset F \subset \bar{k}$ we let $X(F)$ denote the set of points on X that are defined over F i.e. they have their coordinates in F. We have, of course that $X(F) \subset X(\bar{k})$, the set of "all" points on X.

The degree $\deg(P)$ of a point P is the degree over k of the field of definition of P. A divisor D on X is a formal sum $D = \sum_P n_P P$ where the sum runs over the points $P \in X(\bar{k})$ and the $n_p \in \mathbf{Z}$ are almost all 0. The divisors form a free abelian group $\mathrm{Div}(X)$. Examples of divisors are provided by the divisors of rational functions on X: for $f \in \bar{k}(X)$ we let $\mathrm{div}(f)$ denote the divisor $\sum_{P \text{ zero of } f} P - \sum_{P \text{ pole of } f} P$ where one should count with multiplicities.

A divisor D is said to be defined over $F \subset \bar{k}$ if it is left fixed by the Galois group $\mathrm{Gal}(\bar{k}/F)$. The divisors that are defined over a subfield F of $\bar{k}$ form a subgroup $\mathrm{Div}_F(X)$ of the divisor group $\mathrm{Div}(X)$. It is immediate that $\mathrm{div}(f)$ is defined over F whenever $f \in F(X) \subset \bar{k}(X)$. Clearly the divisors of functions form a subgroup of $\mathrm{Div}(X)$: the *principal* divisors. The quotient group is called the *Picard group of* X:

$$\mathrm{Pic}(X) = \mathrm{Div}(X)/\{(f) : f \in \bar{k}(X)\}$$

When k is a finite field, then, for a finite extension F of k inside $\bar{k}$, one has that $\mathrm{Pic}_F(X) = \mathrm{Div}_F(X)/\{(f) : f \in F(X)\}$ is equal to the subgroup of $\mathrm{Gal}(\bar{k}/F)$-invariants of $\mathrm{Pic}(X)$.

The degree of a divisor $D = \sum_P n_P P$ is given by $\sum_P n_P$. The degree is a homomorphism $\mathrm{Div}(X) \longrightarrow \mathbf{Z}$. Its kernel is denoted by $\mathrm{Div}^0(X)$. We'll say that $D \leq D'$ for two divisors $D = \sum_P n_P P$ and $D' = \sum_P m_P P$ if and only if $n_P \leq m_P$ for all points P. Divisors D that satisfy $D \geq 0$ are called *effective* . To distinguish a point P from the divisor determined by it, we will sometimes write (P) for this divisor.

When k is not algebraically closed, it is sometimes convenient to modify the notion of a point somewhat: a *place* v of a curve X is a Galois conjugacy class of points in $X(\bar{k})$. The degree $\deg(v)$ of a place v is its cardinality. The divisor (v) associated to a place v is the sum of its points. It is automatically defined over k. Every divisor D that is defined over k can be written as a sum of places. One has for $D = \sum_v n_v v$ that $\deg(D) = \sum_v n_v \deg(v)$. A place is the analog of a prime ideal in number theory. They will prove convenient in our study of ζ-functions in section 5 and in section 9.

A morphism from a curve X to another curve Y is an algebraic map f which is defined everywhere. The degree $\deg(f)$ of f is defined to be the degree of the corresponding extension $k(X) \hookleftarrow k(Y)$. Since X is non-singular, one can extend every rational map $X \longrightarrow \mathbf{P}^n$ to a morphism. We will use the following Theorem from Hartshorne's book:

Theorem (3.1). *Let $f : X \longrightarrow Y$ be a non-constant map of curves. Let P be a point of Y. Then $f^{-1}(P)$ is a divisor on X and its degree satisfies* $\deg(f^{-1}(P)) = \deg(f)$.

Proof. This is Hartshorne's Proposition 6.9 in Chapter II of [12].

Corollary (3.2). *The degree* $\deg(\mathrm{div}(f))$ *of the divisor of a function f is 0.*

Proof. We can view f as a rational function $X \longrightarrow \mathbf{A}^1$. Since X is smooth, it can be extended to a morphism $X \longrightarrow \mathbf{P}^1$. Now we apply Theorem 3.1 to the point $P = 0$ and the point $P = \infty$. We find that the degree of the zero-divisor of f is equal to the degree of its pole-divisor. This proves the Corollary.

In other words, the subgroup of principal divisors is contained in $\mathrm{Div}^0(X)$.We will denote the quotient by $\mathrm{Pic}^0(X)$. For finite ground fields k one can show, that for every finite extension F of k inside $\bar{k}$, the group $\mathrm{Pic}^0_F(X) = \mathrm{Div}^0_F(X)/\{(f) : f \in F(X)\}$ is equal to the $\mathrm{Gal}(\bar{k}/F)$-invariants of $\mathrm{Pic}^0(X)$.

For every divisor D of a curve X we put

$$L(D) = \{f \in k(X)^*: (f) \geq -D\} \cup \{0\}.$$

These k-vectorspaces $L(D)$ are very important. In order to discuss their properties, it is convenient to use the adèle ring $\mathbf{A}_K$ associated to K see [20].

For every place v we let O_v denote the completion of any of the local rings O_P where P is a point "of" v. This is a discrete valuation ring. By K_v we denote its quotient field. The ring of *adèles* $\mathbf{A}_K$ is the restricted product of the fields K_v with respect to the O_v:

$$\mathbf{A}_K = \{(x_v)_v : x_v \in K_v \text{ and } x_v \in O_v \text{ for almost all } v\}.$$

We let $\widehat{O}$ denote the subring of *integral* adèles $\prod_v O_v$. One can show that the k-vector space $\mathbf{A}_K/(\widehat{O}+K)$ is finite dimensional. Its dimension g is called the *genus* of the curve. For any k-rational divisor D of X we let $\widehat{O}(D)$ denote $\{x = (x_v)_v \in \mathbf{A}_K : v(x_v) \geq -v(D) \text{ for every valuation } v\}$. When D is the trivial divisor, then $\widehat{O}(D)$ is just the ring $\widehat{O}$. The *Euler characteristic* $\chi(D)$ of a divisor D is defined by

$$\chi(D) = \dim_k L(D) + \dim_k \mathbf{A}_K/(\widehat{O}(D) + K).$$

Proposition (3.3). *Let D be a k-rational divisor of X and let v be a k-rational place with residue class field k_v.*

(i) There is a natural exact sequence

$$0 \longrightarrow L(D) \longrightarrow L(D+v) \longrightarrow k_v \longrightarrow \mathbf{A}_K/(\widehat{O}(D)+K) \longrightarrow \mathbf{A}_K/(\widehat{O}(D+v)+K) \longrightarrow 0.$$

(ii) $\chi(D+v) = \chi(D) + \deg v$.

(iii) $\chi(D) = \deg D + 1 - g$.

(iv) $\deg D + 1 - g \leq \dim_k L(D) \leq \deg D + 1$.

Proof. To prove (*i*), we observe that $L(D)$ is contained in $L(D+v)$. Suppose that v occurs with multiplicity m in D. Let t be a uniformizing parameter at P. We define a map $L(D+v) \longrightarrow k_v$ as follows: Let $f \in L(D+v)$. In the local ring at P we have that $f = a_{-m-1}t^{-m-1} + a_{-m}t^{-m} + \ldots$. Now map f to its coefficient $a_{-m-1} \in k_v$. The next arrow is defined by sending $\alpha \in k_v$ to the adèle that has αt^{-m-1} at v and zero at all other places. The last arrow is the canonical one. It is straightforward to check that the sequence is exact.

Part (*ii*) follows immediately from (*i*). From (*ii*) we conclude that $\chi(D) = \deg D + \chi(0)$. Since $L(0) = k$ and, by definition, $\dim_k \mathbf{A}_K/(\widehat{O}+K) = g$, it follows that $\chi(0) = 1 - g$ and (*iii*) follows. Since $0 \leq \dim_k \mathbf{A}_K/(\widehat{O}(D)+K) \leq g$, part (*iv*) follows from (*iii*).

The problem of determining the dimensions of the spaces $L(D)$ is called the Riemann-Roch problem. Note that the dimension of $L(D)$ only depends on the divisor *class* of $D \in \mathrm{Pic}(X)$. We will formulate the theorem of Riemann-Roch which is an important statement about these dimensions.

For a curve X over k we define the module of rational or Kähler differentials Ω_X as follows: it is generated over $k(X)$ by symbols *df* where $f \in k(X)$ subject to the usual rules for derivations:

(*i*) $d(f+g) = df + dg$ for all $f, g \in k(X)$;

(*ii*) $d(fg) = fdg + gdf$ for all $f, g \in k(X)$;

(*iii*) $da = 0$ for all $a \in k$.

The module Ω_X is a vector space of dimension 1 over the function field $k(X)$; see for instance [26,III.4.Thm.3]. Fix a non-zero Kähler differential ω. We are going to associate a divisor $\mathrm{div}(\omega)$

to ω. Since the space Ω_X is one dimensional, for every point P of X and uniformizing parameter t at P there is a unique function $g \in k(X)$ such that $\omega = gdt$. The order of g at P depends only on ω and P and does not depend on the choice of t. Therefore we will write $\mathrm{ord}_P\omega$ for it. For almost all points P it is zero. This justifies the following definition:

Definition (3.4). *Let $\omega \in \Omega_X$ be a Kähler differential. The divisor* $\mathrm{div}(\omega)$ *associated to ω is*

$$\mathrm{div}(\omega) = \sum_P \mathrm{ord}_P(\omega)(P).$$

Since Ω_X has dimension 1 over $k(X)$, the *divisor class* $\kappa \in \mathrm{Pic}(X)$ does not depend on the choice of ω. This class, *the canonical class* , is a completely intrinsically defined invariant of the curve. The divisors it contains are called canonical divisors. The degree of any canonical divisor does not depend on the divisor either. It is another invariant of the curve.

Examples. Consider $X = \mathbf{P}^1$. If t is a coordinate on $\mathbf{P}^1$, then, because $d(\frac{1}{t}) = -t^{-2}dt$, we have that $\mathrm{div}(dt) = -2(\infty)$ and that $\deg(\kappa) = -2$.

Next consider a smooth cubic curve X given by the equation $y^2 = (x-e_1)(x-e_2)(x-e_3)$. This curve has genus 1. Writing P_i for the point $(e_i, 0)$, one finds that $\mathrm{div}(dx) = (P_1)+(P_2)+(P_3)-3(\infty)$. This happens to be the divisor of the function y and we conclude that $\mathrm{div}(dx/y) = 0$ and that $\deg(\kappa) = 0$.

Theorem (3.5). *(Serre duality) Let D be a divisor of the curve X and let ω be a non-trivial k-rational Kähler differential. The natural pairing*

$$L(\kappa - D) \times \mathbf{A}_K/(\widehat{O}(D) + K) \longrightarrow k$$

given by

$$(f, x) \mapsto \sum_v \mathrm{Res}_P(f \cdot x_v \cdot \omega)$$

is non-degenerate. Here the residue $\mathrm{Res}_P(\omega')$ of a differentail ω' at a point P of a place v is the coefficient a_{-1} of an expansion $\omega' = \sum_i a_i t^i dt$ in a uniformizing parameter t at P.

Proof. For a proof see [20,Ch.II]. Note that it is not even clear that the residues do not depend on the choice of the local parameters!

As a consequence we obtain that

$$g = \dim_k \mathbf{A}_K/(\widehat{O} + K) = \dim_k L(\kappa)$$

and we recover the usual definition of the genus g: the number of independent "holomorphic" differentials.

Theorem (3.6). *(Theorem of Riemann-Roch.) Let X be a curve of genus g over k and let D be a divisor of X defined over k. We have that*

$$\dim_k L(D) - \dim_k L(\kappa - D) = \deg(D) + 1 - g.$$

Proof. This follows from Prop. 3.3(*iii*) and Theorem 3.5.

Corollary (3.7).
(i) The degree of κ is $2g-2$.
(ii) If $\deg(D) > 2g-2$ then $\dim_k L(D) = \deg(D) + 1 - g$.

Proof. (i) take $D \in \kappa$. (ii) If $\deg(D) > 2g-2 = \deg(\kappa)$, then $L(\kappa - D) = 0$ by Proposition 3.3(iv). This implies the result at once.

Next we'll see how certain divisors give rise to morphisms from X to projective spaces. Let D be a divisor on X. By $|D|$ we denote the complete linear system associated to D, i.e. the set of *effective* divisors in the same divisor class as D. The projective space $\mathbf{P}(L(D))$ associated to the vectorspace $L(D)$ is in one-to-one correspondence with the set $|D|$ via $f \mapsto (f) + D$. A subspace of this projective space is called a linear system. A point P is called a *base point* of a linear system if it is in the support of all divisors in it. If a complete linear system $|D|$ has no base points then the morphism $X \longrightarrow \mathbf{P}^{d-1}$ given by $P \mapsto (f_1(P) : \ldots : f_n(P))$ is well defined. Here the f_i are a basis for $L(D)$ over k. The choice of the basis does not affect the isomorphism class of the morphism. Moreover, if D is a divisor defined over a finite extension F of k inside $\bar{k}$ then so is the induced map to projective space.

Proposition (3.8). *Let D be a K-rational divisor on a curve X. Then :*

(i) The linear system $|D|$ has no base points if and only if $\dim_k L(D-P) = \dim_k L(D) - 1$ for every point P.

(ii) Suppose that D has no base points. The map from X to projective space induced by D is a closed immersion if and only if $\dim_k L(D-P-Q) = \dim_k L(D) - 2$ for all points P and Q.

Proof. (i) Clearly the inclusion $L(D-P) \hookrightarrow L(D)$ is an equality if and only if P is a base point of D. This proves (i). To prove (ii) we must show that the morphism from X to projective space separates points and tangents. Well, it separates points if and only if for every two distinct points P and Q of X, the point Q is not a base point of $|D-P|$. By (i) this is equivalent to $\dim_k L(D-P-Q) = \dim_k L(D) - 2$. The morphism separates tangents if and only if for each point P there is a divisor $D' \in |D|$ in which P occurs with multiplicity one. This just says that P is not a base point of $|D-P|$ which by (i) is equivalent to $\dim_k L(D-2P) = \dim_k L(D) - 2$. This proves the proposition.

Proposition (3.9). *Let D be a k-rational divisor on a curve X of genus g. We have*

(i) If $\deg(D) \geq 2g$ then $|D|$ has no base points.

(ii) If $\deg(D) \geq 2g+1$ then D induces a closed immersion $X \hookrightarrow \mathbf{P}^n$.

Proof. Using Corollary 3.7 to the Theorem of Riemann-Roch one computes the dimensions of the spaces $L(D)$, $L(D-P)$ and $L(D-P-Q)$. The result then follows at once from Proposition 3.8.

We will now investigate the linear system $|\kappa|$. Recall that a curve is called *hyperelliptic* if it admits a morphism of degree 2 to $\mathbf{P}^1$.

Proposition (3.10). *Let X be a curve over k of genus $g > 0$. Then the canonical system $|\kappa|$ has no base points. It induces a closed immersion $X \hookrightarrow \mathbf{P}^{g-1}$ if and only if X is not hyperelliptic.*

Proof. Since $g \neq 0$, one has for every point P that $L(P) = k$ (Ex.3.A). By Riemann-Roch we then have that $\dim_k L(\kappa - P) = g - 1$. The first statement now follows from Proposition 3.8. To prove the second statement we consider two points P and Q on X. By Riemann-Roch we have that

$$\dim_k L(P+Q) - \dim_k L(\kappa - P - Q) = 3 - g.$$

Therefore, by Proposition 3.8, we see that the canonical system $|\kappa|$ induces a closed immersion if and only if $L(P+Q) = k$ for all points P and Q. If X is hyperelliptic and $f\colon X \longrightarrow \mathbf{P}^1$ is a morphism of degree 2, we have $L(P+Q) > k$ whenever P and Q make up some fiber of f. On the other hand, if $L(P+Q) > k$ we obtain a morphism of degree 2 to $\mathbf{P}^1$. This proves the proposition.

Exercises.

(3.A) Show that a curve X has genus 0 if and only if there is a point P on X with $\dim_k L((P)) = 2$. Show that a curve over a field k of genus 0 is isomorphic to $\mathbf{P}^1$ over k if and only if it has a k-rational point. Give an example of a curve of genus 0 over $\mathbf{R}$ which is not isomorphic to $\mathbf{P}^1$ over $\mathbf{R}$.

(3.B) Show that a curve of genus 1 can be embedded in $\mathbf{P}^2$ as a smooth cubic. Fix a point P_0 on X. Show that every divisor class of degree 0 contains a divisor of the form $P - P_0$. Use this fact to define a group structure on the set of points $X(\bar{k})$ of X.

(3.C) Show that a curve of genus 2 is hyperelliptic.

(3.D) Show that a curve of genus 3 is either hyperelliptic or a smooth quartic in $\mathbf{P}^2$. Show that a curve of genus 4 is either hyperelliptic or is a smooth curve of degree 6 in $\mathbf{P}^3$. (In the latter case the curve is the intersection of an irreducible smooth quadric and an irreducible cubic surface. This last statement is proved in [12,IV.Ex.5.2.2].).

4. Goppa Codes.

In this section we will explain how to construct codes using linear systems on curves. The basic ideas are due to the Soviet mathematician V.D. Goppa, who explained them in two papers [9,10] in 1981. See [11,30] or Lachaud's Bourbaki talk [15] for an exposition of this work.

Let X be a curve over a finite field $\mathbf{F}_q$. Let D be a divisor on X defined over $\mathbf{F}_q$ and let $P_1, \ldots, P_n$ be a collection of points in $X(\mathbf{F}_q)$ which do not occur in the divisor D. We will contruct linear codes over $\mathbf{F}_q$. We define a map θ from the $\mathbf{F}_q$-vectorspace $L(D)$ to $\oplus_{i=1}^n \mathbf{F}_q$ by $f \mapsto (f(P_1), \ldots, f(P_n))$. The Goppa code $\Gamma(D, \sum_i (P_i))$ associated to the curve X and the divisors D and $\sum_i (P_i)$ will be the image of θ.

We now want to estimate the parameters of this Goppa code. Its length is clearly n. Suppose the image of $f \in L(D)$ has weight d. This means that f vanishes in $n-d$ points. So by Prop.3.3(iv) we must have that $\deg(D) - (n-d) \geq 0$. hence $d \geq n - \deg(D)$. This gives a lower bound for the minimum distance of the code. We will assume that the bound is positive i.e we will assume that $\deg(D) < n$. Prop.3.3(iv) implies then at once that the kernel of θ, which is equal to $L(D - \sum_i (P_i))$ is trivial. Therefore the dimension k of the code is just $\dim L(D)$ which is at least $\deg(D) - g + 1$ by Prop.3.3(iv). We conclude that we have the following estimates for the parameters of the code $\Gamma(D, \sum_i (P_i))$:

$$d \geq n - \deg(D),$$
$$k \geq \deg(D) + 1 - g.$$

Let's work out an easy example. Consider the cubic curve given by $y^2 + y = x^3 + x$ over $\mathbf{F}_2$. It is a smooth curve of genus 1. It has 5 points over $\mathbf{F}_2$: $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$ and the point (∞) at infinity. Since $L(\infty)$ contains only constant functions, we take $D = 2(\infty)$; then we have $L(D) = \{0, 1, x, x+1\}$ and evaluating these functions in $L(D)$ on the remaining four rational points we obtain the following Goppa code:

$$\begin{array}{l} (0,0): \\ (0,1): \\ (1,0): \\ (1,1): \end{array} \qquad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

a hardly impressing $[4,2,2]$-code. Using $D = 3(\infty)$ we have that $L(D)$ is generated over by 1, x and y. It is easily seen that with this D one obtains the parity-check code.

One finds back the binary Hamming codes as follows: Let q be a power of 2. Take $X = \mathbf{P}^1$ over $\mathbf{F}_q$ and let t be a parameter on X. With $D = (\infty) - (0)$ we have that $L(D)$ is one dimensional, generated by t. For the rational points divisor we take the formal sum of all the other points on $\mathbf{P}^1(\mathbf{F}_q)$. This gives a one-dimensional code in $\mathbf{F}_q^{q-1}$. Its elements are $\mathbf{F}_q$-linear multiples of $(1, \alpha, \alpha^2, \ldots, \alpha^{q-2})$ where α is a generator of the multiplicative group $\mathbf{F}_q^*$. The dual of this code can easily be identified with the ideal generated by $X - \alpha$ in the ring $\mathbf{F}_q[X]/(X^{q-1} - 1)$. The restriction of this code (Ex.1.C) is precisely the Hamming code.

Taking $X = \mathbf{P}^1$ over $\mathbf{F}_q$ and $D = r(\infty) - s(0)$ and the sum of all other $\mathbf{F}_q$-rational points as the rational points divisor one finds, in a similar way, the BCH-codes of length $q-1$ and designed distance $r - s + 2$ as the restrictions of the duals of the Goppa codes.

Many of the classical codes can be viewed as Goppa codes constructed by means of certain divisors on $\mathbf{P}^1$. One can obtain good codes by considering curves of higher genus. We present one example due to A.M. Barg, S.L. Katsman and M.A. Tsfasman [2].

Example (4.1). The Klein curve X is a smooth curve of genus 3 given by the equation $x^3y + y^3z + z^3x = 0$. It has the three points $(0:0:1)$, $(0:1:0)$ and $(1:0:0)$ over $\mathbf{F}_2$. These are the only points with $xyz = 0$. We will consider X over $\mathbf{F}_8$. Let $\alpha \in \mathbf{F}_8$ be a primitive element satisfying $\alpha^3 + \alpha + 1 = 0$. Suppose $(x:y:z) \in \mathbf{F}_8$ with $xyz \neq 0$. We put $z = 1$, $y = \alpha^i$ and $x = \alpha^{3i}\xi$. We find that $\xi^3 + \xi + 1 = 0$. So $\xi = \alpha$, α^2 or α^4 and we find 21 points. Altogether there are 24 points in $X(\mathbf{F}_8)$. Let $P = (0:0:1)$ and take $D = 10(P)$. We take the sum of the remaining 23 points over $\mathbf{F}_8$ as the rational points divisor. Now we estimate the parameters of this $\mathbf{F}_8$-code: it has length 23, dimension $10 - g + 1 = 8$ and minimum distance at least $23 - \deg(D) = 13$.

Now we apply some tricks from coding theory to get a nice code over $\mathbf{F}_2$. Since $\mathbf{F}_8$ has dimension 3 over $\mathbf{F}_2$, we can view the vectors in the code as 3×23-matrices over $\mathbf{F}_2$. We now extend the code by adding a fourth "parity-check"-row and obtain 4×23-matrices. Clearly the minimum distance of this binary code is at least $2 \times 13 = 26$. We have constructed a binary $[92, 24, \geq 26]$-code. By leaving out one bit we find a binary $[91, 24, \geq 25]$-code. This beats the best known code with $n = 91$ and $d \geq 25$ (See [16,appendix A]).

From the inequalities above and the Singleton bound it is easily seen that the parameters of a $[n, k, d]$-Goppa code satisfy

$$1 + \frac{1-g}{n} \leq \frac{k}{n} + \frac{d}{n} \leq 1 + \frac{1}{n}.$$

It follows that we can make *good* codes i.e. codes with both k/n and d/n large, when we have a curve with small g and a large number n of rational points. For this reason (and others) it is interesting to try and search for curves over finite fields that have many rational points with respect to their genus. Therefore we will in the next sections, discuss the ζ-function of a curve. We will obtain a bound on the number of rational points in terms of the genus. In the last section we will exhibit some examples of curves that have many rational points.

Exercises.

(4.A) Realize the repitition codes and parity check codes as Goppa codes on $\mathbf{P}^1$.

(4.B) Show that the curve $y^2 + y = x^3 + x$ over $\mathbf{F}_2$ has genus 1 and has 25 points over $\mathbf{F}_{16}$ (use Prop.5.2). Let P be one of these 25 points and let $D = 13(P)$. We take 21 of the remaining points as the rational points divisor. Show that the corresponding Goppa code has the parameters $[21, \geq 8, \geq 13]$. Next pick an $\mathbf{F}_2$-basis of $\mathbf{F}_{16}$ and add a parity-check bit to each $\mathbf{F}_{16}$-coordinate. Show that the resulting binary code is a $[105, \geq 32, \geq 26]$-code. Deleting one bit gives a $[104, \geq 32, \geq 25]$-code which beats the code in [16, Appendix A].

(4.C) Let $\omega \in \mathbf{F}_4^*$ be a primitive third root of unity. Consider the curve $x^2y + \omega y^2z + \bar{\omega}z^2x = 0$. Show that its genus is 1 and that it has precisely 9 rational points over $\mathbf{F}_4^*$. Let $Q_1 = (\omega : 1 : 1)$, $Q_2 = (1 : \omega : 1)$ and $Q_3 = (1 : 1 : \omega)$ and let D be the divisor $2(Q_1) + (Q_2)$. Find a basis for $L(D)$. (use the function $x + y + \bar{\omega}z$; its zero-divisor is D). For the rational points divisor we take the sum of the remaining six rational points $P_1, \ldots, P_6$. Compute the $\mathbf{F}_4$-Goppa code $\Gamma(D, \sum_{i=1}^{6}(P_i))$. What are its parameters?

5. Curves over finite fields.

In this section we will study the ζ-function $\zeta_X(s)$ associated to an algebraic curve X over a finite field $\mathbf{F}_q$. We will prove that it satisfies a functional equation, similar to the one satisfied by the ordinary Riemann ζ-function. We also show that the most interesting part of the ζ-function is a polynomial in q^{-s}. The analogue of the Riemann hypothesis will be proved in the next section.

Let X be a curve over a finite field $k = \mathbf{F}_q$. We are going to associate a meromorphic complex function to X: its ζ-function $\zeta_X(s)$. It will be similar to the well-known Riemann ζ-function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} (1 - \frac{1}{p^s})^{-1}.$$

The sum and the product converge for $s \in \mathbf{C}, \operatorname{Re} s \geq 1$. It is well-known and easy to prove that $\zeta(s)$ admits a meromorphic extension to all of $\mathbf{C}$. It has only one pole. It is of order 1 at $s = 1$. The analogue of a prime number, or rather of the prime ideal in $\mathbf{Z}$ generated by it, is a place i.e. a Galois conjugacy class of points on X. Every k-rational divisor can be written as a sum of places in a unique way. The analogue of an ideal $n\mathbf{Z} \subset \mathbf{Z}$ is an effective divisor of X which is defined over k. Therefore we put for $s \in \mathbf{C}$, $\operatorname{Re} s \geq 1$:

$$\zeta_X(s) = \sum_{D \geq 0} \frac{1}{ND^s} = \prod_v (1 - \frac{1}{Nv^s})^{-1}.$$

Here the product runs over the places v of X. The norm Nv of a place v is defined to be the cardinality of the field of definition of a point P in v, in other words $Nv = q^{\deg(v)}$. Similarly we put $ND = q^{\deg(D)}$.

We will prove certain properties of the function $\zeta_X(s)$. They are similar to properties enjoyed by the Riemann ζ-function. Before stating the main result, we will first, as an example, calculate the ζ-function associated to $\mathbf{P}^1$ over $\mathbf{F}_q$. Let a_d denote the number of places on $\mathbf{P}^1$ of degree d. So, $a_1 = q + 1$ and for $d > 1$ one simply has that a_d is equal to the number of irreducible polynomials of degree d over $\mathbf{F}_q$. Obviously we have that $\sum_{d|m} da_d = \#\mathbf{P}^1(\mathbf{F}_{q^m}) = q^m + 1$.

The ζ-function of $\mathbf{P}^1$ is given by

$$\zeta_{\mathbf{P}^1}(s) = \prod_{d \geq 1} (1 - q^{-sd})^{-a_d}.$$

Therefore we consider the power series $\prod_{d \geq 1}(1 - T^d)^{-a_d}$. It is straightforward to verify that

$$\log(\prod_{d \geq 1}(1 - T^d)^{-a_d}) = \sum_{m \geq 1} \frac{1}{m} \sum_{d|m} da_d T^m = \sum_{m \geq 1} \frac{q^m + 1}{m} T^m$$

and this easily implies that

$$\zeta_{\mathbf{P}^1}(s) = \frac{1}{(1 - q^{-s})(1 - q^{1-s})}.$$

We see that this ζ-function is periodic modulo $2\pi i/\log q$. Modulo the period it has only poles at $s = 0$ and $s = 1$. It satisfies $\zeta(1 - s) = q^{1-s}\zeta(s)$.

For arbitrary curves X we have the expansion

$$\zeta_X(s) = \sum_{n=0}^{\infty} \#\{D \in \operatorname{Div}(X)\colon\ D \text{ is effective and } \deg(D) = n\} \cdot q^{-ns}$$

and therefore we put

$$Z_X(T) = \sum_{n=0}^{\infty} \#\{D \in \operatorname{Div}(X)\colon D \text{ is effective and } \deg(D) = n\} \cdot T^n \quad \in \mathbf{Z}[[T]].$$

We have the following general result:

Theorem (5.1). *Let X be a curve of genus g over $\mathbf{F}_q$. Then*
(i) $Z_X(T)$ satisfies the functional equation:

$$Z_X(\frac{1}{qT}) = q^{1-g}T^{2-2g}Z_X(T).$$

(ii) On has that

$$Z_X(T) = \frac{P_X(T)}{(1-T)(1-qT)}.$$

where $P_X(T) \in \mathbf{Z}[T]$ is a polynomial of degree $2g$.

Proof. Let δ denote the gcd of the degrees of all divisors in $\mathrm{Div}_k(X)$ or, equivalently, of all effective divisors of X. We have that δ divides $2g-2 = \deg(\kappa)$. We write $\mathrm{Pic}^n(X)$ for the set of divisor classes of degree n and we let

$$Z_X(T) = \sum_{n=0}^{\infty} \#\{D \in \mathrm{Div}(X) \colon D \text{ is effective and } \deg(D) = n\} \cdot T^n \ = \sum_{n=0}^{\infty} b_n T^n.$$

Clearly

$$b_n = \sum_{D \in \mathrm{Pic}^n(X)} \frac{q^{\dim L(D)} - 1}{q-1}.$$

Now either $\mathrm{Pic}^n(X)$ is empty or it is in one-to-one correspondence to $\mathrm{Pic}^0(X)$ via $D \mapsto D - D_0$ for some fixed $D_0 \in \mathrm{Pic}^n(X)$. Since there are effective divisors D with $L(D) > 0$ it follows from the above formula for b_n that every $\mathrm{Pic}^n(X)$ is finite. We put

$$h = \#\mathrm{Pic}^0(X).$$

By Cor. 3.7 to the Riemann-Roch theorem we have for divisors D with $\deg(D) = n > 2g-2$ that $\dim_{\mathbf{F}_q} L(D) = n+1-g$. It follows that for n a multiple of δ we have that

$$b_n = h\frac{q^{n+1-g} - 1}{q-1}.$$

So

$$Z_X(T) = \sum_{n=0}^{2g-2} b_n T^n \ + \sum_{n > (2g-2)/\delta} h\frac{q^{n\delta+1-g} - 1}{q-1} T^{\delta n}.$$

It is easy to sum the geometric series. One finds

$$Z_X(T) = \sum_{n=0}^{2g-2} b_n T^n \ + (\frac{q^{g-1+\delta}T^{2g-2+\delta}}{1-(qT)^\delta} - \frac{T^{2g-2+\delta}}{1-T^\delta})\frac{h}{q-1}.$$

So, $Z_X(T)$ is a rational function. It has poles of order 1 at $T^\delta = 1$ and at $T^\delta = q^{-\delta}$. Using the product expansion of the ζ-function it is very easy to show that for every positive integer m one has the relation

$$Z_{X/\mathbf{F}_{q^m}}(T^m) = \prod_{\zeta^m = 1} Z_{X/\mathbf{F}_q}(\zeta T)$$

where the product runs over all m-th roots of unity ζ. We apply this with $m=\delta$. Since $Z_X(T) = \sum_n b_{n\delta}T^{n\delta}$ we have that $Z_X(T) = Z_X(\zeta T)$ and hence $Z_{X/\mathbf{F}_{q^\delta}}(T^\delta) = Z_{X/\mathbf{F}_q}(T)^\delta$. Since $Z_{X/\mathbf{F}_{q^\delta}}$ has only poles of order 1, we conclude that $\delta = 1$. We now have

$$
\begin{aligned}
Z_X(T) &= \sum_{\substack{D\in \mathrm{Pic}(X)\\ 0\le \deg(D)\le 2g-2}} \frac{q^{\dim L(D)}-1}{q-1}T^{\deg(D)} - \frac{h}{q-1}\left(\frac{T^{2g-1}}{1-T} - \frac{q^gT^{2g-1}}{1-qT}\right)\\
&= \frac{1}{q-1}\left(\sum_{n=0}^{2g-2} T^n \sum_{\deg(D)=n} q^{\dim L(D)} + h\left(q^{1-g}\frac{(qT)^{2g-1}}{1-qT} - \frac{1}{1-T}\right)\right)
\end{aligned}
.
$$

The functional equation in (*i*) now follows at once from the theorem of Riemann-Roch. Part (*ii*) is obvious from the explicit expression for $Z_X(T)$ above. This proves the theorem.

To a curve of genus g over $\mathbf{F}_q$ we now associate $2g$ complex numbers ϕ_i: the reciprocal zeroes of the polynomial $P_X(T)$: we have that $P_X(T) = \prod_{i=1}^{2g}(1-\phi_i T)$. It follows easily from the above theorem that we can order the ϕ_i in such a way that $\phi_{g+i} = q/\phi_i$ for $1\le i\le g$.

Proposition (5.2). *Let X be a curve of genus g over $\mathbf{F}_q$. Let the $\phi_i\in\mathbf{C}$ be as introduced above. We have for every positive integer m that*

$$P_{X/\mathbf{F}_{q^m}}(T) = \prod_{i=1}^{2g}(1-\phi_i^m T)$$

and

$$\#X(\mathbf{F}_{q^m}) = q^m + 1 - \sum_{i=1}^{2g}\phi_i^m.$$

Proof. The first formula follows at once from the relation

$$Z_{X/\mathbf{F}_{q^m}}(T^m) = \prod_{\zeta^m=1} Z_{X/\mathbf{F}_q}(\zeta T)$$

where the product runs over all m-th roots of unity. We apply this with $m=\delta$ that we used in the proof above. The second formula follows by inspection of the linear coefficient of $Z_X(T)$ and the theorem above. This proves the proposition.

Exercises.

(5.A) Prove that $h = \#\mathrm{Pic}^0(X) = \prod_{i=1}^{2g}(1-\phi_i) = P_X(1)$.
(5.B) Compute the number of points on the curve $x^3+y^3+z^3=0$ over every finite extension of $\mathbf{F}_2$.
(5.C) Compute the ζ-function of the Klein curve $x^3y+y^3z+z^3x=0$ over $\mathbf{F}_2$. (Use the example in section 4)
(5.D) Compute the ζ-function of the hyperelliptic curve given by $y^2+y=x^5+1$ over $\mathbf{F}_2$.

6. Weil's Theorem.

We will prove the analogue of the Riemann hypothesis for ζ-functions associated to curves over finite fields. This was first proved by André Weil [35], who published his proof in 1948. This result and the program started by Weil to generalize it, has had a profound influence on the development of algebraic geometry in the second half of this century. It stimulated Grothendieck to develop his powerful theory of schemes. Using these techniques, a generalization of Weil's Theorem was proved in 1973 by Deligne. The classical Riemann hypothesis remains unproven, however.

Let X be a curve over $\mathbf{F}_q$ of genus g. Let p denote the characteristic of $\mathbf{F}_q$. Let ϕ_i for $i = 1, 2, \ldots, 2g$ denote the reciprocal zeroes of the polynomial $Z_X(T)$ introduced in Theorem 5.1. Our proof follows Bombieri's exposé [3] of Stepanov's proof. The idea is to construct a rational function f on X that vanishes to high order m at the $\mathbf{F}_q$-rational points of X except possibly one but whose number of poles can be bounded. By Cor. 3.2 we will then have that

$$m(\#X(\mathbf{F}_q) - 1) \leq \#\{\text{ zeroes of } f \text{ }\} = \#\{\text{ poles of } f \text{ }\}$$

which implies a bound on $\#X(\mathbf{F}_q)$. Using the functional equation we can then prove that $|\phi_i| = \sqrt{q}$ which is easily seen to be equivalent to the analogue of the Riemann hypothesis.

We let $\pi: X \longrightarrow X$ denote the Frobenius morphism which for the function fields is just the inclusion $k(X) \hookleftarrow k(X)^q \cong k(X)$. The Frobenius morphism acts on the points $X(\bar{k})$ in a natural way: it "raises the coordinates of the points to the q-th power". By Galois theory, the fixed points of π^m make up the subset $X(\mathbf{F}_{q^m})$.

Proposition (6.1). *If q is a square and $q > (g+1)^4$ then we have that*

$$\#X(\mathbf{F}_q) < q + 1 + (2g+1)\sqrt{q}.$$

Proof. We may assume that there is a point P in $X(\mathbf{F}_q)$. Otherwise the proposition is obvious. We will study the spaces $L(m(P))$ for X over $\bar{k}$. For a positive integer μ we let $L(k(P))^{p^\mu}$ denote the space of functions f^{p^μ} where $f \in L(m(P))$. By $L(k(P))^\pi$ we denote the space of functions $f \cdot \pi = f(X^q)$ where $f \in L(m(P))$. We first prove a lemma.

Lemma (6.2). *Let n and m be two positive integers and let μ be a positive integer with $np^\mu < q$. The natural homomorphism*

$$L(n(P))^{p^\mu} \otimes_{\bar{k}} L(m(P))^\pi \longrightarrow L(n(P))^{p^\mu} L(m(P))^\pi$$

is an isomorphism.

Proof. By Prop.3.3(*ii*) we have that $\dim_{\bar{k}} L((\nu+1)(P)) \leq \dim_{\bar{k}} L(\nu(P)) + 1$ and we conclude that there is a $\bar{k}$-basis $e_1, e_2, \ldots, e_r$ of $L(m(P))$ such that $\mathrm{ord}_P(e_1) < \mathrm{ord}_P(e_2) < \ldots < \mathrm{ord}_P(e_r)$. The homomorphism in the statement of the lemma is clearly surjective. We must show it is injective i.e. we must show that whenever $\sum_{i=1}^r g_i^{p^\mu}(e_i \cdot \pi) = 0$ for certain $g_i \in L(n(P))$ then all the g_i are 0 themselves. So assume that i_0 is the smallest integer in $\{1, 2, \ldots, r\}$ for which $g_{i_0} \neq 0$. We have

$$\begin{aligned}
\mathrm{ord}_P(g_{i_0}^{p^\mu}(e_i \cdot \pi) &= \mathrm{ord}_P(-\sum_{i>i_0}^{r} g_i^{p^\mu}(e_i \cdot \pi)) \\
&\geq \min_{i>i_0} \mathrm{ord}(g_i^{p^\mu}(e_i \cdot \pi)) \\
&\geq -np^\mu + q\,\mathrm{ord}_P(e_{i_0+1})
\end{aligned}$$

Therefore

$$
\begin{aligned}
p^\mu \mathrm{ord}_P(g_{i_0}) &\geq -np^\mu + q(\mathrm{ord}_P(e_{i_0+1}) - \mathrm{ord}_P(e_{i_0})) \\
&\geq -np^\mu + q > 0
\end{aligned}
$$

and we see that g_{i_0} vanishes at P. But g_{i_0} being an element of $L(n(P))$ has no poles outside P. Therefore it must be constant and hence 0. This proves the Lemma.

Since $\dim_{\bar{k}} L(n(P))^{p^\mu} = \dim_{\bar{k}} L(n(P))$ and $\dim_{\bar{k}} L(m(P))^\pi = \dim_{\bar{k}} L(m(P))$ we obtain, as a consequence, that

$$
\dim_{\bar{k}} L(n(P))^{p^\mu} L(m(P))^\pi = \dim_{\bar{k}} L(n(P)) \times \dim_{\bar{k}} L(m(P)).
$$

We have a natural well-defined homomorphism $L(n(P))^{p^\mu} L(m(P))^\pi \to L(n(P))^{p^\mu} \otimes L(m(P))^\pi \to L(n(P))^{p^\mu} \otimes L(m(P)) \to L(n(P))^{p^\mu} L(m(P))$ mapping $f = \sum_i g_i^{p^\mu}(e_i \cdot \pi)$ to $\sum_i g_i(e_i \cdot \pi)$ because the first arrow is the inverse of the isomorphism of the lemma, while the second arrow is the inverse of an obvious isomorphism. Since $L(n(P))^{p^\mu} L(m(P)) \subset L((np^\mu + m)(P))$ we obtain a homomorphism

$$
\theta : L(n(P))^{p^\mu} L(m(P))^\pi \longrightarrow L((np^\mu + m)(P)).
$$

By Prop 3.3(iv) we have that $\dim L(n(P)) \geq n + 1 - g$ and $\dim L(m(P)) \geq m + 1 - g$. If n and m are at least g, we have by Cor.3.7(ii) to the Riemann-Roch theorem that $\dim L((np^\mu + m)(P)) = np^\mu + m + 1 - g$. Therefore we find that, whenever $m, n \geq g$,

$$
\begin{aligned}
\dim \ker_{\bar{k}}(\theta) &\geq \dim_{\bar{k}} L(n(P)) \times \dim_{\bar{k}} L(m(P)) - \dim_{\bar{k}} L((np^\mu + m)(P)) \\
&\geq (n + 1 - g)(m + 1 - g) - (np^\mu + m + 1 - g).
\end{aligned}
$$

Suppose that $f = \sum_i g_i^{p^\mu}(e_i \cdot \pi)$ is a function in $\ker(\theta)$ and $Q \in X(\mathbf{F}_q)$ is a point not equal to P. We have that

$$
f(Q) = \sum_i g_i^{p^\mu}(Q)(e_i \cdot \pi)(Q) = \sum_i g_i^{p^\mu}(Q)(e_i)(Q) = (\theta f)(Q) = 0.
$$

We conclude that f vanishes at every point of $X(\mathbf{F}_q)$ except P. But, since f is a p^μ-th power, f has zeroes of order at least p^μ at these points. Therefore any f in $\ker(\theta)$ has at least $p^\mu(\#X(\mathbf{F}_q) - 1)$ zeroes. Since

$$
f \in L(n(P))^{p^\mu} L(m(P))^\pi \subset L((np^\mu + mq)(P))
$$

such a function f has, on the other hand, at most $np^\mu + mq$ poles. We conclude that if the assumptions $np^\mu < q$ and $n, m \geq g$ are fulfilled and if $\ker \theta > 0$ i.e. if

$$
(n + 1 - g)(m + 1 - g) > (np^\mu + m + 1 - g)
$$

then one has that

$$
\#X(\mathbf{F}_q) \leq n + \frac{mq}{p^\mu} + 1.
$$

We will choose the parameters p^μ, n and m as follows: $p^\mu = \sqrt{q}$, $m = \sqrt{q} + 2g$ and $n = [\frac{g}{g+1}\sqrt{q}] + g + 1$. Since $q > (g+1)^4$, it is readily verified that with these choices, the assumptions are fulfilled. This proves the Proposition.

Theorem (6.3). *Let X be a curve of genus g over a finite field $\mathbf{F}_q$. Assume that q is a square and that $q > (g+1)^4$. Then for k large enough*

$$
\#X(\mathbf{F}_{q^k}) = q^k + O(q^{k/2})
$$

where the O-symbol depends only on X over $\overline{\mathbf{F}}_q$.

Proof. Use any non-zero function in $\mathbf{F}_q(X)$ to construct a morphism $X \longrightarrow \mathbf{P}^1$. The corresponding extension of function fields $\mathbf{F}_q(X) \hookleftarrow \mathbf{F}_q(\mathbf{P}^1)$ is not necessarily Galois. Let L denote the Galois closure of this extension. The field L is the function field of a smooth irreducible curve Y of genus g_Y over $\mathbf{F}_q$. Let G denote $\mathrm{Gal}(L/\mathbf{F}_q(\mathbf{P}^1))$ and let H denote the subgroup $\mathrm{Gal}(L/\mathbf{F}_q(X))$.

Let A denote the set of unramified points $P \in Y(\bar{\mathbf{F}}_q)$ whose image in $\mathbf{P}^1$ is rational over $\mathbf{F}_{q^k}$. Since $\mathbf{P}^1$ has q^k+1 rational points over $\mathbf{F}_{q^k}$, it is immediate that

$$\#A = \#G(q^k+1)\#G + O(1)$$

where the O-symbol is due to the finitely many ramification points of the covering Y of $\mathbf{P}^1$. It is independent of the degree k.

For every unramified point $P \in A$ the point $\pi(P)$ maps to the same point in $\mathbf{P}^1$ as P does. Therefore there is a unique $\sigma \in G$ such that $\pi(P) = \sigma(P)$. The automorphism σ is called the *Frobenius substitution* of P. For every $\sigma \in G$ put

$$A_\sigma = \{P \in A\colon \pi(P) = \sigma(P)\}.$$

The set A is a disjoint union of the A_σ's. Since $q^k > g_Y$ for k large enough, we can argue as in the proof of Proposition 6.2, but now with $P \in A_\sigma$. We now have tha map

$$\theta_\sigma\colon L(n(\sigma^{-1}P))^{p^\mu} L(m(P))^\pi \to L(n(\sigma^{-1}P))^{p^\mu} L(m(P))^\sigma \to L((np^\mu+m)(\sigma^{-1}P)).$$

We obtain easily that

$$\#A_\sigma \le q^k + 1 + (2g_Y+1)q^{k/2}$$

and hence

$$\#A = \sum_{\sigma\in G} \#A_\sigma = (q^k+1)\#G + O(q^{k/2}).$$

Combining this with the formula for $\#A$ that we deduced above, we find that for each $\sigma \in G$

$$\#A_\sigma = q^k + O(q^{k/2}).$$

By Galois theory we have that

$$\bigcup_{\sigma\in H} A_\sigma = \{P \in Y\colon \text{ the image of } P \text{ in } X \text{ is rational over } \mathbf{F}_{q^k}\}$$

Therefore

$$\begin{aligned}\sum_{\sigma\in H} \#A_\sigma &= \#H\#X(\mathbf{F}_q) + O(1)\\ &= \#Hq^k + O(q^{k/2})\end{aligned}$$

and the result follows at once.

Theorem (6.4). *(A. Weil, 1948)* Let X be a curve of genus g over a finite field $\mathbf{F}_q$. Then the reciprocal roots $\phi \in \mathbf{C}$ of the function $Z_X(T)$ satisfy

$$|\phi| = \sqrt{q}.$$

Proof. By Proposition 5.2, it suffices to give the proof for a power of q. We will call this power q again and choose it so large that the condition of Proposition 6.1 is satisfied: q is a square exceeding $(g+1)^4$. We deduce from these propositions that for large enough k we have that

$$\#X(\mathbf{F}_{q^k}) = q^k + 1 + O(q^{k/2}).$$

and therefore, with the usual notation, that

$$\sum_{j=1}^{2g} \phi_j^k = O(q^{k/2}).$$

This implies that the function $f(z) = \sum_{i=1}^{2g}(1-\phi_i z)^{-1}$ has a radius of convergence at least as large as $q^{-1/2}$. Therefore we have that $|\phi_i| \leq \sqrt{q}$ for $1 \leq i \leq 2g$. The theorem now follows from the functional equation (Theorem 5.1(*ii*)) satisfied by $Z_X(T)$: when ϕ is a reciprocal root, so is q/ϕ and it follows that $|\phi| = \sqrt{q}$ for all ϕ as required.

Corollary (6.5). *Let X be a curve of genus g over $\mathbf{F}_q$. Then*

$$|q^m + 1 - \#X(\mathbf{F}_{q^m})| \leq 2gq^{m/2}.$$

Proof. This is immediate from Proposition 5.2 and Theorem 6.4.

Exercises.

(6.A) Show that a curve of genus 0 over a finite field is isomorphic to $\mathbf{P}^1$.

(6.B) Show that a curve of genus 1 over a finite field always has a point rational over that field.

(6.C) Prove the Riemann Hypothesis for the ζ-function $\zeta_X(s)$ of a curve X over a finite field: If $\zeta_X(s) = 0$ then the real part of s is $1/2$.

(6.D) Let X be the projective curve given by $y^2 + xy = x^3 + x$ over $\mathbf{F}_2$. Show that its genus is 1 and compute the zeroes of the ζ-function $\zeta_X(s)$.

(6.E) Let p be a prime and let $\chi\colon \mathbf{F}_p^* \to \{\pm 1\}$ be the quadratic character mod p: $\chi(x) = 1$ whenever x is a square in $\mathbf{F}_p^*$ and $\chi(x) = -1$ otherwise. By convention we put $\chi(0) = 0$. Show that for every $A, B \in \mathbf{Z}$ one has that $|\sum_{x\in\mathbf{F}_p} \chi(x^3 + Ax + B)| < 2\sqrt{p}$.

(6.F) Let $\mathbf{F}_q$ be a finite extension of degree m of $\mathbf{F}_2$ and let $\mathrm{Tr}\colon \mathbf{F}_q \to \mathbf{F}_2$ denote the Trace map: $\alpha \mapsto \alpha + \alpha^2 + \ldots + \alpha^{2^{m-1}}$. Show that for every $\alpha, \beta \in \mathbf{F}_q$ one has that $\sum_{x\in\mathbf{F}_q} |(-1)^{\mathrm{Tr}(x^3+\alpha x+\beta)}| < 2\sqrt{q}$.

(6.G) Using the same notation as in Exercise 6.F, let $f(X) \in \mathbf{F}_q[X]$ have odd degree r. Then one has that $\sum_{x\in\mathbf{F}_q} |(-1)^{\mathrm{Tr}(f(x))}| < (r-1)\sqrt{q}$

(6.H) Using the same notation as in Exercise 6.F, consider the map $\theta\colon \mathbf{F}_q \times \mathbf{F}_q \longrightarrow V$ defined by $\theta(\alpha,\beta) = (\mathrm{Tr}(\alpha x + \beta x^{-1})_{x\in\mathbf{F}_q}$. Here V is the $\mathbf{F}_2$-vectorspace with basis indexed by the elements of $\mathbf{F}_q$. Show that θ is injective. View the image of θ as a code in V. Estimate its parameters (cf. [18]).

7. The Theorem of Drinfeld and Vlădut.

Since asymptotically good Goppa-codes can be constructed using families of curves over finite fields that have many points with respect to their genus, it was natural to try and find such curves (cf. [17]). It was soon found, rather surprisingly, that the bounds on the number of points that follow from Weil's theorem are not sharp when the genus is very large. In this section we will prove the asymptotic estimate by Drinfeld and Vlădut [7] and mention an unpublished theorem due to Oesterlé [22] that is somewhat more precise.

Let X be a curve of genus g over $\mathbf{F}_q$. We will be interested in the ratio $\#X(\mathbf{F}_q)/g$. Curves for which this ratio is large can be used to construct good codes. First we will look a little bit at curves with small genus and then we will study the behaviour of this ratio as the finite field $\mathbf{F}_q$ is fixed and the genus tends to infinity. We will only consider the case $\mathbf{F}_2$. It is the most interesting case for coding theory and the easiest field when one wishes to do computations.

By Exercise 6.A the only curve, upto isomorphism, of genus 0 is $\mathbf{P}^1$. It has 3 points and the bound $\#X(\mathbf{F}_q) \leq q+1+2g\sqrt{q}$ of Weil is sharp. The curve $y^2+y=x^3+x$ is a curve of genus 1. It has been used in section 4 to construct a simple Goppa code. It has 5 points rational over $\mathbf{F}_2$. This is best possible since the Weil-bound is $3+2\sqrt{2}$ here. Since all curves of genus 2 are hyperelliptic (Ex.3.C) they can have at most 6 points over $\mathbf{F}_2$. This is best possible as the example $y^2+y=(x^2+x)/(x^3+x+1)$ shows. The Weil bound is not sharp; it is only $3+4\sqrt{2}\approx 8.65$ here. For curves of genus 3 the Weil bound is approximately 11.48. However, by Exercise 3.D a curve of genus 3 is either hyperelliptic or isomorphic to a smooth quartic curve in $\mathbf{P}^2$. Over $\mathbf{F}_2$ the projective plane has only 7 points! So curves of genus 3 cannot have more than 7 points. This is the best possible bound: the curve

$$x^3y+y^3z+z^3x+x^2y^2+x^2z^2+y^2z^2+x^2yz+xy^2z=0$$

passes through all seven rational points of $\mathbf{P}^2$.

According to the remark in Ex.3.E a curve of genus 4 is either hyperelliptic or can be realized as the intersection of a smooth irreducible quadric and an irreducible cubic surface in $\mathbf{P}^3$. One can show that quadrics in $\mathbf{P}^2$ can have at most 9 points over $\mathbf{F}_2$. This is not sharp as we will see later, but it is a lot better than the Weil bound which is approximately 14.31 for curves of genus 4. The curve with affine equation

$$x^2y^3+x^3y^2+xy^3+x^3y+x^2y^2+x^2+y^2+1=0$$

in $\mathbf{P}^1\times\mathbf{P}^1$ passes through every point except $(0,0)$. This is best possible. It follows from Oesterlé's theorem below that a curve of genus 4 over $\mathbf{F}_2$ can have at most 8 rational points.

So we see that the bounds that follow from the Riemann hypothesis are by no means sharp when the genus becomes somewhat large. Next we'll investigate what happens when the genus becomes *very* large. Put

$$A(q)=\limsup_{g\to\infty}\frac{\#X(\mathbf{F}_q)}{g}.$$

It follows from Weil's Theorem that

$$A(q)\leq 2\sqrt{q}.$$

It is rather easy to improve this somewhat.

First we discuss an improvement due to Serre [24].

Theorem (7.1). *For every curve X of genus g over $\mathbf{F}_q$ one has that*

$$\#X(\mathbf{F}_q) \leq q + 1 + [2\sqrt{q}]g.$$

Proof. We have that $\#X(\mathbf{F}_q) = q + 1 - \sum_{i=1}^{g}(\phi_i + \bar{\phi}_i)$ where $|\phi_i| = \sqrt{q}$ for $1 \leq i \leq g$. Put

$$x_i = [2\sqrt{q}] + 1 + \phi_i + \bar{\phi}_i \qquad \text{for } 1 \leq i \leq g.$$

By Theorem 6.4 the numbers x_i are totally positive algebraic integers. Therefore their product is at least 1. It follows from the arithmetic-geometric-mean inequality that

$$\frac{1}{g}\sum_{i=1}^{g} x_i \geq (\prod_{i=1}^{g} x_i)^{1/g} \geq 1.$$

So we have that $\sum_i x_i \geq g$ which is easily seen to imply the result.

An immediate corollary is that

$$A(q) \leq [2\sqrt{q}].$$

A second improvement is due to Ihara [13]: Consider X also over $\mathbf{F}_{q^2}$. We have that

$$\begin{aligned}
\#X(\mathbf{F}_q) \leq \#X(\mathbf{F}_{q^2}) &= q^2 + 1 - \sum_{i=1}^{g}(\phi_i^2 + \bar{\phi}_i^2) \\
&= q^2 + 1 + 2qg - \sum_{i=1}^{g} t_i^2 \\
&\leq q^2 + 1 + 2qg - \frac{1}{g}(\sum_{i=1}^{g} t_i)^2
\end{aligned}$$

where $t_i = \phi_i + \bar{\phi}_i$. The last inequality is the inequality of Cauchy-Schwartz. We conclude that

$$\#X(\mathbf{F}_q) \leq q^2 + 1 + 2qg - \frac{1}{g}(\#X(\mathbf{F}_q) - q - 1)^2$$

which easily implies that

$$A(q) \leq \sqrt{2q} - \frac{1}{2}.$$

Now we come to the best known estimate due to Drinfeld and Vlădut [7]. The proof is an extension of Ihara's argument. It involves a consideration of all finite extensions of $\mathbf{F}_q$. We will write a_d for the number of places of X of degree d. Equivalently, a_d is the number of points in $X(\overline{\mathbf{F}}_q)$ of degree d upto Galois conjugacy. So, we have that $\#X(\mathbf{F}_{q^m}) = \sum_{d|m} da_d$.

Let $\Psi(T) = \sum_{n=1}^{\infty} c_n T^n$ be a polynomial with non-negative coefficients c_n for which

$$\Psi(t) + \Psi(\bar{t}) + 1 \geq 0 \qquad \text{for all } t \in \mathbf{C} \text{ with } |t| = 1.$$

By $\Psi_d(T)$ we denote the polynomial $\sum_{n \equiv 0 \pmod{d}} c_n T^n$.

Theorem (7.2). *Let X be a curve over $\mathbf{F}_q$ of genus g and let $\Psi(T)$ be a polynomial as above. We have*

$$\sum_{d\geq 1} da_d\Psi_d(q^{-1/2}) \leq g + \Psi(q^{1/2}) + \Psi(q^{-1/2}).$$

Proof. As usual, let ϕ_j denote the reciprocal zeroes of the function $Z_X(T)$. By the Riemann hypothesis we have that $\phi_j = \sqrt{q}e^{i\theta_j}$ with $\theta_j \in \mathbf{R}$. By Theorem 5.1(*i*) We can order the ϕ_j in such a way that $\theta_{g+j} = -\theta_j$. We have that

$$\#X(\mathbf{F}_{q^n}) = q^n + 1 - q^{n/2}\sum_{j=1}^{g}(e^{in\theta_j} + e^{-in\theta_j}).$$

So

$$\begin{aligned}
0 \leq \sum_{j=1}^{g}(\Psi(e^{in\theta_j}) + \Psi(e^{-in\theta_j}) + 1) &= g + \sum_{j,n} c_n(e^{in\theta_j} + e^{-in\theta_j}) \\
&= g + \sum_{n\geq 1} q^{-n/2}c_n(q^n + 1 - \#X(\mathbf{F}_{q^n})) \\
&= g + \Psi(q^{1/2}) + \Psi(q^{-1/2}) - \sum_{d\geq 1}\sum_{d|n} q^{-n/2}da_dc_n
\end{aligned}$$

and the inequality follows.

The theorem clearly implies that

$$(\#X(\mathbf{F}_q) - 1)\Psi(q^{-1/2}) \leq g + \Psi(q^{1/2})$$

and therefore that

$$A(q) \leq \frac{1}{\Psi(q^{-1/2})}.$$

We must choose our polynomial $\Psi(T) = \sum_{n=1}^{\infty} c_nT^n$ in order to get an estimate for $A(q)$. The larger we can take the c_n, the better the estimate will be. However, we have that

$$0 \leq \frac{1}{\pi}\int_0^{2\pi}(1 + \Psi(e^{i\theta}) + \Psi(-e^{i\theta}))(1 - \cos n\theta)d\theta = 1 - c_n$$

So $c_n \leq 1$ and we cannot choose all $c_n = 1$ since they should be zero for large n. We will, instead, choose a sequence of polynomials Ψ whose coefficients approach 1 viz:

$$1 + \Psi(T) + \Psi(T^{-1}) = \frac{1}{N+1}(1 + T + \ldots + T^N)(1 + T^{-1} + \ldots + T^{-N})$$

It is easy to see that this gives a sequence of bounds on $A(q)$, the limit of which is

$$A(q) \leq \sqrt{q} - 1.$$

One can show that this estimate is sharp when q is a square [13,29]. The curves that meet the bound are Shimura curves. In the case $q = p^2$ the Shimura curves are the modular curves $X_0(\ell)$ and the proof then briefly runs as follows:

For every supersingular j-invariant j in characteristic p, there exists over $\mathbf{F}_{p^2}$ an elliptic curve with j-invariant j whose Frobenius endomorphism can be identified with an integer. Therefore all the finite cyclic subgroups of its points are defined over $\mathbf{F}_{p^2}$. Consider the modular curves $X_0(\ell)$ for prime $\ell \equiv 11 \pmod{12}$ over the field $\mathbf{F}_{p^2}$. The curve $X_0(\ell)$ parametrizes elliptic curves together with an isogeny of degree ℓ. The genus of this curves is $(\ell+1)/12$ and its degree over the j-line is $(p-1)/2$ and by the above, all the points lying over the supersingular j-invariants are rational over $\mathbf{F}_{p^2}$. Since there are roughly $p/12$ supersingular j-invariants in characteristic p, which are all in the field $\mathbf{F}_{p^2}$, there are, upto a slight error due to the cusps and the j-values 0 and 1728, at least $(\ell+1)\times(p-1)/12$ rational points on $X_0(\ell)$. We see that the ratio of the number of point on $X_0(\ell)$ to its genus is approximately $p-1$. It is not so difficult to check that it actually approaches $p-1$ as ℓ tends to infinity.

For finite fields $\mathbf{F}_q$ for which q is not a square, much less is known. Th. Zink [37] has shown that for prime p, one has that $A(p^3) \geq 2(p^2-1)/(p+2)$. Serre [24], using an infinite class field tower of the function field of a hyperelliptic curve over $\mathbf{F}_q$ showed that there exists a positive constant $c \in \mathbf{R}$ such that $A(q) > c\log q$. For $q=2$ one can show that $2/9 \leq A(2) \leq \sqrt{2}-1$, so $0.222 < A(2) < 0.415$. These results are proved in section 9.

It is possible to do the above estimates in such a way that one obtains information for "finite" values of g as well. This was done by Oesterlé in 1982. We state his result without proof.

Theorem (7.3). *Let X be a curve of genus g over $\mathbf{F}_q$ with $L+1$ points rational over $\mathbf{F}_q$. Then*

$$g \geq \sup_{\Psi}\{L\Psi(q^{-1/2}) - \Psi(q^{1/2})\} \geq \frac{(L-1)\sqrt{q}\cos\theta_0 + q - L}{q+1-2\sqrt{q}\cos\theta_0}$$

Moreover, if $q \geq 3$ the second inequality is actually an equality. Here the supremum is taken over the polynomials $\Psi(T)$ with non-negative coefficients satisfying $\Psi(t)+\Psi(\bar{t})+1 \geq 0$ for $t \in \mathbf{C}, |t|=1$. The value of θ_0 is defined as follows: Let m be the unique integer for which $\sqrt{q}^m < L \leq \sqrt{q}^{m+1}$. Put

$$u = \frac{\sqrt{q}^{m+1} - L}{L\sqrt{q} - \sqrt{q}^m} \in [0,1)$$

and let θ_0 denote the unique solution $\theta = \theta_0$ of the equation $\cos\frac{m+1}{2}\theta + u\cos\frac{m-1}{2}\theta = 0$ in the interval $[\frac{\pi}{m+1}, \frac{\pi}{m})$.

See Table 8.8 for the bounds one obtains from Theorem 7.3 for small values of g in the case $q=2$.

Exercises.

(7.A) We use the notation of Theorem 7.2. Show that the choice $c_1 = 1/2$ and $c_n = 0$ for $n > 1$ leads to the Weil bound: $\#X(\mathbf{F}_q) \leq q+1+2g\sqrt{q}$.

(7.B) Verify the bounds given in table 8.8.

(7.C) Show that the inequality in Theorem 7.3 is equivalent to

$$\frac{g-1}{L+1} \geq \mathrm{Re}\left(\frac{1}{\sqrt{q}e^{i\theta_0}-1}\right).$$

Deduce that $A(q) \leq \sqrt{q}-1$.

(7.D) (cf. [24]) Let $q=2$. We will choose Ψ such that $1+\Psi(t)+\Psi(t^{-1}) = c^{-1}(1+d_1(t+t^{-1})+\ldots+d_m(t^m+t^{-m}))^2$ with $d_i \geq 0$ and $c = 1+2d_1^2+\ldots+2d_m^2$. Show that Theorem 7.1 with $m=3$ and the choice $d_1=1$, $d_2=0.7$ and $d_3=0.2$ gives rise to the bound $\#X(\mathbf{F}_2) \leq 0.826g+5.346$. Compute which bounds it gives for $2 \leq g \leq 12$. Compute the bound one gets with $m=5$ and the choice $d_1=1$, $d_2=0.8$, $d_3=0.6$, $d_4=0.4$ and $d_5=0.1$.

8. Class Field Theory.

Following J.-P. Serre, we use class field theory over function fields [20] to construct curves that have many points with respect to their genus. Examples of rather high genus that meet the bounds of Oesterlé are presented. Most of the results in this section are due to Serre who lectured on this in October 1982 at the Collège de France [22,23,24,25]. For the cohomology theory of groups that we will use see [4,21] and see [1,14] for a cohomological approach to class field theory of number fields as well as function fields.

Let X be a curve over a finite field $\mathbf{F}_q$. Recall that we only consider projective, absolutely irreducible non-singular curves. Let $K = \mathbf{F}_q(X)$ denote the function field of X. A *place* of X or K is a $\mathrm{Gal}(\overline{\mathbf{F}}_q/\mathbf{F}_q)$-conjugacy class of points in $X(\overline{\mathbf{F}}_q)$. For every place v we let O_v denote the completion of any of the local rings O_P where P is a point "of" v. This is a discrete valuation ring. By K_v we denote its quotient field.

The ring of *adèles* $\mathbf{A}_K$ is the restricted product of the fields K_v with respect to the O_v:

$$\mathbf{A}_K = \{(x_v)_v : x_v \in K_v \text{ and } x_v \in O_v \text{ for almost all } v\}.$$

The unit group $\mathbf{A}_K^*$ is called the group of *idèles* of K. It contains K^* as a subgroup. The quotient group $C_K = \mathbf{A}_K^*/K^*$ is called the idèle class group of K. The idèles admit a natural surjective homomorphism $\mathbf{A}_K^* \longrightarrow \mathrm{Div}(X)$ induced by the valuation maps. The kernel is denoted by U. It consists of the idéles that have trivial valuation at all places. The intersection of U and K^* inside $\mathbf{A}_K^*$ is precisely equal to $\mathbf{F}_q^*$. Writing P for the group of principal divisors and Q for the quotient $U/\mathbf{F}_q^*$ it is easy to see that we have the following diagram with exact rows and columns:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathbf{F}_q^* & \longrightarrow & K^* & \longrightarrow & P & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & U & \longrightarrow & \mathbf{A}_K^* & \longrightarrow & \mathrm{Div}(X) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & Q & \longrightarrow & C_K & \longrightarrow & \mathrm{Pic}(X) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

Class field theory gives a description of the idèle class group C_K. It relates its open subgroups to abelian extensions of the function field K or, equivalently, to abelian coverings of the curve X. The main result is contained in the following theorem.

Theorem (8.1). *Let X be a curve over $\mathbf{F}_q$ with function field K.*

(i) (Artin-Reciprocity.) For every finite extension L of K with $\pi = \mathrm{Gal}(L/K)$ abelian, there are canonical isomorphisms

$$\widehat{H}^q(\pi, C_L) \cong \widehat{H}^{q-2}(\pi, \mathbf{Z}).$$

(ii) (Existence Theorem.) For every subgroup of finite index M of C_K there exists a unique finite abelian extension L for which $N(C_L) = M$. Here N denotes the norm map from L to K.

Proof. For the proof of this theorem we refer to standard texts on class field theory [1,4,14]

Applying Theorem 8.1(i) with $q = 0$ to a finite extension L over K with $\pi = \mathrm{Gal}(L/K)$ abelian, one obtains the classical reciprocity map

$$C_K/N(C_L) \xrightarrow{\cong} \pi.$$

A place v of K is unramified in L if and only if the unit group O_v^* is contained in $N(C_L)$. In this case we obtain a map $\mathbf{Z} \cong K_v^*/O_v^* \longrightarrow C_K/N(C_L) \cong \pi$. The image of 1 is called the Frobenius element ϕ_v of v in π. It determines the splitting behavior of v in the extension L over K. For instance, v is totally split if and only if its Frobenius element ϕ_v is trivial i.e. if and only if $K_v^*/O_v^* \subset N(C_L)$.

We will study some special abelian extensions of a function field K, the so-called *ray class fields* . Let D be an effective divisor which is defined over $\mathbf{F}_q$. So $D = \sum_v n_v v$ where the sum runs over the places v of K, the n_v are non-negative and only finitely many of them satisfy $n_v > 0$. Let $U_D = \{(x_v)_v \in U\colon x_v \equiv 1 \pmod{t_v^{n_v}}\}$; here t_v denotes a uniformizing element at v. The group $U_D/\mathbf{F}_q^*$ is a subgroup of the idèle class group C_K. The abelian extensions of K corresponding to the finite quotients of C_K/U_D are called *ray class fields.* Let L be a finite extension of K and let $M \subset C_K$ be the *ray class group* that corresponds to it according to Theorem 8.1.(ii). There is an effective divisor D, minimal with respect to division, such that $U_D/\mathbf{F}_q^* \subset M$. This is called the *conductor* of L

For $D = (0)$ the group $U_D/\mathbf{F}_q^*$ is just $U/\mathbf{F}_q^*$ and the quotient C_K/U_D is precisely the Picard group $\mathrm{Pic}(X)$. For arbitrary D we have an exact sequence

$$0 \longrightarrow U/U_D \longrightarrow C_K/U_D \longrightarrow \mathrm{Pic}(X) \longrightarrow 0.$$

For the first group we have the following explicit description:

$$U/U_D \cong \bigoplus_{v \text{ in } D} \mathbf{F}_q[[t_v]]^*/\{x : x \equiv 1 \pmod{t_v^{n_v}}).$$

From the exact sequence

$$0 \longrightarrow \mathrm{Pic}^0(X) \longrightarrow \mathrm{Pic}(X) \xrightarrow{\deg} \mathbf{Z} \longrightarrow 0$$

we see that for each degree n there is a ray class field corresponding to the cyclic quotient of order n of $\mathbf{Z}$. This is the constant field extension $\mathbf{F}_{q^n}(X)$. In general the degree of the field of constants of a finite abelian extension L of K corresponding to $M \subset C_K$ is equal to the index of $\deg(M)$ in $\mathbf{Z}$.

Finally we mention *Dirichlet characters* i.e. continuous homomorphisms $\chi : \mathrm{Gal}(\overline{K}/K) \longrightarrow \mathbf{C}^*$. Here $\overline{K}$ denotes a separable closure of K. A character χ has a finite image and the fixed field of $\ker \chi$ is a cyclic extension L of K. We define the *conductor* of χ to be the conductor of L. We recall the following result, very useful to calculate the genus of an abelian covering.

Proposition (8.2). *(Führerproduktdiskriminantformel) Let X be a curve of genus g_X over $\mathbf{F}_q$ and let $f : Y \longrightarrow X$ be a covering of genus g_Y. Suppose that the corresponding extension of function fields has a finite abelian Galois group G. Then*

$$2g_Y - 2 = \deg(f)(2g_X - 2) + \sum_{\chi} \deg(\mathrm{cond}\chi).$$

Here the product runs over the characters χ of G.

Proof. See [35.Ch.V].

In the remainder of this section we will use class field theory to show the existence of curves over $\mathbf{F}_2$ with certain properties.

Example (8.3). Consider $\mathbf{P}^1$ over $\mathbf{F}_2$. Let K denotes its function field and let P be a point of degree 3 on $\mathbf{P}^1$. Consider the ray class fields of conductor P. Since $\mathrm{Pic}(\mathbf{P}^1) = \mathbf{Z}$ and $\mathbf{F}_2^* = \{1\}$, we have an exact sequence

$$0 \longrightarrow \mathbf{F}_8^* \longrightarrow C_K/U_P \longrightarrow \mathbf{Z} \longrightarrow 0.$$

Next we pick an $\mathbf{F}_2$-rational point Q on $\mathbf{P}^1$ and we let L be the fixed field of the Frobenius element of Q and let X be the curve corresponding to it. Clearly L has degree $2^3 - 1 = 7$ over K. Its genus g is easily computed using Prop.8.5: $2g - 2 = -2 \cdot 2 + 6 \cdot 3$. Since all the points over Q on X are rational we see that we have found a curve over $\mathbf{F}_2$ of genus 3 and at least 7 rational points. It is rather easy to see that there are, in fact, precisely 7 rational points.

Example (8.4). Let (∞) be a rational point on $\mathbf{P}^1$ over $\mathbf{F}_2$ and let 0 and 1 denote the others. Let $D = 4(\infty)$. We have an exact sequence

$$0 \longrightarrow \mathbf{F}_2[[t]]^*/\{f\colon f \equiv 1 \pmod{t^4}\} \longrightarrow C_K/U_D \longrightarrow \mathbf{Z} \longrightarrow 0.$$

It follows easily that, as an abelian group, $C_K/U_D \cong \mathbf{Z} \oplus \mathbf{Z}/4\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z}$. Let L be the invariant field of the two Frobenius elements of the other two points 0 and 1. It is easy to see that the degree of L over the function field K of $\mathbf{P}^1$ is 2. The genus g of the curve E corresponding to L follows from $2g - 2 = -2 \cdot 2 + 4$. By construction the points 0 and 1 are split and (∞) is ramified in E over $\mathbf{P}^1$. We have found a curve over $\mathbf{F}_2$ of genus 1 with 5 rational points.

Example (8.5). This time we use the curve E of Example 8.3 as a basis. Its genus is 1 and it has five $\mathbf{F}_2$-rational points $P_1, P_2, \ldots, P_5$. Its class group $\mathrm{Pic}^0(E)$ has order 5. We will only consider 2-extensions of E. Take $D = 2(P_1) + 4(P_2)$. We have that

$$\begin{aligned} U/U_D &\cong \mathbf{F}_2[[t]]/\{1 \pmod{t^2}\} \oplus \mathbf{F}_2[[t]]/\{1 \pmod{t^4}\} \\ &\cong \mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/4\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z} \end{aligned}$$

Let L denote the fixed field of the Frobenius elements of the three other rational points P_3, P_4 and P_5. This is a quadratic extension. The genus g of the corresponding curve is given by $2g - 2 = 0 + 6$. We have found a curve over $\mathbf{F}_2$ of genus 4 with 8 points.

These examples show that one can find curves with many points with respect to their genus by means of class field theory. All examples are best possible: there do not exist curves over $\mathbf{F}_2$ of genus 1, 3 or 4 that have more points than the curves above. This follows from Oesterlé's estimates that have been explained in the previous section. Serre used this idea to find curves over $\mathbf{F}_2$ with rather high genus and a large number of rational points. We present two of his examples. Others can be found in the exercises.

Example (8.6). *(Serre)* Consider $\mathbf{P}^1$ over $\mathbf{F}_2$. Let P_2 be the unique place of degree 2 and let P_3 denote one of the two places of degree 3. Consider the full ray class field of conductor $P_2 + P_3$. Let K denote the subfield in which the point ∞ of $\mathbf{P}^1$ is completely split. This field has its Galois group over $\mathbf{P}^1$ isomorphic to $\mathbf{F}_4^* \times \mathbf{F}_8^*$. There are two characters of conductor P_2; they have degree 2. The six characters of conductor P_3 have degree 3 and the 12 remaining non-trivial characters have conductors of degree 2+3=5. We conclude from Proposition 8.5 that

$$2g_X - 2 = 21(2 \cdot 0 - 2) + 2 \cdot 2 + 6 \cdot 3 + 12 \cdot 5$$

and hence that $g_X = 21$. Since all points on X that lie over ∞ are rational, we see that X is a curve of genus 21 with 21 rational points. This is best possible.

Example (8.7). *(Serre)* Let E be the curve of genus 1 and five points from Example 8.3. Let $P_1, \ldots, P_5$ denote the rational points on E. We will consider ray class fields of conductor $k(P_1)$ and in particular the subfields in which the other points $P_2, \ldots, P_5$ are completely split. It is easy to see that there exists such a subfield K of conductor $12(P_1)$ and Galois group over $\mathbf{F}_2(E)$ isomorphic to $\mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z} \oplus \mathbf{Z}/2\mathbf{Z}$. There is one character of conductor $8(P_1)$, two of conductor $10(P_1)$ and the remaining four non-trivial ones have conductor $12(P_1)$. Let X be a smooth curve with function field K. By Proposition 8.5 we have that

$$2g_X - 2 = 2(2 \cdot 1 - 2) + 1 \cdot 8 + 2 \cdot 10 + 4 \cdot 12$$

and hence that $g_X = 39$. The unique point on X over P_1 is rational. Over every other point P_i there are precisely 8 rational points. So, the curve X has genus 39 and $1+8\cdot 4 = 33$ rational points. This is the maximal possible number for a curve of genus 39.

We conclude this section with a table. In the first column we have the genus g. In the second column the bound b from Theorem 7.2 is listed: every curve of genus g over $\mathbf{F}_2$ has at most b points. In the third column one finds the maximal number n for which a curve of genus g over $\mathbf{F}_2$ with n rational points is *known* to exist. Most entries in the table follow from the examples or the exercises in this section. They can also be found in [22,23,24]. The entries with an asterisk were found by Serre more recently.

Table (8.8).

g	b	n	g	b	n	g	b	n	g	b	n
1	5	5	7	11	10	13	15	15*	19	20	20
2	6	6	8	11	11	14	16	15	20	21	19
3	7	7	9	12	12	15	17	17	21	21	21
4	8	8	10	13	12	16	18	16	22	22	21
5	9	9	11	14	14*	17	18	17	39	33	33
6	10	10	12	15	14	18	19	18	50	40	40

Exercises.

(8.A) Let P_d be a point of degree $d \geq 2$ on $\mathbf{P}^1$ over $\mathbf{F}_2$. Show that the degree of the maximal subfield of 2-power degree inside the ray class field of conductor $2(P_d)$ in which the three rational points are split has degree 2^{d-2}. Show that the corresponding curves have genus 2 and 6 rational points for $d = 3$, genus 9 and 12 points for $d = 4$ and genus 28 and 24 points for $d = 5$.

(8.B) Use ray class fields of conductor $k(\infty)$ of $\mathbf{F}_2(\mathbf{P}^1)$ to show that there exists curves over $\mathbf{F}_2$ of genus 5 with 9 points and of genus 17 with 17 points.

(8.C) Let E be the curve of genus 1 over $\mathbf{F}_2$ from Example 8.3. Study its ray class fields of conductor $k(P_1)$ where P_1 is a rational point and show that there exist curves over $\mathbf{F}_2$ of genus 6 or 7 with 10 points, of genus 15 with 17 points. Study its ray class fields of 2-power degree of conductor $2(P_d)$ where P_d denotes a place of degree d. Show that there exist curves over $\mathbf{F}_2$ of genus 19 with 20 points and of genus 50 with 40 points.

(8.D) As Exercise 8.C with E replaced by a curve of genus 2 with 6 points. Show that there exist curves over $\mathbf{F}_2$ of genus 8 with 11 points, genus 9 or 10 with 12 points, of genus 22 with 21 points and of genus 26 with 24 points.

(8.E) As Exercise 8.C with E replaced by a curve of genus 3 with 7 points. Show that there exist curves over $\mathbf{F}_2$ of genus 11 with 13 points and of genus 12 or 13 with 14 points.

(8.F) Show there exist curves of genus g over $\mathbf{F}_2$ with n rational points, for the following pairs (g, n): (14,15), (15,16), (16,16), (17,17), (18,18), (20,19) and (21,20).

(8.G) (*Serre [22]*) Show that there is no curve over $\mathbf{F}_2$ of genus 7 with 11 points. (Hint: show that such a curve has no "new" points over small extensions of $\mathbf{F}_2$; use this to calculate its ζ-function and show that it factors into two polynomials that are coprime in $\mathbf{Z}[X]$; show that this contradicts the irreducibility of the Θ-divisor on the Jacobian of X.)

9. Class Field Towers.

In this section we show how to construct infinite class field towers of function fields; see also [4,Ch.IX]. We will use freely the main results of class field theory that have been explained in the previous section. As an application we prove Serre's result that $A(q) > c \log q$ for some absolute constant $c > 0$. We will moreover show that $A(2) \geq 2/9$.

The only way known to construct infinite class field towers is by means of a group theoretical result that will be proved below. Let ℓ be a prime and let G be a finite ℓ-group. We let $d = \dim H_1(G, \mathbf{Z}/\ell\mathbf{Z})$ denote the *number of generators* of G and we let $r = \dim H_2(G, \mathbf{Z}/\ell\mathbf{Z})$ denote the *number of relations of* G. Here the dimensions are $\mathbf{F}_l$-dimensions. From the homology of the exact sequence $0 \longrightarrow \mathbf{Z} \longrightarrow \mathbf{Z} \longrightarrow \mathbf{Z}/\ell\mathbf{Z} \longrightarrow 0$ we see that

$$H_1(G, \mathbf{Z})/\ell H_1(G, \mathbf{Z}) \cong H_1(G, \mathbf{Z}/\ell\mathbf{Z})$$

and we have a surjective map

$$H_2(G, \mathbf{Z}/\ell\mathbf{Z}) \longrightarrow H_1(G, \mathbf{Z})[\ell].$$

This shows that $r \geq d$.

Theorem (9.1). *(Golod and Shafarevič, 1965) Let ℓ be a prime. For every finite ℓ-group G one has that*

$$r > \frac{1}{4}d^2.$$

Proof. Let I denote the augmentation ideal in the group ring $\mathbf{F}_l[G]$. Since $I/I^2 \cong H_1(G, \mathbf{Z}/\ell\mathbf{Z})$ we see, using Nakayama's lemma, that the minimal number of $\mathbf{F}_l[G]$-generators of I is d. Let F_d be a free $\mathbf{F}_l[G]$-module of rank d admitting a surjective map $F_d \longrightarrow I$. From the long homology sequence it is easy to see that the minimal number of $\mathbf{F}_l[G]$-generators of the kernel of this map is precisely r. Therefore there exists an exact sequence of $\mathbf{F}_l[G]$-modules:

$$F_r \longrightarrow F_d \longrightarrow I \longrightarrow 0$$

where F_r is free of rank r. Because d is the *minimal* number of generators of I the map $F_d \longrightarrow I$ is an isomorphism modulo I. So the image of F_r must be contained in IF_d. We conclude that we have exact sequences

$$F_r/I^{k-1}F_r \longrightarrow F_d/I^k F_d \longrightarrow I/I^{k+1} \longrightarrow 0$$

for every $k \geq 1$. We obtain the following inequalities for $0 \leq t < 1$:

$$\sum_{k=1}^{\infty} \dim_{\mathbf{F}_l}(F_d/I^k F_d)t^k \leq \sum_{k=1}^{\infty} \dim_{\mathbf{F}_l}(I/I^{k+1})t^k + \sum_{k=1}^{\infty} \dim_{\mathbf{F}_l}(F_r/I^{k-1}F_r)t^k.$$

Observe that the group ring $\mathbf{F}_l[G]$ is finite, so the ideal $I^k = 0$ for k large enough. Therefore the sums converge for $0 \leq t < 1$. We define the "Poincaré polynomial" $P(t)$ by

$$P(t) = \sum_{k=0}^{\infty} \dim_{\mathbf{F}_l}(I^k/I^{k+1})t^k.$$

It is easy to see that for a free $\mathbf{F}_l[G]$-module M of rank m one has that

$$\sum_{k=0}^{\infty} \dim_{\mathbf{F}_l}(M/I^{k+1}M)t^k = \frac{mP(t)}{1-t}.$$

Using this, the inequality above becomes

$$\frac{tdP(t)}{1-t} \leq \frac{P(t)-1}{1-t} + \frac{rt^2P(t)}{1-t}$$

for $0 \leq t < 1$ and hence we obtain

$$rt^2 - dt + 1 \geq \frac{1}{P(t)} \qquad \text{for } 0 \leq t < 1.$$

Since $P(t)$ has all its coefficients positive, we have that $rt^2 - dt + 1 > 0$ for all $0 \leq t < 1$. We have already seen that $r \leq d$ and hence that $0 < d/2r < 1$. Substituting this value for t gives the required result.

Theorem 9.1 was proved by Golod and Shafarevič in 1965. They used it to solve the "class field tower problem". More precisely, they showed that there exist algebraic number fields that possess infinite class field towers. Their proof applies to function fields over finite fields as well. We will now study the function field case in more detail.

Let X be a curve over $\mathbf{F}_q$ and let K denote its function field. Let S denote a non-empty set of places on X. Recall that places are Galois conjugacy classes of points. The *S-Hilbert class field* of K is the maximal unramified abelian extension $H(K)$ of K in which all places in S split completely. This is a finite extension K_1 of $K = K_0$ and by class field theory, the Galois group $\mathrm{Gal}(K_1/K_0)$ is isomorphic to $\mathrm{Pic}_S(X) = \mathrm{Pic}(X)/B$ where B is the subgroup generated by the points in S. This can be repeated: Let S' be the set of places of H that lie over S. Let $K_2 = H(H(K))$ be the S'-Hilbert class field of $K_1 = H(K)$. Etcetera. In this way one obtains a sequence of fields

$$K = K_0 \subset K_1 \subset \ldots \subset K_n \subset \ldots$$

all unramified over K in which the places in S are totally split. This is the *S-class field tower* of K or X. It is said to be finite if it stabilizes i.e. if for some n one has that $K_m = K_n$ for all $m \geq n$, and infinite otherwise.

Rather than class field towers we will consider ℓ-class field towers since these are easier to handle. For a prime ℓ one defines the (ℓ, S)-class field tower of a function field K in a similar way: The *(ℓ, S)-Hilbert class field* of K is the maximal unramified abelian ℓ-extension $H_\ell(K)$ of K in which all places in S split completely. This is a finite extension K_1 of $K = K_0$ and by class field theory, the Galois group $\mathrm{Gal}(K_1/K_0)$ is isomorphic to the ℓ-part of $\mathrm{Pic}_S(X) = \mathrm{Pic}(X)/B$ where B is the subgroup generated by the points in S. As before, one can repeat this and obtain the (ℓ, S)-class field tower of K. It is easy to see that a field K has an infinite S-class field tower whenever it has an infinite (ℓ, S)-class field tower for some prime ℓ.

We are mainly interested in infinite S-class field towers because of the following application:

Proposition (9.2). *Let X be a curve of genus g over $\mathbf{F}_q$. Let S be a set of $\mathbf{F}_q$-rational points of X. If X has an infinite S-class field tower then*

$$A(q) \geq \frac{\#S}{g-1}.$$

Proof. There is no loss in assuming that S is not empty. Consider the curves

$$X = X_0 \longleftarrow X_1 \longleftarrow X_2 \longleftarrow \ldots$$

whose function fields are the layers in the infinite S-class field tower of K. Since the points in S are totally split in every covering $X \longleftarrow X_n$, we see that X_n has field of constants equal to $\mathbf{F}_q$ and that it has at least $d_n\#S$ points rational over $\mathbf{F}_q$. Here d_n denotes the degree of X_n over X. Since X_n is unramified over X, the Hurwitz-Zeuthen formula for its genus g_n becomes

$$2g_n - 2 = d_n(2g-2).$$

Since the tower is infinite, we conclude that

$$A(q) \geq \lim_{n\to\infty} \frac{d_n\#S}{1+d_n(g-1)} = \frac{\#S}{g-1}$$

as required.

We will freely use the notation introduced in the previous section. Before deriving a criterion for a function field K to have an infinite (ℓ, S)-class field tower, we introduce some more notation. Let X be a curve with function field K and let S be a finite set of places on X. By O_S we denote the subring of K consisting of the functions that have at most poles at places in S. let O_S^* denote its unit group. The group of S-divisors $\mathrm{Div}_S(X)$ is the subgroup of divisors with support outside S. The kernel of the canonical map $\mathbf{A}_K^* \longrightarrow \mathrm{Div}_S(X)$ will be denoted by U_S. Finally we let P_S denote the group of principal S-divisors and we let $Q_S = U_S/O_S^*$. We have the following diagram with exact rows and columns:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & O_S^* & \longrightarrow & K^* & \longrightarrow & P_S & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & U_S & \longrightarrow & \mathbf{A}_K^* & \longrightarrow & \mathrm{Div}_S(X) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & Q_S & \longrightarrow & C_K & \longrightarrow & \mathrm{Pic}_S(X) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

For a finitely generated abelian group A and a prime number ℓ we let $d_\ell A$ denote the ℓ-rank of A, i.e. $d_\ell A$ is the $\mathbf{F}_l$-dimension of $A/\ell A$.

Proposition (9.3). *Let X be a curve over $\mathbf{F}_q$ and let S be a non-empty finite set of places of X. If for a prime number ℓ one has that*

$$d_\ell \mathrm{Pic}_S(X) \geq 2 + 2\sqrt{d_\ell O_S^* + 1}$$

then X has an infinite (ℓ, S)-class field tower.

Proof. Suppose X has a finite (ℓ, S)-class field tower. Let L denote the union of all the layers of the tower of function fields. The Galois group $G = \mathrm{Gal}(L/K)$ is a finite ℓ-group. Its maximal abelian quotient is the ℓ-part of $\mathrm{Pic}_S(X)$. Therefore its number of generators d satisfies

$$d = d_\ell \mathrm{Pic}_S(X).$$

From the exact sequence $0 \longrightarrow \mathbf{Z} \longrightarrow \mathbf{Z} \longrightarrow \mathbf{Z}/\ell\mathbf{Z} \longrightarrow 0$ we obtain the exact sequence

$$0 \longrightarrow H_2(G, \mathbf{Z})/\ell H_2(G, \mathbf{Z}) \longrightarrow H_2(G, \mathbf{Z}/\ell\mathbf{Z}) \longrightarrow H_1(G, \mathbf{Z})[\ell] \longrightarrow 0$$

and the inequality

$$r - d \le d_\ell H_2(G, \mathbf{Z}).$$

Next we compute the group $H_2(G, \mathbf{Z}) = \widehat{H}^{-3}(G, \mathbf{Z})$. By class field theory we have that $\widehat{H}^{-3}(G, \mathbf{Z}) \cong \widehat{H}^{-1}(G, C_L)$. To compute the latter group we consider the diagram above with field L and the set of places S' of places of L over S: by the maximality of L, the ℓ-part of $\mathrm{Pic}_S(Y)$ is trivial and so are its G-cohomology groups. Since L is unramified over K the G-cohomology groups of U_L are zero and we find that $\widehat{H}^{-1}(G, C_L) \cong \widehat{H}^{-1}(G, Q_L) \cong \widehat{H}^0(G, O_{S'}^*)$. we conclude that

$$r - d \le d_\ell H^0(G, O_{S'}^*) \le d_\ell O_S^*.$$

The result now follows easily from the theorem of Golod and Shafarevič and the fact, mentioned above, that $d = d_\ell \mathrm{Pic}_S(X)$.

To apply this proposition, we need to know $d_\ell O_S^*$ and $d_\ell \mathrm{Pic}_S(X)$. It is well-known that

$$d_\ell O_S^* = \begin{cases} \#S, & \text{if } \ell | q-1; \\ \#S - 1, & \text{otherwise.} \end{cases}$$

To estimate the ℓ-rank of $\mathrm{Pic}_S(X)$ we consider only a special case:

Proposition (9.4). *Let X be a curve over $\mathbf{F}_q$ with function field K. Suppose that K is a cyclic extension of degree ℓ of a field F. Let S' denote the set of places of F over which the places in S lie and let ρ denote the number of places of F that are ramified in K. Then*

(i) If

$$\rho \ge 3 + d_\ell O_{S'}^*/(O_{S'}^* \cap NU_S) + 2\sqrt{d_\ell O_S^* + 1}$$

then K has an infinite class field tower.

(ii) If

$$\rho \ge \begin{cases} 3 + \#S' + 2\sqrt{\#S + 1}, & \text{if } \ell | q-1; \\ 2 + \#S' + 2\sqrt{\#S}, & \text{otherwise.} \end{cases}$$

then K has an infinite class field tower.

Proof. Using class field theory we obtain the following inequalities from the diagram above:

$$\begin{aligned} d_\ell \mathrm{Pic}_S(X) &\ge d_\ell \widehat{H}^{-1}(\pi, \mathrm{Pic}_S(X)) \ge d_\ell \widehat{H}^0(\pi, Q_S) - d_\ell - \widehat{H}^0(\pi, C_K) \\ &\ge d_\ell \widehat{H}^0(\pi, U_S) - d_\ell O_{S'}^*/(O_{S'}^* \cap NU_S) - d_\ell \widehat{H}^{-2}(\pi, \mathbf{Z}) \\ &\ge \rho - d_\ell O_{S'}^*/(O_{S'}^* \cap NU_S) - 1. \end{aligned}$$

The last inequality follows, because we have that $U_S \cong \prod_{v \notin S} O_v^* \prod_{v \in S} K_v^*$ where O_v is the completed local ring at v and K_v its quotient field. Therefore, by local class field theory, $d_\ell H^0(\pi, U_S)$ is equal to the number of ramified places ρ of F plus the number of inert places in S'. We will neglect the latter contribution. A combination with proposition 9.3 yields (i). Part (ii) follows from this together with the estimate

$$d_\ell O_{S'}^*/(O_{S'}^* \cap NU_S) \leq \begin{cases} \#S'+1, & \text{if } \ell | q-1; \\ \#S'; & \text{otherwise.} \end{cases}$$

This proves the Proposition.

We will use this Proposition to prove some results concerning the function $A(q)$ that was introduced in section 7.

Corollary (9.5). *For every finite field $\mathbf{F}_q$ and every prime ℓ there exists a curve X over $\mathbf{F}_q$ and a set $S \neq \emptyset$ of rational places of X such that X has an infinite (ℓ, S)-class field tower. For every finite field $\mathbf{F}_q$ one has that $A(q) > 0$.*

Proof. Let F be $\mathbf{F}_q(T)$, the function field of $\mathbf{P}^1$. Let K be a cylic extension of degree ℓ of F in which at least 8 places are ramified one of which is an $\mathbf{F}_q$-rational point P. Using class field theory, it is a trivial matter to exhibit such an extension. Now take $S' = \{P\}$ and S equal to the unique point over P in K. By Proposition 9.4 the field K has an infinite (ℓ, S)-class field tower and it follows from Proposition 9.2 that $A(q) > 0$.

Proposition (9.6). *There exists an absolute constant $c > 0$ such that $A(q) > c \log q$.*

Proof. Let us first suppose that q is odd. In view of Corollary 9.5 there is no loss in assuming that q is large. We let A and B be two subsets of $\mathbf{F}_q$ with the property that $\alpha - \beta \in (\mathbf{F}_q^*)^2$ for every $\alpha \in A$ and $\beta \in B$. Next we let X be the hyperelliptic curve given by $Y^2 = \prod_{\alpha \in A}(T-\alpha)$. This is a curve of genus $\#A/2 - 1$ or $(\#A-1)/2$. All places $(T-\alpha)$ of $\mathbf{P}^1$ with $\alpha \in A$ are ramified in the function field K of X. For S we choose the set of places lying over the places $(T-\beta)$ with $\beta \in B$. Since $\alpha - \beta$ is a square for every $\alpha \in A$ and $\beta \in B$, the places in S are $\mathbf{F}_q$-rational. Moreover, we have, in the notation of Proposition 9.4(i), that $O_{S'}^* \subset NU_S$ and we conclude that X admits an infinite $(2, S)$-class field tower when $\#A \geq 3 + 2\sqrt{\#B+1}$. Choosing $\#B$ as large as possible with respect to $\#A$ i.e. $\#B \sim (\#A)^2$, we have in this case that

$$A(q) \geq c\frac{\#B}{\#A} \sim \#A$$

for some $c > 0$. When q is even there is a similar argument with quadratic Artin-Schreier extsensions.

It remains to see how large A can be chosen. The following combinatorial lemma gives an estimate for this. It should be applied with $\Omega = \mathbf{F}_q$ and $R = \{(x,y) \in \Omega \times \Omega : x - y \in (\mathbf{F}_q^*)^2\}$. One has that $m = (q-1)/2$. It follows easily from the lemma that one can find sets A and B with $a = \#A \sim \log q$ and $b = \#B \sim \log^2 q$. This proves the Proposition.

Lemma (9.7). *Let Ω be a finite set of cardinality ω and let $R \subset \Omega \times \Omega$. Suppose that*

$$\#\{x \in \Omega \colon (x,y) \in R\} \geq m \qquad \text{for all } x \in \Omega.$$

If

$$b\binom{\omega}{a} \leq \omega \binom{m}{a}$$

then there exist two subsets A of cardinality a and B of cardinality b such that $A \times B \subset R$.

Proof. Let $T = \{(A, y): A \subset \Omega, \#A = a, A \times \{y\} \subset R\}$. Each fiber of the natural projection $T \longrightarrow \Omega$ given by $(A, y) \mapsto y$, contains at least $\binom{m}{a}$ elements. Therefore $\#T \geq \omega\binom{m}{a}$.

Now let $P_a(A)$ denote the collection of subsets of Ω of cardinality a and consider the other projection $T \longrightarrow P_a(\Omega)$ given by $(A, y) \mapsto A$. Since the cardinality of $P_a(\Omega)$ is $\binom{\omega}{a}$, we see that there must exist a fiber with at least

$$\#T/\binom{\omega}{a} \geq \omega\binom{m}{a}/\binom{\omega}{a} \geq b$$

elements. So, it suffices to take the set A corresponding to this fiber and $B = \{y \in \Omega: (A, y) \in R\}$. This proves the Lemma.

Finally we restrict our attention to the field $\mathbf{F}_2$. We know that $A(2) \leq \sqrt{2} - 1 \approx 0.414$. Here we will prove a lower bound.

Theorem (9.8). *One has that*

$$A(2) \geq \frac{2}{9} \approx 0.222.$$

Proof. We will give two proofs. The first one is due to Serre. In both proofs a curve X and a set S of places are exhibited such that the function field K of X has an infinite $(2, S)$-class field tower.

(i) Serre starts with a curve of genus 1 with precisely two $\mathbf{F}_2$-rational points, e.g. $Y^2 + XY = X^3 + X + 1$. This curve has three points with field of definition equal to $\mathbf{F}_4$ and four with field of definition $\mathbf{F}_8$. Let F denote the function field of E and let S' be the set of two rational points on E. The unit group $O_{S'}^*$ is infinite and cyclic generated by ε say. For each of the points Q of degree 2 and 3 there exists a quadratic extension of conductor $2(Q)$ where ε is a local norm. Therefore there exists a quadratic extension K of F of conductor $\sum_Q 2(Q)$ in which the points in S' are totally split and for which $\varepsilon \in NU_S$. Here S denotes the set of places over S'. Since $d_2 O_S^* = 3$ and $O_{S'}^* \subset NU_S$, it follows from Prop.9.4(i) that K has an infinite $(2, S)$-class field tower. The genus g_X of the corresponding smooth curve X is given by Prop.8.2.:

$$2g_X - 2 = 2(2 \cdot 1 - 2) + 3 \cdot 4 + 4 \cdot 6.$$

So $g_X = 19$ and $\#S = 2\#S' = 4$ and we conclude from Prop.9.2 that

$$A(q) \geq \frac{4}{19 - 1} = \frac{2}{9}.$$

(ii) This time we start with a curve E of genus 1 and 5 rational points $P_1, P_2, \ldots, P_5$ over $\mathbf{F}_2$. It has been constructed in section 8 as a quadratic cover of $\mathbf{P}^1$ of conductor $4(\infty)$ in which the other two rational points (0) and (1) are split. We will make two quadratic extensions of the function field of E:

- F: A quadratic extension of conductor $2Q$ in which all points $P_1, P_2, \ldots, P_5$ split. Here Q is a point of degree 5 of E. The curve Y corresponding to this field has genus 6 and 10 points rational over $\mathbf{F}_2$. It has been constructed before in section 8.
- F': This field is the composite of $\mathbf{F}_2(E)$ and the following extension of the function field of $\mathbf{P}^1$: a quadratic extension of conductor $2(0) + 2(1)$ in which (∞) splits. As a consequence F' has conductor $2\sum_{i=1}^{4} P_i$ and P_5, the unique place over (∞), splits.

We let X denote the curve corresponding to the function field $F'F$. It admits a map of degree two to Y. Now we apply Proposition 9.4(ii). For S' we take the two points on Y over P_5. They are split in the covering $X \longrightarrow Y$. The remaining 8 rational points on Y lie over $P_1, \ldots, P_4$ and they are all ramified. We conclude that $\#S = 4$ and that $\rho = 8$. Therefore

$$\rho = 2 + \#S' + 2\sqrt{\#S}$$

and we see that $K = \mathbf{F}_2(X)$ has an infinite $(2, S)$-class field tower. The degrees of the conductors of the characters of the Galois group of X over E are 10, 8 and 18 respectively. It follows from Proposition 8.2 that the genus of X is 19. Since $\#S = 4$ we deduce from Proposition 9.2 that $A(q) \geq 4/(19-1) = 2/9$ as required.

Exercises.

(9.A) Let G be a finite abelian p-group with d independent generators. Show that it has precisely $\binom{d+1}{2}$ independent relations.

(9.B) Let p be a prime. Show that the group presented as a pro-p-group as $< x, y : x^p = 1, [x, y] = y^p >$ is a finite p-group. It is a group with two independent generators x and y and two relations.

(9.C) Let p be a prime and let G be a group of order p^a. Show that G has at most a independent generators and at most $\binom{a+1}{2}$ independent relations.

(9.D) (*Mennicke*) Let p be a prime and let $d \geq 3$ be an integer. Let G be the pro-p-group generated by $x_1, x_2, \ldots, x_d$ with the following relations: $[x_i, x_{i+1}] = x_i^p$ for $1 \leq i \leq d$ and $[x_i, x_j] = 1$ for all i and j for which $i - j \neq \pm 1$. Here all indices should be taken modulo d. Show that G is a finite p-group with d generators and $\binom{d}{2}$ relations.

(9.E) Exhibit for each prime p and each $d \leq 3$ a finite abelian p-group with d generators and d relations. Show that no such group exists when $d > 3$. Give an example of a p-group with 4 generators and 6 relations. It is not known whether there exist finite p-groups with 4 generators and 5 relations.

(9.F) (*Kostrikin*) Let p be a prime and let $d > 1$ be an integer. Let G_d be the pro-p-group generated by $x_1, x_2, \ldots, x_d$ and $y_1, y_2, \ldots, y_d$ with the following relations: $x_i^p = y_i^p = 1$ for all $1 \leq i \leq d$, $[x_i, x_j] = [y_i, y_j]$ for all $i \neq j$ and $[x_i, y_j] = 1$ for all $i \neq j$. Show that G_d is a finite p-group with $2d$ independent generators and at most $\frac{3}{2}d(d+1)$ independent relations. Conclude that for every prime p one has that $1/4 \leq \limsup_{\#G, d\to\infty}(r/d^2) \leq 3/8$. Here the limit is taken over finite p-groups G with d generators and r relations. It has been shown by J. Wisliceny [36] that for $p > 2$ one actually has that $\limsup_{\#G, d\to\infty}(r/d^2) = 1/4$.

Bibliography

[1] Artin, E. and Tate, J.T.: *Class field theory*, Math. Lecture Notes, Benjamin, New York 1967.

[2] Barg, A.M., Katsman S.L. and Tsfasman, M.A.: Algebraic geometric codes from curves of small genus, *Probl. of Information Transmission* **23** (1987), 34–38.

[3] Bombieri, E.: Counting points on curves over finite fields (d'après Stepanov), Sém. Bourbaki **430**, juin 1973.

[4] Cassels J.W.S. and Fröhlich A.: *Algebraic number theory*, Academic Press, London 1967.

[5] Chevalley, C.: *Introduction to the theory of algebraic functions of one variable*, Math. Surveys **6**, AMS, New York 1951.

[6] Cusack, E.: *Codes*, Open University TM 361–14, Open Univ. Press, Walton Hall, Milton Keynes 1982.

[7] Drinfeld, V.G. and Vlădut, S.G.: The number of points on an algebraic curve, *Funct. Anal. i Ego Pril.*, **17** (1983), 68–69. (=*Functional Analysis* **17** (1983), 53–54.)

[8] Fulton, W.:*Algebraic curves*, W.A.Benjamin, New York 1969.

[9] Goppa, V.D.: Codes on algebraic curves, *Doklady Akad. Nauk SSSR*, **259** (1981), 1289–1290. (=*Soviet Math. Dokl.* **24** (1981), 170–172.)

[10] Goppa, V.D.: Algebraico-geometric codes, *Izvestia. Akad. Nauk SSSR*, **46** (1982), 209–257. (=*Math. USSR Izvestiya* **21** (1983), 75–91.)

[11] Goppa, V.D.: *Geometry and codes*, Mathematics and its applications (Soviet Series), Kluwer Ac. Publ., Dordrecht 1988.

[12] Hartshorne, R.: *Algebraic geometry*, Graduate Texts in Math. **52**, Springer-Verlag, Berlin 1977.

[13] Ihara, Y.: Some remarks on the number of rational points of algebraic curves over finite fields, *J. Fac. Sc. Tokyo*, **28** (1981), 721–724.

[14] Iyanaga, S.: *The theory of numbers*, North Holland Math. Lib. **8**, Amsterdam 1975.

[15] Lachaud, G.: Les codes géométriques de Goppa, Sém. Bourbaki **641**, Février 1985.

[16] MacWilliams F.J. and Sloane N.J.A.: *The theory of error-correcting codes*, North-Holland, Amsterdam 1978.

[17] Manin, Yu.I.: What is the maximum number of points on a curve over $\mathbf{F}_2$?, *J. Fac. Sci. Univ. Tokyo I A* **28** (1981), 715–720.

[18] Schoof, R. and Van der Vlugt, M.: Hecke operators and the weight distributions of certain codes, To appear in *J. of Comb. Theory A* (1991).

[19] Schoof, R.: Algebraic curves over $\mathbf{F}_2$ with many rational points, Preprint RUU 1990.

[20] Serre, J.-P.: *Groupes algébriques et corps de classes*, Hermann, Paris 1959.

[21] Serre, J.-P.: *Corps locaux*, Hermann, Paris 1968.

[22] Serre, J.-P.: Nombre de points sur une courbe sur un corps fini $\mathbf{F}_q$, notes taken by M. Waldschmidt (19 pages), octobre 1982.

[23] Serre, J.-P.: Nombre de points des courbes algébriques sur $\mathbf{F}_q$, *Sém. Th. de nombres, Bordeaux* exp. **22** (1982–1983).

[24] Serre, J.-P.: Sur le nombre des points rationnels d'une courbe algébrique sur un corps fini, *Comptes Rendus Acad. Sci. Paris* **296** (1983), 397–402.

[25] Serre, J. -P.: Résumé des cours de 1983–1984, Annuaire du Collège de France (1984), 79–83.

[26] Shafarevič, I.R.: *Basic algebraic geometry*, Springer-Verlag, Berlin 1977.

[27] Silverman, J.H.: *The arithmetic of elliptic curves*, Graduate Texts in Math. **106**, Springer-Verlag, Berlin 1986.

[28] Stewart, I.N. and Tall, D.O.: *Algebraic number theory*, Chapman and Hall, London 1987.

[29] Tsfasman, M.A., Vlădut, S.G. and Zink, Th.: Modular curves, Shimura curves and Goppa codes better than the Varshamov-Gilbert bound, *Math. Nachr.* **109** (1982), 21–28.

[30] Tsfasman, M.A. and Vlădut, S.G.: *Algebrogeometritskie codi*, Moskou 1988.

[31] Van der Waerden, B.L.: *Algebra I und II*, Heidelberger Taschenbücher **12,23**, Springer-Verlag, Berlin 1967,1971.

[32] Van Lint, J.H.: *Introduction to coding theory*, Graduate Texts in Math. **86**, Springer-Verlag, Berlin 1982.

[33] Van Lint, J.H. and Van der Geer, G.: *Introduction to coding theory and algebraic geometry*, DMV Seminar **12**, Birkhäuser Verlag, Berlin 1988.

[34] Van Wijngaarden, A. et al.: *Revised report on the algorithmic language Algol 68*, Springer-Verlag, Berlin 1976.

[35] Weil, A.: *Sur les courbes algébriques et les variétés qui s'en déduissent*, Hermann, Paris 1948.

[36] Wisliceny, J.: Zur Darstellung von pro-p-gruppen und Lieschen Algebren durch Erzeugenden und Relationen, *Math. Nachr.* **102** (1981), 57–78.

[37] Zink, Th.: Degeneration of Shimura surfaces and a problem in coding theory, in Budach, L.: *Fundamentals of Computation Theory*, Cottbus, GDR 1985. Lecture Notes in Computer Science **199**, Springer-Verlag, Berlin 1986.