Proceedings Trim Size: 9in x 6in
Text Area: 7.35in (include runningheads) x 4.5in
Main Text is 10/13pt

For Half-Title Page (prepared by publisher)

Publishers' page — (Blank page)

For Full Title Page (prepared by publisher)

For Copyright Page (prepared by publisher)

To Gilles Lachaud on the occasion of his 60th birthday

This volume, as well as the Tahiti meeting itself, is dedicated to our dear friend Gilles Lachaud on the occasion of his 60-th birthday.

The educational and research interests of Gilles Lachaud are very broad. He began his research in Paris under the guidance of Roger Godement and the first object of his interest was automorphic forms. Then, following the discovery of unknown books of Diophantos that he wrote the commentaries for, he became interested in the history of mathematics. In the early 1980s, after the discovery of algebraic geometry codes, he gives the first expository talk about them at the Bourbaki seminar, and thus created the French school of the geometric theory of error-correcting codes. His papers on the subject vary from classical questions for error-correcting codes - like those on Kloosterman sums, cyclic codes and bent-functions - to purely algebraic geometry problems. Among the latter let us mention those concerning the number of points of a Jacobian over a finite field, those about the asymptotic number of points on surfaces and multi-dimensional varieties, and the number of points on Grasmannians. In the paper on points on surfaces, he cleverly applied his - nowadays rare - knowledge of 19th century analysis. His early interest in continuous fractions led him to the theory of sails, where (after Klein and Arnold) he is one of the pioneers.

Gilles has many PhD students, and many of them have developed into working mathematicians. Gilles has the ability of giving extremely clear talks, and his more general expository papers are read even by non-mathematicians.

Gilles' non-mathematical interests include philosophy, literature, art, history, etc. At the age of 50 he started to learn Sanskrit.

The CIRM (International Centre for Mathematical Meetings in Luminy), as we see it now, owes a lot to the times when Gilles Lachaud was its director. After that, he joined the CNRS and created the Laboratory of Discrete Mathematics (LMD) in Marseille that grew into the Luminy Institute of Mathematics (IML), of which he is now the director. The mathematics

of Guadeloupe and Tahiti owes a lot to his efforts.

It is Gilles' idea to organize the recurrent AGCT (Arithmetic, Geometry, and Coding Theory) meeting we have in Luminy every two years. This meeting is in a sense the heart of the geometry codes community.

His attitude to colleagues, friends, and people in general, is not just benevolent, but also actively helpful.

We wish him many fruitful years of research and many happy returns of the day.

R.Rolland, M.Tsfasman

PREFACE

During five days from 7 to 11 May 2007, the mathematics symposium SAGA devoted to the applications of algebraic geometry, was held in Tahiti in the honor of Gilles Lachaud, Director of the *Institut de Mathématiques de Luminy*, bringing together researchers from about fifteen countries of Europe, Asia, America and Oceania. It made it possible to give a progress report on the current state of knowledge and research in the field of algebraic geometry and its applications to information theory.

The first conference of mathematics ever held in French Polynesia was organized jointly by the *Université de la Polynésie Française (UPF)* and the *Institut de Mathématiques de Luminy (IML)*. It is the result of a partnership developed between the team ATI of Luminy and the young research team GAATI, of the *Université de la Polynésie Française*. Let me thank the members of the team GAATI who by their scientific work, their cohesion and especially their good results, allowed the behaviour of this conference.

Among the fifty-nine participants who accept to come to Tahiti were two academicians, Mr Christophe Soulé of the *Académie des Sciences de Paris*, and Mr Igor Shparlinsky of the *Australian Academy of Science* and many of the best specialists of the domain.

Special thanks are due to all the authors who submitted papers and to the conference participants from all over the world. Among them, there were seven invited speakers : Jean-Marc Couveignes, *(Institut de Mathématiques de Toulouse, Université de Toulouse-Le Mirail, France* and *CNRS)*, Gerhard Frey, *(Institut für Experimentelle Mathematik, Universität Duisburg-Essen, Germany)*, Marc Girault, *(France Télécom R&D)*, Everett Howe, *(IDA Center for Communication Research,USA)*, Gilles Lachaud, *(Institut de Mathématiques de Luminy)*, Christophe Soulé *(IHES, Bures-sur-Yvette, France)*, and Felipe Voloch, *(University of Texas, Austin, USA)*

The timetable of the week devoted to the conference was distributed over four working days: Monday, Tuesday, Thursday and Friday. Wednesday was free and reserved to the visit of Mooréa. As part of the side events

of the scientific meeting, three conferences for the general public were given by Pierre Bathélemy, *(Institut de Mathématiques de Luminy)*, Marc Girault, *(France Télécom R&D)*, Robert Rolland, *(Institut de Mathématiques de Luminy)*. These three conferences aimed at informing the public on the current use of mathematical results in everyday life (smart cards, encoding, security) and at communicating with local engineers and technicians responsible for networks.

The session was opened by Louise Peltzer, president of the *Université de la Polynésie Française*, in the presence of local ministry officials in charge of research, at the Sheraton hotel of Papeete. Thirty five conferences showed the vitality, dynamism and richness of research in algebraic geometry and related topics. Conferences of two types were held: one-hour conferences suggested by the guests made it possible to present an survey of the current state of knowledge in the various fields touched upon in this symposium, and half-hour talks were reserved for the presentations of recent findings by lecturers. Out of the thirty-five conferences, twenty-seven were submitted and selected for publication in this volume of the Proceedings of SAGA in *Series on Number Theory and its Applications*, published by World Scientific.

The topics announced at the time of the call for papers were related to arithmetic and algebraic geometry over finite fields, combinatorial geometry over finite fields, algorithms over finite fields, error correcting codes, cryptography, boolean functions and sequences.

In this book of proceedings, many very interesting theoretical and applicated new results, and surveys are presented. For example, one can find new results on a conjecture of J.-P. Serre, which answer a question setted in a letter from J.-P. Serre also published in this book. New results in the domain of applications of the dicsrete logarithm problem and elliptic curves to cryptography, including discrete logarithm computation also appear in these proceedings. Other domains are covered as towers of function fields, new designs of classes of boolean cryptographic functions ... Some of the papers are surveys giving an overview of the state of the art in these related domains of research.

The working sessions were a powerful channel for the presentation of research in a field where popularization is not over-abundant, and they strengthened the links between various research centers with an interest in those topics. This symposium could not have been born without the effective support of the french national *Ministère de l'Éducation Nationale*, the

Ministère de l'Éducation, de l'Enseignement Supérieur et de la Recherche de la Polynésie Française, the *Fonds franais Économique, Social et Culturel pour le Pacifique*, the *Institut de Mathématiques de Luminy*, and the *Université de la Polynésie Française*.

The participants appreciated the Polynesian welcome which touched them deeply; a significant number of them extended their stay in French Polynesia, visiting the islands and discovering more about this country.

We would like to thank the sponsors, the corporate outfits which trustingly gave us support, namely OSB (*Océanienne de Services Bancaires*), ISS (Isis, Sigma, Spin), MANA (the Polynesian Internet provider). They showed us the interest of Polynesian companies in contacts with the world of research. Information concerning the conference was accessible on the site of the University of French Polynesia http://www.upf.pf as well as on that of the Mathematics Institute of Luminy: http://iml.univ-mrs.fr/ati/saga2007/welcome.html

Our thanks go to all the personnel of the *Institut de Mathématiques de Luminy*, of the *Université de la Polynésie Française*, the Sheraton Hotel in Tahiti and the various companies which contributed to the smooth running of the conference.

We would like to take this opportunity to thank the Program Committee members and the experts for their help in producing the conferences program. We also wish to thank Jean Chaumine, James Hirshfeld, Robert Rolland for their involvement in the editorial process. Let us thank again all the members of the team GAATI for their deep implication in the organization of this symposium and particularly Jean Chaumine who was the main coordinator.

J.-M. Goursaud

ORGANIZING COMMITTEES

EDITORS

- Jean Chaumine, Université de la Polynésie Française.
- James Hirschfeld, University of Sussex, Brighton, United Kingdom.
- Robert Rolland, Institut de Mathématiques de Luminy, Marseille.

PROGRAM COMMITTEE

- James Hirschfeld, University of Sussex, Brighton, United Kingdom.
- Igor Shparlinski, Fellow of the Australian Academy of Science, Macquarie University, Sydney, Australia.
- Philippe Langevin, Université de Toulon et du Var.
- Dominique Le Brigand, Université Pierre et Marie Curie, Paris.
- Marc Perret, Université du Mirail, Toulouse.
- Robert Rolland, Institut de Mathématiques de Luminy, Marseille.
- Serge Vladut, Institut de Mathématiques de Luminy, Marseille.
- Stéphane Ballet, Université de la Polynésie française.
- Régis Blache, Université de la Polynésie française.

ORGANIZATION COMMITTEE

- Aurélia Lozingot, Institut de Mathématiques de Luminy, Marseille.
- Eric Lozingot, Institut de Mathématiques de Luminy, Marseille.
- Corinne Roux, Institut de Mathématiques de Luminy, Marseille.
- Yves Aubry, Université de Toulon et du Var.
- Pierre Barthélemy, Institut de Mathématiques de Luminy, Marseille.
- Jean Chaumine, Université de la Polynésie française.

- Jean-Pierre Cherdieu, Université des Antilles et de la Guyane.
- Eric Férard, Université de la Polynésie française.
- Jean-Marie Goursaud, Université de la Polynésie française.
- Jean-Francis Michon, Université de Rouen.

SUPPORTED BY

Public institutions

- French national *Ministère de l'Éducation Nationale*
- French Polynesian *Ministère de l'Éducation, de l'Enseignement Supérieur et de la Recherche de la Polynésie Française*
- *Fonds français Économique, Social et Culturel pour le Pacifique*
- French *Centre National de la Recherche Scientifique - CNRS*
- *Université de la Polynésie Française*
- *Université de la Méditerranée*
- French polynesian team *Géométrie Algébrique, Arithmétique et Théorie de l'Information - GAATI* (Papeete)
- Laboratory *Institut de Mathématiques de Luminy -IML* (Marseille)

Companies

- *Océanienne de Services Bancaires (OSB)*
- *Isis Sigma Spin (ISS)*
- MANA

CONTENTS

Fast addition on non-hyperelliptic genus 3 curves

Stéphane Flon

UFR de mathématiques, Cité scientifique,
F-59655 Villeneuve d'Ascq, France
E-mail: Stephane.Flon@math.univ-lille1.fr

Roger Oyono

Department of Combinatorics and Optimizations
University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada
E-mail: royono@math.uwaterloo.ca

Christophe Ritzenthaler*

Institut de Mathématiques de Luminy,
UMR 6206 du CNRS, Luminy, Case 907,
13288 Marseille, France
E-mail: ritzenth@iml.univ-mrs.fr

We present a fast addition algorithm in the Jacobian of a genus 3 non-hyperelliptic curve over a field k of any characteristic. When the curve has a rational flex and k is a finite field of characteristic greater than 5, the computational cost for addition is $163M+2I$ and $185M+2I$ for doubling. We study also the rationality of intersection points of a line with a quartic and give geometric characterizations of $C_{3,4}$ curves and Picard curves. To conclude, an appendix gives a formula to compute flexes in all characteristics.

Keywords: Jacobians, non-hyperelliptic curves, addition, rationality, quartic, flex

Introduction

In this article, we present a simple geometric algorithm for addition in the Jacobian of non-hyperelliptic genus 3 curves, represented as smooth plane quartics. Several articles have been written on the subject (see Section

*The third author acknowledges the financial support provided through the European Community's Human Potential Programme under contract HPRN-CT-2000-00114, GTEM

5 for a discussion) and the present one continues this work by providing a straightforward generalization of [3,8,23]. Our contribution is at three different levels:

(1) we have devoted special care in writing the algorithm to minimize the number of operations. Thus, our algorithm is to date the fastest one for arithmetic in the Jacobian of a 'general' (see below) non-hyperelliptic genus 3 curve over a finite field k of characteristic greater than 5. As in previous articles, we measure the complexity by counting the number of multiplications M and inversions I that need to be performed in k. The computational cost for addition is $163M + 2I$ and $185M + 2I$ for doubling. Note that [22] has announced $117M + 2I$ for addition and $129M + 2I$ for doubling for $C_{3,4}$ curves which makes it the fastest algorithm for this special case.
(2) we present several mathematical results on the arithmetic of plane quartics. Indeed, the efficiency of our algorithm depends on the existence of a rational line l^∞ cutting the quartic in rational points only. We announce in this article that if $\#k \geq 127$ there always exists such a line (Theorem 2.1) and if $\#k \geq 66^2 + 1$ and $\mathrm{char}(k) \neq 2$ then l^∞ can be chosen tangent to the quartic C (Theorem 2.2). We then study the remaining cases: we show heuristically that any quartic has a line l^∞ such that $(l^\infty \cdot C) = 3P + Q$ with $P, Q \in C(k)$ with probability about 0.63 (The point P is called a flex). We call this case the 'general case'. We finally show that quartics with a rational hyperflex (i.e. $P = Q$ with the previous notations) represent exactly the case of $C_{3,4}$ curves (Proposition 2.1) and we characterize among them Picard curves as the curves with a rational Galois point (Proposition 2.2).
(3) To confirm our heuristic probability, we made tests which required the computation of flexes. As far as we know, the most general method was due to Abhyankar [1] which works for all but characteristic 2. In this article, we present the first formula to compute the flexes in all characteristics.

Due to recent progress in index calculus attacks (see [6]), it appears unlikely that genus 3 non-hyperelliptic curves may be used for building discrete logarithm cryptosystems. However, as in [22], we point out that the results presented in this paper still may be useful for cover attacks on discrete logarithms of other curves particularly in connection with Weil descent. Moreover, as illustrated in Section 4, fast addition algorithms can be useful in some recent point counting algorithm, like the AGM or those based on

modular curves.
The article is organized as follows. In the first section we present the geometric description of our algorithm. Section 2 deals with the rationality issues of the intersection of a line and a plane quartic over a finite field. Section 3 deals with the translation of the geometry in an algebraic language, thanks to Mumford representation. We write down the operations performed in the tangent case and we optimize our algorithm in the flex case. Section 4 shows examples of application of our algorithm. The conclusion summarizes and compares complexities of already existing methods. Finally an appendix proves our formula to compute the flexes and gathers tables which describe in details the operations for addition and doubling in the 'general' case.

1. Geometric description of the algorithm

Let C be a non-singular curve of genus g over a field k. Let D^∞ be an effective k-rational divisor of degree g. A consequence of Riemann-Roch theorem is the following representation of divisors:

Fact 1 (Representation of divisors). *Let D be a rational degree 0 divisor of C. Then there exists a rational effective divisor D^+ of degree g such that $D^+ - D^\infty \sim D$. Generically, the divisor D^+ is unique.*

We now restrict ourselves to the case where C is a genus 3 non-hyperelliptic curve. Thanks to the canonical embedding, we may assume that C is a smooth plane quartic. Conversely, any smooth plane quartic is a genus 3 non-hyperellipic curve. We denote by x, y, z (or sometimes x_1, x_2, x_3) coordinates in $\mathbb{P}^2$.

We denote by $(*)$ the following condition: *There is a rational line l^∞ which crosses C in four (not necessarily distinct, but with multiplicity then) k-points $P_1^\infty, P_2^\infty, P_3^\infty, P_4^\infty$.*

Until the end of this section, we assume that condition $(*)$ is fulfilled (see Section 2 for a discussion on this topic when k a finite field).

We choose D^∞ to be the divisor $P_1^\infty + P_2^\infty + P_3^\infty$.

By abuse of language we say that a curve C' *goes through* nP if $i(C, C'; P) = n$, where $i(C, C'; P)$ denotes the intersection multiplicity of C and C' at P.

Proposition 1.1. *Let $D_1, D_2 \in \mathrm{Jac}(C)(k)$. Then $D_1 + D_2$ is equivalent to a divisor $D = D^+ - D^\infty$, where the points in the support of D^+ are given*

by the following algorithm:

(1) Take a cubic E defined over k which goes (with multiplicity) through the support of D_1^+, D_2^+ and $P_1^\infty, P_2^\infty, P_4^\infty$. This cubic also crosses C in the residual effective divisor D_3.

(2) Take a conic Q defined over k which goes through the support of D_3 and P_1^∞, P_2^∞. This conic also crosses C in the residual effective divisor D^+.

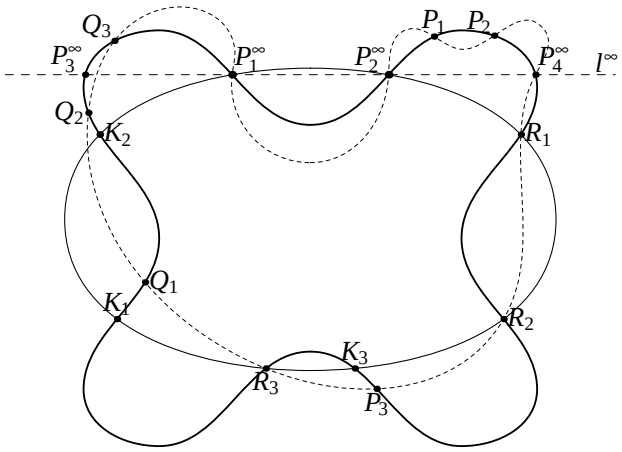


Fig. 1. Description of the algorithm

Proof. C being canonically embedded, $(E \cdot C) \sim 3\kappa$ where $\kappa = \kappa_C$ is the canonical divisor of C. Therefore we have

$$D_1^+ + D_2^+ + P_1^\infty + P_2^\infty + P_4^\infty + D_3 \sim 3\kappa.$$

Similarly, $(Q \cdot C) \sim 2\kappa$ so

$$D_3 + P_1^\infty + P_2^\infty + D_e \sim 2\kappa$$

and $(l^\infty \cdot C) = P_1^\infty + P_2^\infty + P_3^\infty + P_4^\infty \sim \kappa$. Combining these three relations, we obtain

$$D_1^+ + D_2^+ + P_1^\infty + P_2^\infty + P_4^\infty + D_3 \sim D_3 + P_1^\infty + P_2^\infty + D_e + P_1^\infty + P_2^\infty + P_3^\infty + P_4^\infty$$

so

$$D_1^+ + D_2^+ \sim D_e + D^\infty.$$

Now we subtract $2D^\infty$ on both sides:

$$D_1 + D_2 \sim D_e - D^\infty \sim D$$

So $D_e = D^+$.
The cubic E and conic Q are both defined over the field k because of the k-rationality of P_i^∞, D_1^+ and D_2^+. □

Remark 1.1. Actually, we need a milder hypothesis than $(*)$: it would be enough to have a line l^∞ such that P_3^∞, P_4^∞ are rational (this is always true over a finite field $k = \mathbb{F}_q$ with $q > 26$ see [7]). However, we need $(*)$ in order to simplify the equations of the different curves involved and optimize the algorithm (see Section 3).

2. Rationality of the points on a canonical divisor

2.1. *Structure of the canonical divisor*

Let C be a smooth plane quartic defined over an algebraically closed field $\bar{k}$. There are 5 possibilities for the intersection divisor $(l \cdot C) = P_1+P_2+P_3+P_4$ of a line l with C:

(1) The four points are pairwise distinct. This is the generic position.
(2) $P_1 = P_2$, then l is tangent to C at P_1.
(3) $P_1 = P_2 = P_3$. The point P_1 is then called a *flex*. As a linear intersection also represents the canonical divisor, these points are exactly the ones where a regular differential has a zero of order 3. They are thus the Weierstrass points of C. The quartic C has infinitely many flexes if and only if $\mathrm{char}(\bar{k}) = 3$ and C is isomorphic to the Fermat quartic $x^4 + y^4 + z^4 = 0$ (see [25, p.28]).
(4) $P_1 = P_2$ and $P_3 = P_4$. The line l is called a *bitangent* of the curve C and the points P_i *bitangence points*. It is well known (see for example [19]) that if $\mathrm{char}(\bar{k}) \neq 2$ then C has exactly 28 bitangents. If $\mathrm{char}(\bar{k}) = 2$, then C has respectively $7, 4, 2$, or 1 bitangents, according to the 2-rank of its Jacobian (resp. $3, 2, 1, 0$).
(5) $P_1 = P_2 = P_3 = P_4$. The point P_1 is called a *hyperflex*. Generically, such a hyperflex does not exist (*i.e.* the set of quartics with at least one hyperflex is of codimension 1 in the space of quartics). The number of hyperflexes is less than 12 if C is not isomorphic to the Fermat quartic over a field of characteristic 3. Moreover in this later case, the number of hyperflexes of C is equal to 28 (all the bitangence points are hyperflexes) (see [25, p.30]).

The efficiency of the algebraic version of the algorithm will depend on the choice of l^∞ (see Section 3). Roughly speaking, 'the more special the faster'. However, it is not clear for which choice of l^∞, the condition $(*)$ is fulfilled. We now study this condition when k is a finite field. For the general and tangent cases, we only state results whose proofs will be given in a forthcoming article [9].

In this section, we assume that k is a finite field $\mathbb{F}_q$ (with $q = p^n$ for a certain prime p).

2.2. *The general and tangent case*

Using the same techniques as in [6], we can prove the following result.

Theorem 2.1 ([9]). *Let C be a smooth plane quartic over $\mathbb{F}_q$. If $q \geq 127$, there exists a line which cuts C at rational points only, i.e. C satisfies the condition $(*)$.*

For the tangent case, we had to build a more elaborate strategy based on correspondence curves.

Theorem 2.2 ([9]). *Let C be a smooth plane quartic over $\mathbb{F}_q$ and assume that $char(\mathbb{F}_q) \neq 2$. If $q \geq 66^2 + 1$, there exists a tangent at C which cuts C at rational points only, i.e. C satisfies $(*)$ for a tangent line.*

Remark 2.1. We have been so far unable to extend the proof to the characteristic 2 case. We hope to solve this problem in a near future.

2.3. *The flex case*

Let us assume that C has a rational flex. Then the tangent at this point is a line satisfying $(*)$. Unfortunately, we do not know how to compute the probability for a quartic to have at least a rational flex. But we can have a guess on that number, coming from heuristic remarks on one side, and relying on numerical evidences on the other side.

Conjecture 2.1. *The probability that a smooth plane quartic has at least one rational flex is asymptotically, when q tends to ∞, equal to $1-e^{-1}+\alpha > 0.63$, with $|\alpha| \leq 10^{-25}$.*

Proof. *(heuristic)* Here we suppose that $\mathrm{char}(k) > 3$. Let $C : f = 0$ be the curve and $H(f) : h = 0$ its Hessian (see Appendix). The curve $H(f)$ is of degree 6 and the $(C \cdot H(f))$ are the 24 flexes with multiplicities. Generically,

when $q >> 0$ we may suppose that no two flexes have the same abscissae. Then there is a rational flex if and only if the polynomial $\mathrm{Res}(f, h, y)$ has a root in k. If we suppose that these polynomials are uniformly distributed among the polynomials of degree 24, then one only has to compute the probability that a polynomial of degree 24 has at least one linear factor in $\mathbb{F}_q$. Let $(\alpha_i)_{i\in\{1,\ldots,q\}}$ be an enumeration of $\mathbb{F}_q$.

Let S be the set of all monic polynomials of degree n and S_i the subset of S of polynomials having one or more factors of the form $x-\alpha_i, i = 1, \ldots, q$. Then $\#S = q^n$ and $\#S_i = q^{n-1}$. By inclusion-exclusion principle, the number $N(n, q)$ of monic polynomials of degree n with one or more linear factors is equal to

$$N(n,q) = \sum_{i=1}^{n} \binom{q}{i} q^{n-i}(-1)^{i-1} \quad \text{if } n < q,$$

and

$$N(n,q) = \sum_{i=1}^{q} \binom{q}{i} q^{n-i}(-1)^{i-1} \quad \text{if } n \geq q.$$

After straightforward computations, one computes that the probability $P(n, q)$ that a monic polynomial of degree n has at least a linear factor in $\mathbb{F}_q$ is

$$P(n,q) = 1 - \left(1 - \frac{1}{q}\right)^q - \beta_n(q)\,,$$

where

$$|\beta_n(q)| \leq \frac{1}{(n+1)!} \text{ and hence } \lim_{\substack{n<q \\ n,q\to\infty}} \beta_n(q) = 0\,.$$

Already for $n = 24$ and $q = 25$ we have $|\beta_n(q)| = 25^{-25} \leq 1.13 \cdot 10^{-35}$. □

We made numerical experiments to support our heuristic argument as well as to check the conjecture in characteristics 2 and 3. In these two cases, we have indeed $H(f) \equiv 0$. However, we have been able to find a good substitute for $H(\mathrm{f})$, see Section 6.
Computations realized with a bench of 10^6 non-singular quartics give the right percentage. Thus the conjecture seems to hold.

p	n	Probabilities
2	17	$632074/10^6 = 0.632074$
3	11	$632344/10^6 = 0.632344$
1009	2	$631358/10^6 = 0.631358$
$2^{17}+29$	1	$632921/10^6 = 0.632921$

2.4. *The hyperflex case*

We recall that a generic non-singular quartic has no hyperflex. If C has a rational hyperflex, then we find special curves already treated in the literature, namely $C_{3,4}$ curves. Recall that a C_{ab} curve is a non-singular curve X/k for which there exists a cover $\varphi : X \to \mathbb{P}^1$ in which a k-rational point P is totally ramified. Such a curve admit a plane affine model

$$X : \alpha_{0,a}\, y^a + \alpha_{b,0}\, x^b + \sum_{ia+jb<ab} \alpha_{i,j}\, x^i y^j = 0,$$

with $\alpha_{i,j} \in k$ and $\alpha_{b,0}, \alpha_{0,a} \neq 0$.

Proposition 2.1. *A non-singular plane quartic C with a rational hyperflex P is k-isomorphic to a $C_{3,4}$ curve of genus* 3.

Proof. By a k-linear rational transformation, we may suppose that P is the point $(0:1:0)$ and that the tangent at this point is the line at infinity, i.e. the line with equation $z = 0$. Therefore the equation of C is of the form

$$y^3 + h_1 y^2 + h_2 y = f_4\,,$$

where h_i is a degree i polynomial and f_4 is a degree 4 monic polynomial. □

Remark 2.2. We can wonder whether a plane quartic with a hyperflex generically has a rational hyperflex. This is actually the case: indeed, according to [26] and if $\mathrm{char}(k) > 3$, the locus of plane quartics with more or equal than two hyperflexes has codimension one in the locus of plane quartic with a hyperflex. So plane quartics with exactly one hyperflex (thus a rational hyperflex since it has to be Galois invariant) are generic. However, one can find rational families of quartics with at least two hyperflexes which are not defined over k. For instance $x^4 + (y^2 - \alpha z^2) \cdot Q(x, y, z)$ where $Q \in k[x, y, z]$ is a homogeneous degree 2 polynomial and α is not a square in k has $(0 : \sqrt{\alpha} : 1)$ and $(0 : -\sqrt{\alpha} : 1)$ as conjugate hyperflexes.

One may like to characterize Picard curves among $C_{3,4}$ curves. Recall that if $\mathrm{char}(k) \neq 3$, a Picard curve is a genus 3 curve which admits an affine model of the form $y^3 = f_4(x)$. Clearly the four points $(\alpha_i : 0 : 1) \in C$ are flexes whose tangent goes through $P = (0 : 1 : 0)$. Conversely, it is easy to see that Picard curves are exactly the smooth plane quartics with one rational hyperflex P and 4 distinct collinear flexes $(P_i)_{i=1,\dots,4}$ whose tangents are all concurrent at P (take $P = (0 : 1 : 0)$ and the line defined by the P_i's as the $y = 0$ line). Another characterization, maybe more natural, is in terms of Galois point. Such points have been studied in [18] over a field of characteristic 0 and are defined as follows. Let $P \in C(\overline{k})$ and $\phi_P : C \to |\kappa_C - P| = \mathbb{P}^1$ the degree 3 morphism induced by the linear system $|\kappa_C - P|$ (i.e. the lines going through P). A point P is called a *Galois point* if the geometric cover defined by ϕ_P is Galois. One has the following characterization.

Proposition 2.2. *Let* $\mathrm{char}(k) \neq 3$. *A smooth plane quartic* C *is a Picard curve if and only if there exists* $P \in C(k)$ *such that* P *is a Galois point.*

Proof. If C is a Picard curve, it admits a projective model $(y/z)^3 = f_4(x/z)$. Let $P = (0 : 1 : 0)$ and replace $x = tz$ for $t \in \overline{k}$. We obtain

$$\left(\frac{y}{z}\right)^3 = f_4(t)\,.$$

This clearly defines a Galois extension of $k(\mathbb{P}^1) = k(t)$. Conversely, let assume that C is a smooth plane quartic with a Galois point $P \in C(k)$. First we show that P is a hyperflex. If the cover ϕ_P is Galois then there exists an automorphism $\alpha : C \to C$ of order 3 such that $\phi_P : C \to C/\langle\alpha\rangle$. As C is canonically embedded, α induces a projective automorphism of $\mathbb{P}^2$. We show that $\alpha(P) = P$. Let $R_1 + R_2 + R_3 = \phi_P^{-1}(t_0)$ for a generic t_0. The line $\overline{\alpha(R_1)\alpha(R_2)}$ goes through $\alpha(P)$. The morphism α permutes the R_i so $\overline{\alpha(R_1)\alpha(R_2)} = \overline{R_1R_2}$ and $\alpha(P) = P$. The point P is then ramified in the cover ϕ_P and then is completely ramified. Thus, the tangent line to C at P cuts the divisor $4P$, i.e. P is a hyperflex.
Now if a point $Q \neq P$ is ramified then Q is completely ramified and it is then a flex. As $\mathrm{char}(k) \neq 3$, Hurwitz formula shows that there must be exactly 4 such flexes associated to P. We can assume that $P = (0 : 1 : 0)$ with tangent $z = 0$, and that two of them are the points $P_1 = (0 : 0 : 1)$ and the point $P_2 = (1 : 0 : 1)$. As P is a hyperflex, Proposition 2.1 shows that C admits an affine model

$$y^3 + (a_1x + a_0)y^2 + (b_2x^2 + b_1x + b_0)y = x(x-1)(x-r_1)(x-r_2)\,.$$

We have the following facts:

- since the tangent at P_1 (resp. P_2) goes through P, $b_0 = 0$ (resp. $b_2 = -b_1$);
- since P_1 (resp. P_2) is a flex, the tangent at P_1 (resp. P_2) cuts the curve only at P_1 and P (resp. at P_2 and Q). So $a_0 = 0$ (resp. $a_1 = 0$).

Then we actually get a model of the form

$$y^3 + bxy(x-1) = x(x-1)(x-r_1)(x-r_2)$$

and we are done if we show that $b = 0$.
We consider separately the case $\mathrm{char}(k) > 3$ and the case $\mathrm{char}(k) = 2$.
If $\mathrm{char}(k) > 3$, letting $x = tz$ we get the following equation for the cover

$$y^3 + (bt^2 - bt)y + (-t^4 + t_3(r_1 + r_2 + 1) - t^2(r_1r_2 + r_1 + r_2) + tr_1r_2)\,.$$

It is classical that this extension is Galois if and only if its discriminant $\Delta \in \overline{k}(t)$ is a square (here we need that $\mathrm{char}(k) \neq 2$). Now,

$$\Delta = -27t^2 \cdot (t-1)^2 \cdot [t^4 - 2(r_1 + r_2)t^3 + (r_1^2 + r_2^2 + 4r_1r_2 + \frac{4}{27}b^3)t^2$$
$$-2(r_1r_2^2 + r_2r_1^2 + \frac{2}{27}b^3)t + (r_1r_2)^2].$$

Thus Δ is a square if and only if the last factor is a square, i.e. can be written $(s_2t^2 + s_1t + s_0)^2$. It is easy to check that this implies $b = 0$.
If $\mathrm{char}(k) = 2$, let $P_3 = (x_3 : y_3 : z_3) \in C$ be a third flex such that its tangent goes trough P. In particular

$$\partial h/\partial y(P_3) = y_3^2 z_3 + bx_3z_3(x_3 - z_3) = 0.$$

We replace $y_3^2 = bx_3(x_3 - 1)$ in the equation of C and we get

$$x_3(x_3 - 1)(x_3 - r_1)(x_3 - r_2) = 0.$$

Thus let say $x_3 = r_1$. Let $x = r_1z$ then replacing in the equation of C, we get $zy^3 + br_1z^3y(r_1 - 1) = 0$. The point $(r_1 : 0 : 1)$ is then a flex if and only if $b = 0$. □

3. Algebraic description

In section 1, we gave a general geometric description of our algorithm. In this section, we will give an algebraic description in the tangent case and a completely optimized one, for implementation, in the flex case.

3.1. *Mumford representation and typical divisors*

We need a simple representation for the effective divisors D^+. Let C be a smooth plane quartic satisfying $(*)$. We may suppose (after a k-linear transformation) that P_1^∞ is a point at infinity (*i.e.* such that its z-coordinate is 0), and that l^∞ is the line $z = 0$. Let $f(x, y) = 0$ be an affine equation of C. As in [8], we work with *Mumford representation*. A divisor $D \in \mathrm{Jac}(C)(k)$ is represented by a couple $[u, v]$ of polynomials in $k[x]$. Recall that this representation is unique under the following generic assumptions on D, which define a *typical divisor*:

(1) The three points in the support of D^+ are non-collinear. In this case D^+ is unique: in fact if $P_1 + P_2 + P_3 + (f) = Q_1 + Q_2 + Q_3$ then $f \in \mathcal{L}(P_1 + P_2 + P_3)$ and f has to be constant by the Riemann-Roch theorem.
(2) There is no point at infinity in the support of D^+. Let $P_i = (x_i : y_i : 1)$ $(i = 1, 2, 3)$ be the three points in the support of D^+ and $u = \prod(x - x_i)$. Since D^+ is a rational divisor, $u \in k[x]$.
(3) The $(x_i)_{i=1,2,3}$ are distinct. In this case, there exists a unique polynomial $v \in k[x]$ of degree 2 such that $y_i = v(x_i)$ for $i = 1, 2, 3$ (it is simply the interpolation polynomial).

Conversely, given a couple $[u, v]$ such that

- $u, v \in k[x]$,
- $u = \prod(x - x_i)$ is monic of degree 3 and with simple roots,
- $\deg(v) = 2$,
- $u | f(x, v(x))$,

then $P_1 + P_2 + P_3 - D^\infty$ is a rational typical divisor of C (where, for $i \in \{1, 2, 3\}$, we have $P_i = (x_i : v(x_i) : 1)$).

Proposition 3.1. *Assume that k is algebraically closed, then the locus of non typical divisor is of codimension* 1 *in the Jacobian of C.*

Proof. Clearly if the points in the support of D^+ are collinear $l(\kappa - D^+) \neq 0$, i.e. D^+ is a special divisor. $D^+ - D^\infty$ is then contained in a translate of the theta divisor, i.e. in a variety of codimension 1.
If the second condition is not satisfied then D^+ is contained in the union $\cup_{P \in (C \cdot l^\infty)}(C + C + P)$. The image of this dimension 2 variety in the Jacobian of C is thus of codimension 1.
Let us assume (after a possible change of coordinates) that the point $(0 :$

$1:0$) does not belong to C. Denote $\phi : C \to \mathbb{P}^1$ the projection of the x-coordinate $\phi(x:y:z) = (x:z)$. Let

$$V = \{(P_1, P_2, P_3) \in C^3, \phi(P_1) = \phi(P_2) \text{ or } \phi(P_2) = \phi(P_3) \text{ or } \phi(P_3) = \phi(P_1)\}.$$

If the third condition is not satisfied, then $D^+ - D^\infty$ belongs to the image of V in the Jacobian of C. So again, it belongs to a variety of codimension 1. □

In particular, we see that addition of two typical divisors or doubling of a typical divisor is generically a typical divisor. As we are mainly interested in implementation over large fields where we can assume that the generic hypothesis holds, we will restrict our description of the algorithms to the case of a typical divisor. Note however that the non typical cases can be handled even more efficiently than the generic case since the representation uses polynomials $[u, v]$ of lower degrees.

3.2. *The tangent case*

After a k-linear transformation, we may suppose that $l^\infty : z = 0$ is tangent at $P_1^\infty = P_2^\infty = (0:1:0)$ and goes through $P_4^\infty = (1:0:0)$. An equation for C is then of the form

$$y^3 + h_1 y^2 + h_2 y = f_3,$$

where $h_1, h_2, f_3 \in k[x]$ and $\deg(h_1) \leq 2, \deg(h_2) \leq 3, \deg(f_3) \leq 3$. We then have

Lemma 3.1. *The cubic E from the theorem is generically of the form*

$$y^2 + s \cdot y + t,$$

where s and t are polynomials in $k[x]$, with $\deg(s) \leq 2$ and $\deg(t) \leq 2$. The conic Q is of the form

$$y - v,$$

where $v \in k[x]$ and $\deg(v) = 2$.

Proof. As $P_1^\infty \in E$ we see that an equation of E has no y^3 term. One can then write it in the form

$$y^2 d + sy + t$$

with d (resp. s, resp. t) polynomials in x of degree less than 1 (resp. less than 2, resp. less than 3). Now $l^\infty : z = 0$ is the tangent at E in P_1^∞ so we

can assume $d = 1$. Finally $P_4^\infty \in E$ implies that E has no x^3 term. This gives the form of the cubic.
As for the cubic, the conic Q must have a tangent line at P_1^∞ equal to l^∞. This gives directly the desired form. □

To explicit the coefficients of E and Q, one proceeds similarly as in [8, 2.1.2]]. Note that all the computations are carried over k.

Algorithm 3.1 (Algorithm for Addition).

INPUT: $D_1 = [u_1, v_1]$ and $D_2 = [u_2, v_2]$
OUTPUT: $D_1 + D_2 = [u_{D_1+D_2}, v_{D_1+D_2}]$

1. *Computation of the cubic E*
   Addition
   compute the inverse t_1 of $v_1 - v_2$ modulo u_2
   compute the remainder r of $(u_1 - u_2)t_1$ by u_2
   solve the linear equations given by the following conditions

$$\begin{cases} \deg_x(-v_1(v_1+s) + u_1\delta_1) = 2 & (2 \text{ eq.}) \\ v_1 + v_2 + s \equiv r\delta_1 \quad [u_2] & (3 \text{ eq.}) \end{cases}$$

   where $s, \delta_1 \in k[x]$ with $\deg(s) = 2$ and $\deg(\delta_1) = 1$. Then

$$E = (y - v_1)(y + v_1 + s) + u_1\delta_1$$

   Doubling
   compute $\omega_1 = (v_1^3 + v_1^2 h_1 + v_1 h_2 - f_3)/u_1$
   compute the inverse t_1 of ω_1 modulo u_1
   compute the remainder r of $(3v_1^2 + 2v_1 h_1 + h_2)t_1$ by u_1
   solve the linear equations given by the following conditions

$$\begin{cases} \deg_x(-v_1(v_1+s) + u_1\delta_1) = 2 & (2 \text{ eq.}) \\ 2v_1 + s \equiv r\delta_1 \quad [u_1] & (3 \text{ eq.}) \end{cases}$$

   where $s, \delta_1 \in k[x]$ with $\deg(s) = 2$ and $\deg(\delta_1) = 1$. Then

$$E = (y - v_1)(y + v_1 + s) + u_1\delta_1$$

2. *Computation of the conic Q*
   compute $u' := Res^*(E, C, y)/(u_1 u_2)$
   compute the inverse α_1 of $t - s^2 - h_2 + sh_1$ modulo u'
   compute the remainder v' of $\alpha_1(st - th_1 - f_3)$ by u'

3. *Computation of $D_1 + D_2$*
   $v_{D_1+D_2} := v'$
   $u_{D_1+D_2} := ((v^3 + v^2 h_1 + vh_2 - f_3)/(u'))^*$
   $D_1 + D_2 = [u_{D_1+D_2}, v_{D_1+D_2}]$

For a polynomial g, we used the notation g^* to symbolize the quotient of g by its leading coefficient.

Remark 3.1. One may wonder about the special choice of the divisor D^∞. It was chosen such that the conic Q be of the form $y - v$. It thus gives directly the second part of the Mumford representation $[u, v]$ of the final divisor. Other choices of the points P_1^∞, P_2^∞ imply using an auxiliary conic to find the representation.

3.3. *Flex case*

This case is particularly interesting for fast computations in the Jacobian. Indeed, the expressions involved in Algorithm 3.1 are very similar to those in the Picard curves [8] case, and decrease the number of operations.

As in the tangent case, we can assume (after a linear transformation) that $l^\infty : z = 0$ is tangent at the flex $P_1^\infty = P_2^\infty = P_4^\infty = (0 : 1 : 0)$. An equation of C is

$$y^3 + h_1 y^2 + h_2 y = f_4\,,$$

where $h_1, h_2, f_4 \in k[x]$ with $\deg(h_2) \leq 3, \deg(f_4) \leq 4$. Moreover P_1^∞ is a flex point with tangent $z = 0$ if and only if $\deg(h_1) \leq 1$ (consider the x-coordinates of the intersection $(l^\infty \cdot C)$).
In the same way as for the Lemma 3.1 we obtain

Lemma 3.2. *The cubic E is generically of the form*

$$y^2 + s \cdot y + t,$$

where s and t are polynomials in $k[x]$, with $\deg(s) \leq 1$ and $\deg(t) \leq 3$.
The conic Q is of the form

$$y - v,$$

where $v \in k[x]$ *and* $\deg(v) = 2$.

Let $D_i = [u_i, v_i]$ in Mumford representation. As in the tangent case, division with rest of $y^2 + sy + t$ by $y - v_i$ gives

$$y^2 + sy + t = (y - v_i)(y + v_i + s) + r_i$$

where $r_i \in k[x]$ and $\deg(r_i) \leq 4$. As the support of D_1 (resp. D_2) is contained in the support of $(C \cdot E)$ we have $r_i(x) = u_i(x)\delta_i(x)$ for some $\delta_i(x)$ of degree 1. The computation of E reduces on finding the polynomials s and δ_1 in $k[x]$. The advantage is that s and δ_1 have now degree 1. Computations are thus a lot easier: the linear system in step 1 consists only of four equations, and consequently, the resultant $\mathrm{res}(E, C, y)$ is easier to compute. In Algorithm 3.1 we just have to replace f_3 by f_4.

Furthermore, if $\mathrm{char}(k) \neq 3$, we let $Y = y + h_1(x)/3$ and we can assume that C is of the following form:

$$Y^3 + h_2 Y = f_4\,,$$

with h_2 and f_4 as above. If in addition $\mathrm{char}(k) \neq 2$, then we can assume that f_4 has no x^3 term.

3.4. *Comments on implementation*

We deal in this part with an optimized implementation in the case of the existence of a rational flex. To make the algorithm more efficient, we use the following well known methods:

(1) In order to reduce the number of field inversions, we use Montgomery's trick to compute simultaneous inversions. For the same reason, we compute almost inverses (using Bézout matrix), rather than inverses.
(2) We use either Karatsuba or Toom-Cook (in case $\mathrm{char}(k) \neq 2, 3, 5$) trick to multiply two polynomials, and we compute only the coefficients we need in the algorithm. For instance, as we only need to know the quotient of the resultant of E and C by u_1u_2, the degree ≤ 5 part of this resultant is irrelevant. Note that using Toom-Cook algorithm leads to divisions and multiplications by $2, 3$ and 5. These operations are not counted in the complexity since they are "easy".
(3) As explained in [2], one can try to use -2-adic expansion rather than usual 2-adic expansion, in order to save time for scalar multiplication. But this is only worthwhile if the computation of $-(D_1 + D_2)$ is easier than that of $D_1 + D_2$. This only happens in Theorem 1.1 if $P_1^\infty =$

$P_2^\infty = P_4^\infty$. In that case (and only in that case), this leads to a saving of at least 10% for the computation of scalar multiples mD, assuming a ratio of 10 : 1 for inversions and 2 : 3 for squarings, in relation to multiplications. This saving is not yet included in our algorithm.

We give in Tables 1, 2, 3, 4 and 5 the detailed and optimized operations in the case of existence of a rational flex and $\mathrm{char}(k) > 5$. In that case, an addition requires $148M+15SQ+2I$ and a doubling $165M+20SQ+2I$. The interested reader can find a program in MAGMA at the following webpage:

`http://www.math.uwaterloo.ca/~royono/Quartic.html`

If C has a rational hyperflex and $\mathrm{char}(k) > 5$, the nullity of an extra coefficient saves a couple of other operations. Addition then requires $131M+14SQ+2I$ and a doubling requires $148M+19SQ+2I$. Finally, note that the case of Picard curves has been handled in [8]. However, we point out that thanks to the new remarks made in this paper, we can actually reduce the cost for addition in the case of Picard curves to $116M+14SQ+2I$ and to $133M + 19SQ + 2I$ for doubling.

4. Examples

Fast additions can be useful in modern counting points algorithm and the two following examples are in this trend. The first example illustrates our algorithm in characteristic 2 and in the tangent case. Even without optimization, it is much faster than the existing (general) algorithm of MAGMA. The second case uses the optimized version with a flex.

4.1. *AGM-method*

In [21], a quasi-quadratic time algorithm for computing the Frobenius polynomial $\chi(X)$ of an ordinary non-hyperelliptic genus 3 curve C over $k = \mathbb{F}_{2^n}$ is described. However, the first part of the algorithm only gives $\chi(\pm X)$. Determining this sign can be done by checking for a generic degree 0 k-divisor D whether $\chi(1) \cdot D \sim 0$ or $\chi(-1) \cdot D \sim 0$.

Example 4.1. Let C over $k = \mathbb{F}_{2^n}$ with $n = 100$, be defined by

$$(\omega x^2+(\omega^3+1)y^2+\omega^2 z^2+\omega^4 xy+(\omega^3+\omega^2)xz+\omega^6 yz)^2-xyz(x+y+z) = 0,$$

where the generator ω of k is a root of $(X^{101} - 1)/(X - 1)$. In 2 minutes,

[21] gives us

$$\begin{aligned}\chi_C(\pm X) = X^6 &+ 377276036264709 \cdot X^5 \\ &+ 345535106116904583889422793740 3 \cdot X^4 \\ &+ 92979302197227669130776666646461687227769187 1 \cdot X^3 \\ &+ 345535106116904583889422793740 3 \cdot 2^{100} \cdot X^2 \\ &+ 377276036264709 \cdot 2^{200} \cdot X + 2^{300}.\end{aligned}$$

The line $z = 0$ is a bitangent at C at two rational points. We can now use the algorithm of Section 3.2 to prove in 4 seconds that the correct polynomial is $\chi(X)$. The same computation with MAGMA took 2 minutes.

4.2. 3-*dimensional factors of* $J^{new}(X_0(N))$

Let f be a newform of $X_0(N)$. Following a construction due to Shimura, one may associate to this newform a factor of $J_0(N)$ (the Jacobian of $X_0(N)$), denoted A_f. If $\dim A_f \leq 3$, it is easy to determine whether it is the Jacobian of a 'modular' curve C_f or not (see for example [11] or [13]). In particular, if $\dim A_f = 3$, and if the curve C_f is non-hyperelliptic, an equation of C_f seems to be often given by linear relations in $S_2(f)^{\otimes 4}$. On the other hand, thanks to the Eichler-Shimura relation, fast computation of Hecke operators T_p leads to a fast determination of $\#\tilde{A}_f(\mathbb{F}_p)$ where $\tilde{A}_f = A_f \otimes \mathbb{F}_p$ for primes $p \nmid N$ (c.f. [10]). In order to check that one obtains the right equation for the curve, one can check that the group of rational points of its Jacobian has the expected order n by computing $n \cdot D$ for a random rational degree 0 divisor D.

Example 4.2. We consider the modular curve $X_0(203)$. There is only one simple factor of dimension 3 in $J^{new}(X_0(203))$. We find one quartic relation between the associated cusp forms:

$$C : y^4-(x+3z)y^3+y^2(x^2-3xz+6z^2)+y(4xz^2-3z^3)-x^3z+3x^2z^2-4xz^3+2z^4 = 0$$

We let now $p = 25033$. We denote $\tilde{C} = C \otimes \mathbb{F}_p$ and $\tilde{C_f} = C_f \otimes \mathbb{F}_p$. The computation of the characteristic polynomial of T_p leads to $\#\operatorname{Jac}(\tilde{C_f})(\mathbb{F}_p) = 15692826275509$, which is prime.
The curve $\tilde{C}$ has a rational flex. After a linear transformation, and by denoting new coordinates still by x, y, z, we have

$$\begin{aligned}\tilde{C} \ : \ & y^3z + y^2(5057xz + 22616z^2) + y(6567x^3 + 18877x^2z + 162xz^2 + 14333z^3) \\ & = 8673x^4 + 24517x^3z + 20295x^2z^2 + 17815xz^3 + 3799z^4\end{aligned}$$

Choosing a random rational divisor, and computing its order, we may check in 0.14 seconds that, at least, $\#\operatorname{Jac}(\tilde{C}_f)(\mathbb{F}_p)$ divides $\#\operatorname{Jac}(\tilde{C})(\mathbb{F}_p)$.

5. Conclusion

We summarize here comparisons of the existing algorithms in the special case of genus 3 curves with a rational flex point. In particular, we did not include general algorithms for C_{ab} curves like in [14] since they only give asymptotic complexities.

We assume that $\operatorname{char}(k) > 5$. Such a curve has a rational model $y^3 + h_2 y = f_4$ with $\deg h_2(x) \leq 3$ and $\deg f_4(x) \leq 4$. We sort out the methods according to the degree of h_2.

Operation		hyperelliptic of genus 3	$C_{3,4}$ Picard	$C_{3,4}$ $\deg(h_2)=1$	$C_{3,4}$ $\deg(h_2)=2$	'general' quartic $\deg(h_2)=3$
Our	Add		2I+130M	2I+138M	2I+145M	2I+163M
Methods	Dbl		2I+152M	2I+160M	2I+167M	2I+185M
Previous	Add	I+70M [12]	2I+140M [4]	2I+147M [4]	2I+117M [22], 2I+150M [4]	
Work	Dbl	I+71M [12]	2I+164M [4]	2I+171M [4]	2I+129 M [22], 2I+174M [4]	

Some comments on this table:

- As far as we know, our algorithm is the fastest one for the 'general' genus 3 case.
- The algorithm [22] works also in characteristic 5 and is currently the fastest one for $C_{3,4}$ curves. Their method, which is a special case of [16] and [17], relies on a good choice of Riemann-Roch spaces and then has a geometric/algebraic flavor.
- In [4], the authors work in the function field of the curve, which allows them to use the tools from algorithmic number theory. In order to identify Jacobians and Class groups, they are restricted to work with a unique point at infinity.
- The algorithms [3] and [23] for Picard curves do not appear in this table as their point of view is different: they deal with the more general problem of reduction of divisors and they give only asymptotic complexity. We point out that a generalization of their method for C_{ab} curves based on geometric intersections, has been designed in [5].

6. Appendix

We show here how to compute the flexes of a plane algebraic curve $C : f(x_1, x_2, x_3) = 0$ of degree n over any algebraically closed field k of characteristic $p \geq 0$. Let P be a non-singular point of C. Recall that a point P is a *flex* if the intersection multiplicity at P of the tangent at P with C is greater than or equal to 3. This generalizes the definition given in Section 2.1. Non classical behaviors may appear when the characteristic divides $n-1$. For instance, there exist curves, called *funny curves*, for which all points are flexes (see for instance [15], where it is proved that a funny quartic is isomorphic to the Fermat quartic).

We are here interested in computational aspects of flexes. In characteristic 0, this is done by computing the Hessian.

Definition 6.1. Denote by f_i the derivative of f with respect to x_i. We call the *Hessian matrix* of f the matrix $(f_{ij})_{i,j}$ and we call its determinant $H(f)$ the *Hessian* of f.

The flexes are then the intersection points of the curve $H(f) = 0$ and C (see below Proposition 6.2). However, we shall see that this does not work when p divides $2(n-1)$. In [1], Abhyankar gives a method to overcome the difficulty when $p \neq 2$.

Proposition 6.1 ([1]). *Assume that $p \neq 2$ and that $P = (a : b : 1) \in C$. Then P is a flex if and only if $h(a, b) = 0$ with*

$$h(x_1, x_2) = \begin{vmatrix} f(x_1, x_2, 1) & f_1(x_1, x_2, 1) & f_2(x_1, x_2, 1) \\ f_1(x_1, x_2, 1) & f_{11}(x_1, x_2, 1) & f_{12}(x_1, x_2, 1) \\ f_2(x_1, x_2, 1) & f_{21}(x_1, x_2, 1) & f_{22}(x_1, x_2, 1) \end{vmatrix}.$$

We present here a method which works in any characteristic. We will need the following lemmas.

Lemma 6.1. *Let $g \in GL_3(k)$ be a linear transformation. Then $H(f \circ g^{-1}) = (\det g)^2 \cdot H(f) \circ g^{-1}$.*

Proof. Apply the chain rule. □

Lemma 6.2. $x_1^2 H(f) = \begin{vmatrix} n(n-1)f & (n-1)f_2 & (n-1)f_3 \\ (n-1)f_2 & f_{22} & f_{23} \\ (n-1)f_3 & f_{23} & f_{33} \end{vmatrix}$

Proof. Apply twice the Euler's formula $x_1 f_1 + x_2 f_2 + x_3 f_3 = (\deg f) f$. See for example [20]. □

If $f = 0$ is an equation of C of degree $n \geq 3$, then there exists a linear transformation g which sends a non-singular point $P = (p_1 : p_2 : p_3)$ on $(1 : 0 : 0)$ and its tangent to the line $x_3 = 0$. Then in affine coordinates

$$f \circ g^{-1} = x_2 + r x_2^2 + s x_2 x_3 + t x_3^2 + R(x_2, x_3) \tag{1}$$

and R has only terms of degree greater or equal to 3. Then P is a flex if and only if $r = 0$.

Proposition 6.2. *Suppose that p does not divide $2(n-1)$. Then P is a flex if and only if $H(f)(P) = 0$.*

Proof. Suppose that the x_1-coordinate of P is not 0 (otherwise do the same proof with an other coordinate). We have

$$(x_1^2 H(f) \circ g^{-1})(g(P)) = (\det g)^{-2} (x_1^2 H(f \circ g^{-1}))(g(P))$$

by Lemma 6.1 and because the $x_i x_j$ $(i, j \neq 1)$ terms in $(x_1^2) \circ g^{-1}$ are 0 at $g(P) = (1 : 0 : 0)$. Then by Lemma 6.2 and the form of $f \circ g^{-1}$

$$(x_1^2 H(f))(P) = -(\det g)^{-2} 2(n-1)^2 r.$$

So $H(f)(P) = 0$ if and only if $r = 0$ (i.e. P is a flex).

□

The proof shows also that this method can fail if p divides $2(n-1)$. We then suggest the following strategy. Denote K a complete local field of characteristic 0, $\mathcal{O}$ its ring of integers, $\mathcal{M}$ its maximal ideal such that $\mathcal{O}/\mathcal{M} \simeq k$ ($\mathcal{O}$ may be the ring of Witt vectors of k).

Proposition 6.3. *Let $\mathcal{C}/\mathcal{O}$ be a model of C given by a polynomial $F \in \mathcal{O}[X_1, X_2, X_3]$. We denote $\overline{H}$ the polynomial*

$$\overline{H} = \frac{X_1^2 H(F) - n(n-1) F (F_{22} F_{33} - F_{23}^2)}{2(n-1)^2}.$$

Then $\overline{H}$ is in $\mathcal{O}[X_1, X_2, X_3]$. We call $\overline{h}$ its reduction modulo $\mathcal{M}$.
Let $P = (1 : a : b) \in C$ be a non-singular point. The point P is a flex if and only if $\overline{h}(1, a, b) = 0$.

Proof. First we prove that $\overline{H}$ is in $\mathcal{O}[X_1, X_2, X_3]$. By Lemma 6.2,

$$X_1^2 H(F) - n(n-1)F(F_{22}F_{33} - F_{23}^2) = (n-1)^2(2F_2F_3F_{23} - F_2^2F_{33} - F_3^2F_{22}).$$

So $2(n-1)^2$ divides $X_1^2 H(F) - n(n-1)F(F_{22}F_{33} - F_{23}^2)$.
Since P is non-singular, there exists $\mathcal{P} = (1 : A : B) \in C(\mathcal{O})$ lifting P. Let $g \in \mathrm{GL}_3(\mathcal{O})$ a linear transformation that maps $\mathcal{P}$ on $(1 : 0 : 0)$ with tangent $X_3 = 0$. The reduction of this point is a flex if and only if the corresponding r (of equation (1)) is in $\mathcal{M}$. Now

$$\overline{H}(\mathcal{P}) = \frac{(X_1^2 H(F))(\mathcal{P})}{2(n-1)^2} = -\deg(g)^2 \cdot r$$

by the computations of Proposition 6.2. So P is a flex if and only if $\overline{h}(1, a, b) = 0$. □

Acknowledgment

The formula

$$(2F_2F_3F_{23} - F_2^2F_{33} - F_3^2F_{22})$$

appears already in [24] (Th.0.1). We are thankful to F. Voloch for this reference. We also want to thank J. Hirschfeld for pointing out a mistake in an earlier version and M. Girard for discussions on hyperflexes.

Table 1. **Addition,** $\deg u_1 = \deg u_2 = 3$

Input	$D_1 = [u_1, v_1]$ and $D_2 = [u_2, v_2]$ $u_i = x^3 + u_{i2}x^2 + u_{i1}x + u_{i0}, v_i = v_{i2}x^2 + v_{i1}x + v_{i0}$ $C : y^3 + h(x)y - f(x) = 0$ with $h(x) := h_3x^3 + h_2x^2 + h_1x + h_0,\ f(x) := x^4 + f_2x^2 + f_1x + f_0$
Output	$D = [u_{D_1+D_2}, v_{D_1+D_2}] = D_1 + D_2$ with $u_{D_1+D_2} = x^3 + u_2x^2 + u_1x + u_0$ $v_{D_1+D_2} = v_2x^2 + v_1x + v_0$

Step	Expression	Operations
1.1	compute the inverse t_1 of $v_1 - v_2$ modulo u_2 $a_1 = (v_{12} - v_{22})u_{22} - (v_{11} - v_{21}), a_2 = (v_{12} - v_{22})^2, a_3 = a_2u_{20} - a_1(v_{10} - v_{20})$; $a_4 = a_2(u_{22} + u_{21} + u_{20} + 1) - (v_{12} - v_{22} + a_1)(v_{12} + v_{11} + v_{10} - (v_{22} + v_{21} + v_{20})) - a_3$; $a_5 = a_4(v_{12} - v_{22}), a_6 = a_4(v_{11} - v_{21}) - a_3(v_{12} - v_{22})$; $a_7 = a_4^2, res_1 = a_7(v_{10} - v_{20}) - a_6a_3, t_{10} = a_1a_6, t_{12} = (v_{12} - v_{22})a_5$; $t_{11} = (a_1 + v_{12} - v_{22})(a_6 + a_5) - (t_{10} + t_{12}), t_{10} = t_{10} + a_7$; $t_1 = t_{12}x^2 + t_{11}x + t_{10}$	13M+2SQ
1.2	compute the remainder r of $(u_1 - u_2)t_1$ by u_2 $b_1 = (u_{12} + u_{11} + u_{10} - (u_{22} + u_{21} + u_{20}))(t_{12} + t_{11} + t_{10})$; $b_2 = (u_{12} - u_{11} + u_{10} - (u_{22} - u_{21} + u_{20}))(t_{12} - t_{11} + t_{10})$; $b_3 = (4(u_{12} - u_{22}) + 2(u_{11} - u_{21}) + u_{10} - u_{20})(4t_{12} + 2t_{11} + t_{10})$; $b_4 = (u_{12} - u_{22})t_{12}, b_5 = (u_{10} - u_{20})t_{10}, b_6 = (b_1 + b_2)/2 - (b_5 + b_4)$; $b_7 = ((b_3 + b_2 - b_1 - b_5)/2 - 2(4b_4 + b_6))/3, b_8 = b_1 - (b_5 + b_6 + b_7 + b_4)$; $b_9 = b_7 - b_4u_{22}, r_2 = b_5 - b_9u_{20}$; $b_{10} = b_4 + b_7 + b_6 + b_8 + b_5 - (b_9 + b_4)(u_{22} + u_{21} + u_{20} + 1)$; $r_1 = (b_{10} - (b_4 + b_6 + b_5 - (b_7 + b_8) - (b_9 - b_4)(u_{22} - u_{21} + u_{20} - 1)))/2$; $r_0 = b_{10} - (r_2 + r_1)$; $r = r_0x^2 + r_1x + r_2$	9M
1.3	compute the cubic $E = y^2 + sy + t$ $c_1 = v_{12}^2, c_2 = r_0c_1, c_3 = res_1 \cdot (v_{12} + v_{22}) - (r_1c_1) + (c_2u_{22}), c_4 = c_3 \cdot res_1,\ c_5 = res_1 \cdot r_0, c_6 = r_0c_2, c_7 = r_2c_3 - (c_6u_{20}) - c_5(v_{10} + v_{20})$; $c_8 = (r_0 + r_1 + r_2)(c_2 + c_3) - c_6(1 + (u_{22} + u_{21} + u_{20})) - c_5(v_{22} + v_{21} + v_{20} + v_{12} + v_{11} + v_{10}) - c_7$; $c_9 = c_4 + u_{12}c_5c_1 - v_{12}(c_8 + 2c_5v_{11}), c_{10} = c_5c_9, c_{11} = c_5^2$;	39M+3SQ+I
$*_1$	$c_{12} = c_9^2,\ c_{13} = c_{12} + h_3(-2c_{10} + h_3c_{11}),\ inv_1 = (c_{10}c_{13})^{-1},\ c_{14} = c_{13} \cdot inv_1,\ c_{15} = c_9c_{14},\ c_{16} = c_{12} \cdot inv_1 \cdot c_{10}$;	(7M+SQ+I)
	$s_0 = c_7c_{15},\ s_1 = c_8c_{15},\ c_{17} = c_4c_{15}$; $c_{18} = (1 + u_{12} + u_{11} + u_{10})(c_1 + c_{17}) - (v_{12} + v_{11} + v_{10})(v_{12} + v_{11} + v_{10} + s_1 + s_0),\ \ t_3 = c_9c_{15},\ t_0 = u_{10}c_{17} - v_{10}(v_{10} + s_0)$; $t_2 = (c_{18} + (-1 + u_{12} - u_{11} + u_{10})(-c_1 + c_{17}) - (v_{12} - v_{11} + v_{10})(v_{12} - v_{11} + v_{10} - s_1 + s_0))/2 - t_0$; $t_1 = c_{18} - (t_0 + t_2 + t_3),\ k_1 = c_{11}c_{14},\ c_{19} = t_0k_1,\ c_{20} = t_1k_1,\ c_{21} = t_2k_1$; $E = y^2 + (s_1x + s_0)y + t_3x^3 + t_2x^2 + t_1x + t_0$	
2.1	compute $res(E, C, y)$ and $u' := res(E, C, y)^*/(u_1u_2)$ $d_0 = c_{21}^2,\ d_1 = 3c_{21},\ d_2 = 3(c_{20} + d_0),\ d_3 = c_{21}(6c_{20} + d_0) + 3c_{19}$; $d_4 = s_1^2,\ d_5 = s_0^2,\ d_6 = (s_1 + s_0)^2 - (d_4 + d_5),\ d_7 = (s_1 + s_0)(t_3 + t_2 + t_1 + t_0)$; $d_8 = (s_0 - s_1)(t_2 + t_0 - (t_3 + t_1)),\ d_9 = (2s_1 + s_0)(8t_3 + 4t_2 + 2t_1 + t_0)$; $d_{10} = s_1t_3,\ d_{11} = s_0t_0,\ d_{12} = -(d_{11} + d_{10}) + (d_7 + d_8)/2$; $d_{13} = -2d_{10} + (d_{11} - d_7 + (d_9 - d_8)/3)/2,\ d_{14} = d_7 - (d_{11} + d_{12} + d_{13} + d_{10})$; $d_{15} = s_1d_4,\ d_{16} = 3d_4s_0,\ d_{17} = 1 - 3d_{10},\ d_{18} = d_{15} - 3d_{13}$; $d_{19} = f_2 + d_{16} + (1 - 3d_{10})f_2 - 3d_{12}$;	37M+5SQ
$*_2$	$d_{20} = (t_3 + t_2 + t_1 + t_0)(h_3 + h_2 + h_1 + h_0 + d_4 + d_6 + d_5 - 2(t_3 + t_2 + t_1 + t_0))$; $d_{21} = (-t_3 + t_2 - t_1 + t_0)(2(t_3 - t_2 + t_1 - t_0) - h_3 + h_2 - h_1 + h_0 + d_4 - d_6 + d_5)$; $d_{22} = (8t_3 + 4t_2 + 2t_1 + t_0)(8(-2t_3 + h_3) + 4(d_4 - 2t_2 + h_2) + 2(d_6 - 2t_1 + h_1) + d_5 - 2t_0 + h_0)$;	(15M)

Table 2. **Addition (cont.)**

	$d_{23} = (-8t_3 + 4t_2 - 2t_1 + t_0)(-8(-2t_3 + h_3) + 4(d_4 - 2t_2 + h_2) - 2(d_6 - 2t_1 + h_1) + d_5 - 2t_0 + h_0)$; $d_{24} = (27t_3 + 9t_2 + 3t_1 + t_0)(27(-2t_3 + h_3) + 9(d_4 - 2t_2 + h_2) + 3(d_6 - 2t_1 + h_1) + d_5 - 2t_0 + h_0)$; $d_{25} = t_0(d_5 - 2t_0 + h_0)$, $d_{26} = t_3(-2t_3 + h_3), d_{32} = f_2 s_1$; $d_{27} = -5d_{26} + ((-(d_{20} + d_{21}) + (3d_{25} + (d_{23} + d_{22})/2)/2)/2)/3$; $d_{28} = 15d_{26} + (((5d_{25} - 7d_{20} + (-d_{24} + 7d_{22} - d_{23} - d_{21})/2)/2)/2)/3$; $d_{29} = -3d_{26} + (((d_{20} - d_{25} + (d_{21} - d_{22} + (d_{24} - d_{23})/5)/2)/2)/2)/3$; $d_{33} = (h_3 + h_2 + h_1 + h_0)(d_{26} + d_{29} + d_{27} + d_{28} + s_1 + s_0 + d_{32})$; $d_{34} = (-h_3 + h_2 - h_1 + h_0)(-d_{26} + d_{29} - d_{27} + d_{28} + s_1 - s_0 + d_{32})$; $d_{35} = (8h_3 + 4h_2 + 2h_1 + h_0)(8d_{26} + 4(d_{29} + s_1) + 2(d_{27} + s_0) + d_{28} + d_{32})$; $d_{36} = (-8h_3 + 4h_2 - 2h_1 + h_0)(-8d_{26} + 4(d_{29} + s_1) - 2(d_{27} + s_0) + d_{28} + d_{32})$; $d_{37} = (27h_3 + 9h_2 + 3h_1 + h_0)(27d_{26} + 9(d_{29} + s_1) + 3(d_{27} + s_0) + d_{28} + d_{32})$; $d_{38} = h_0(d_{28} + d_{32})$, $d_{44} = h_3 d_{26}$; $d_{42} = -5d_{44} + ((-(d_{33} + d_{34}) + (3d_{38} + (d_{36} + d_{35})/2)/2)/2)/3$; $d_{41} = 15d_{44} + (((5d_{38} - 7d_{33} + (-d_{37} + 7d_{35} - d_{36} - d_{34})/2)/2)/2)/3$; $d_{43} = -3d_{44} + (((d_{33} - d_{38} + (d_{34} - d_{35} + (d_{37} - d_{36})/5)/2)/2)/2)/3$; $d_{40} = (d_{33} + d_{34})/2 - (d_{38} + d_{42} + d_{44})$; $d_{39} = d_{33} - (d_{38} + d_{40} + d_{41} + d_{42} + d_{43} + d_{44})$;	
	$d_{45} = k_1^3$, $d_{46} = d_{45}(d_{19} + d_{41}) + d_3$, $d_{47} = d_{45}(d_{18} + d_{42}) + d_2$; $d_{48} = d_{45}(d_{17} + d_{43}) + d_1$;	
$*_3$	$d_{46} = d_{46}c_{16}$, $d_{47} = d_{47}c_{16}$, $d_{48} = d_{48}c_{16}$;	(3M)
	$d_{49} = u_{12} + u_{22}$, $d_{50} = u_{21} + u_{11} + u_{12}u_{22}$; $d_{51} = u_{20} + u_{10} + u_{12}u_{21} + u_{11}u_{22}$, $u'_2 = d_{48} - d_{49}$; $u'_1 = d_{47} - d_{50} - d_{49}u'_2$, $u'_0 = -d_{49}u'_1 + d_{46} - d_{51} - d_{50}(d_{48} - d_{49})$; $u' = x^3 + u'_2 x^2 + u'_1 x + u'_0$	
2.2	compute the inverse α_1 of $t - s^2 - h$ modulo u' $g_1 = t_3 - h_3$, $g_0 = g_1(1 + u'_2 + u'_1 + u'_0)$, $g_2 = t_0 - (d_5 + h_0 + g_1 u'_0)$; $g_3 = t_3 + t_2 + t_1 + t_0 - (h_3 + h_2 + h_1 + h_0 + d_4 + d_6 + d_5 + g_0)$; $g_5 = (t_3 + t_2 + t_1 + t_0 - (h_3 + h_2 + h_1 + h_0 + d_4 + d_6 + d_5 + g_0) - (-t_3 + t_2 - t_1 + t_0 + h_3 - h_2 + h_1 - h_0 - d_4 + d_6 - d_5 - g_1(-1 + u'_2 - u'_1 + u'_0)))/2$; $g_6 = g_3 - g_5 - g_2$, $g_7 = g_6 u'_2 - g_5$, $g_8 = g_6^2$, $g_{10} = g_8 u'_0 - g_7 g_2$; $g_{11} = g_8(1 + u'_2 + u'_1 + u'_0) - (g_6 + g_7)(g_6 + g_5 + g_2) - g_{10}$, $g_{12} = g_{11} g_6$; $g_{13} = g_{11} g_5 - g_{10} g_6$, $g_9 = g_{11}^2$, $res_2 = g_9 g_2 - g_{13} g_{10}, \alpha_{10} = g_7 g_{13}$; $\alpha_{12} = g_6 g_{12}, \alpha_{11} = (g_6 + g_7)(g_{12} + g_{13}) - \alpha_{10} - \alpha_{12}$, $\alpha_{10} = \alpha_{10} + g_9$; $\alpha_1 = \alpha_{12} x^2 + \alpha_{11} x + \alpha_{10}$	16M+2SQ
2.3	compute the remainder v of $\alpha_1(st - f_4)$ by u' $i_1 = d_{10} - 1$, $i_2 = d_{13} - (d_{10} - 1)u'_2$, $i_3 = d_{11} - f_0 - i_2 u'_0$; $i_4 = d_{10} + d_{13} + d_{12} + d_{14} + d_{11} - (1 + f_2 + f_1 + f_0) - (i_2 + i_1)(u'_2 + u'_1 + u'_0 + 1)$; $i_5 = (i_4 - ((d_{10} - d_{13} + d_{12} - d_{14} + d_{11} - 1 - f_2 + f_1 - f_0) - (i_2 - i_1)(u'_2 - u'_1 + u'_0 - 1)))/2$, $i_6 = i_4 - i_3 - i_5$, $i_7 = (i_6 + i_5 + i_3)(\alpha_{12} + \alpha_{11} + \alpha_{10})$, $i_9 = i_6 \alpha_{12}$, $i_{10} = i_3 \alpha_{10}$, $i_8 = (i_6 - i_5 + i_3)(\alpha_{12} - \alpha_{11} + \alpha_{10})$, $i_{11} = (i_7 + i_8)/2 - (i_{10} + i_9)$; $i_{12} = (((4i_6 + 2i_5 + i_3)(4\alpha_{12} + 2\alpha_{11} + \alpha_{10}) - i_7 + i_8 - i_{10})/2 - 2(4i_9 + i_{11}))/3$; $i_{13} = i_7 - (i_{10} + i_{11} + i_{12} + i_9)$, $i_{14} = i_9$, $i_{15} = i_{12} - i_9 u'_2$, $i_{16} = i_{10} - i_{15} u'_0$; $i_{17} = (i_9 + i_{12} + i_{11} + i_{13} + i_{10}) - (i_{15} + i_{14})(u'_2 + u'_1 + u'_0 + 1)$; $i_{18} = (i_{17} - (i_9 - i_{12} + i_{11} - i_{13} + i_{10}) + (i_{15} - i_{14})(u'_2 - u'_1 + u'_0 - 1))/2$; $i_{19} = i_{17} - i_{16} - i_{18}$, $inv_2 = (res_2 \cdot i_{19})^{-1}$, $i_{20} = inv_2 \cdot i_{19}$; $v_0 = i_{20} i_{16}$, $v_1 = i_{20} i_{18}$, $v_2 = i_{20} i_{19}$; $v = v_2 x^2 + v_1 x + v_0$	18M+I
3	compute $u := u_{D_1 + D_2}$ $j_1 = inv_2 \cdot res_2^2$, $j_2 = j_1^3$, $j_3 = j_1 v_1$, $j_4 = j_3^2$, $j_5 = j_1 v_0$, $j_6 = j_3(j_4 + 6j_5)$;	16M+3SQ
$*_4$	$j_7 = (v_2 + v_1 + v_0)(h_3 + h_2 + h_1)$, $j_8 = (v_2 - v_1 + v_0)(h_3 - h_2 + h_1)$, $j_9 = v_2 h_3$; $j_{10} = v_0 h_1, j_{11} = (j_7 + j_8)/2 - (j_{10} + j_9)$, $j_{12} = 3j_3 + j_2 j_9$, $j_{14} = j_6 + j_2 j_{11}$; $j_{13} = 3(j_5 + j_4) - j_2 + j_2((((4v_2 + 2v_1 + v_0)(4h_3 + 2h_2 + h_1) - j_7 + j_8 - j_{10})/2 - 2(4j_9 + j_{11}))/3)$;	(8M)
	$u_2 = j_{12} - u'_2$, $u_1 = j_{13} - u'_1 - u'_2 u_2$, $u_0 = -u'_2 u_1 + j_{14} - u'_0 - u'_1(j_{12} - u'_2)$; $u = x^3 + u_2 x^2 + u_1 x + u_0$	
total		148M+15SQ+2I

Table 3. **Doubling,** $\deg u_1 = 3$

Input	$D_1 = [u_1, v_1]$ $u_1 = x^3 + u_{12}x^2 + u_{11}x + u_{10},\ v_1 = v_{12}x^2 + v_{11}x + v_{10}$ $C : y^3 + h(x)y - f(x) = 0$ with $h(x) := h_3x^3 + h_2x^2 + h_1x + h_0,\ f(x) := x^4 + f_2x^2 + f_1x + f_0$
Output	$D = [u_{2D_1}, v_{2D_1}] = 2D_1$ with $u_{2D_1} = x^3 + u_2x^2 + u_1x + u_0$ $v_{2D_1} = v_2x^2 + v_1x + v_0$

Step	Expression	Operations
1.1	compute w_1 such that $u_1w_1 = v_1^3 + h(x)v_1 - f(x)$ $l_1 = (v_{12} + v_{11} + v_{10})^2$, $l_2 = (v_{12} - v_{11} + v_{10})^2$, $l_3 = v_{12}^2$, $l_4 = v_{10}^2$; $l_5 = (l_1 + l_2)/2 - (l_4 + l_3)$; $l_6 = (((4v_{12} + 2v_{11} + v_{10})^2 - l_1 + l_2 - l_4)/2 - 2(4l_3 + l_5))/3$; $l_7 = l_1 - (l_4 + l_5 + l_6 + l_3), l_8 = (v_{12} + v_{11} + v_{10})(l_3 + l_6 + l_5 + l_7 + h_3 + h_2 + h_1),\ l_9 = (v_{12} - v_{11} + v_{10})(-l_3 + l_6 + h_3 - (l_5 + h_2) + l_7 + h_1)$; $l_{10} = (4v_{12} + 2v_{11} + v_{10})(8l_3 + 4(l_6 + h_3) + 2(l_5 + h_2) + l_7 + h_1)$; $l_{11} = (4v_{12} - 2v_{11} + v_{10})(-8l_3 + 4(l_6 + h_3) - 2(l_5 + h_2) + l_7 + h_1)$; $l_{12} = v_{10}(l_7 + h_1), l_{13} = v_{12}l_3, l_{14} = -5l_{13} + ((l_9 - l_8 + (l_{10} - l_{11})/2)/2)/3$; $l_{15} = ((-(l_8 + l_9) + (3l_{12} + (l_{10} + l_{11})/2)/2)/2)/3$; $l_{16} = (l_8 + l_9)/2 - (l_{12} + l_{15}), l_{14} = l_{14} - 1, w_{13} = l_{13}, w_{12} = l_{15} - w_{13}u_{12}$; $w_{11} = l_{14} - w_{13}u_{11} - w_{12}u_{12}, w_{10} = l_{16} - w_{13}u_{10} - w_{12}u_{11} - w_{11}u_{12}$; $w_1 = w_{13}x^3 + w_{12}x^2 + w_{11}x + w_{10}$	12M+5SQ
1.2	compute the inverse t_1 of w_1 modulo u_1 $a_1 = w_{13}$, $a_2 = w_{10} - a_1u_{10}$; $a_3 = w_{13} + w_{12} + w_{11} + w_{10} - a_1(1 + u_{12} + u_{11} + u_{10})$; $a_4 = (a_3 - (-w_{13} + w_{12} - w_{11} + w_{10} - a_1(-1 + u_{12} - u_{11} + u_{10})))/2$; $a_5 = a_3 - a_4 - a_2$, $a_6 = a_5u_{12} - a_4$, $a_7 = a_5^2$, $a_8 = a_7u_{10} - a_6a_2$; $a_9 = a_7(1 + u_{12} + u_{11} + u_{10}) - (a_5 + a_6)(a_5 + a_4 + a_2) - a_8$, $a_{10} = a_9a_5$; $a_{11} = a_9a_4 - a_8a_5$, $a_7 = a_9^2$, $res_1 = a_7a_2 - a_{11}a_8$, $t_{10} = a_6a_{11}$; $t_{12} = a_5a_{10}$, $t_{11} = (a_5 + a_6)(a_{10} + a_{11}) - t_{10} - t_{12}$, $t_{10} = t_{10} + a_7$; $t_1 = t_{12}x^2 + t_{11}x + t_{10}$	16M+2SQ
1.3	compute the remainder r of $(3v_1^2 + h)t_1$ by u_1 $b_1 = 3l_6 + h_3 - 3l_3u_{12}$, $b_2 = 3l_4 + h_0 - b_1u_{10}$; $b_3 = (3l_3 + 3l_6 + h_3 + 3l_5 + h_2 + 3l_7 + h_1 + 3l_4 + h_0) - (b_1 + 3l_3)(u_{12} + u_{11} + u_{10} + 1)$; $b_4 = (b_3 - ((3l_3 - (3l_6 + h_3) + 3l_5 + h_2 - (3l_7 + h_1) + 3l_4 + h_0) - (b_1 - 3l_3)(u_{12} - u_{11} + u_{10} - 1)))/2$; $b_5 = b_3 - b_2 - b_4$, $b_6 = (b_5 + b_4 + b_2)(t_{12} + t_{11} + t_{10})$; $b_7 = (b_5 - b_4 + b_2)(t_{12} - t_{11} + t_{10})$, $b_8 = b_5t_{12}$, $b_9 = b_2t_{10}$; $b_{10} = (b_6 + b_7)/2 - (b_9 + b_8)$; $b_{11} = (((4b_5 + 2b_4 + b_2)(4t_{12} + 2t_{11} + t_{10}) - b_6 + b_7 - b_9)/2 - 2(4b_8 + b_{10}))/3$; $b_{12} = b_6 - (b_9 + b_{10} + b_{11} + b_8)$, $b_{13} = b_{11} - b_8u_{12}$, $r_2 = b_9 - b_{13}u_{10}$; $b_{14} = (b_8 + b_{11} + b_{10} + b_{12} + b_9) - (b_{13} + b_8)(u_{12} + u_{11} + u_{10} + 1)$; $r_1 = (b_{14} - (b_8 + b_{10} + b_9) + (b_{11} + b_{12}) + (b_{13} - b_8)(u_{12} - u_{11} + u_{10} - 1))/2$; $r_0 = b_{14} - (r_2 + r_1)$; $r = r_0x^2 + r_1x + r_2$	13M
1.4	compute the cubic $E = y^2 + sy + t$ $c_1 = l_3$, $c_2 = r_0c_1$, $c_3 = 2res_1 \cdot v_{12} - (r_1c_1 - c_2u_{12})$, $c_4 = c_3 \cdot res_1$, $c_5 = res_1 \cdot r_0$, $c_6 = r_0c_2$, $c_7 = r_2c_3 - c_6u_{10} - 2c_5v_{10}$; $c_8 = (r_0 + r_1 + r_2)(c_2 + c_3) - c_6(1 + u_{12} + u_{11} + u_{10}) - 2c_5(v_{12} + v_{11} + v_{10}) - c_7$, $c_9 = c_4 + u_{12}c_5c_1 - v_{12}(c_8 + 2c_5v_{11})$, $c_{10} = c_5c_9$, $c_{11} = c_5^2$;	39M+2SQ+I
$*_1$	$c_{12} = c_9^2$, $c_{13} = c_{12} + h_3(-2c_{10} + h_3c_{11})$, $inv_1 = (c_{10}c_{13})^{-1}$, $c_{14} = c_{13} \cdot inv_1$, $c_{15} = c_9c_{14}$, $c_{16} = c_{12} \cdot inv_1 \cdot c_{10}$;	(7M+SQ+I)
	$s_0 = c_7c_{15}$, $s_1 = c_8c_{15}$, $c_{17} = c_4c_{15}$; $c_{18} = (1 + u_{12} + u_{11} + u_{10})(c_1 + c_{17}) - (v_{12} + v_{11} + v_{10})(v_{12} + v_{11} + v_{10} + s_1 + s_0),\ t_3 = c_9c_{15}$, $t_0 = u_{10}c_{17} - v_{10}(v_{10} + s_0)$; $t_2 = (c_{18} + (-1 + u_{12} - u_{11} + u_{10})(-c_1 + c_{17}) - (v_{12} - v_{11} + v_{10})(v_{12} - v_{11} + v_{10} - s_1 + s_0))/2 - t_0$; $t_1 = c_{18} - (t_0 + t_2 + t_3)$, $k_1 = c_{11}c_{14}$, $c_{19} = t_0k_1$, $c_{20} = t_1k_1$, $c_{21} = t_2k_1$; $E = y^2 + (s_1x + s_0)y + t_3x^3 + t_2x^2 + t_1x + t_0$	

Table 4. **Doubling (cont.)**

2.1	compute $res(E,C,y)$ and $u' := res(E,C,y)^*/(u_1u_2)$ $d_0 = c_{21}^2$, $d_1 = 3c_{21}$, $d_2 = 3(c_{20}+d_0)$, $d_3 = c_{21}(6c_{20}+d_0)+3c_{19}$; $d_4 = s_1^2$, $d_5 = s_0^2$, $d_6 = (s_1+s_0)^2-(d_4+d_5)$, $d_7 = (s_1+s_0)(t_3+t_2+t_1+t_0)$; $d_8 = (s_0-s_1)(t_2+t_0-(t_3+t_1))$, $d_9 = (2s_1+s_0)(8t_3+4t_2+2t_1+t_0)$; $d_{10} = s_1t_3$, $d_{11} = s_0t_0$, $d_{12} = -(d_{11}+d_{10})+(d_7+d_8)/2$; $d_{13} = -2d_{10}+(d_{11}-d_7+(d_9-d_8)/3)/2$, $d_{14} = d_7-(d_{11}+d_{12}+d_{13}+d_{10})$; $d_{15} = s_1d_4$, $d_{16} = 3d_4s_0$, $d_{17} = 1-3d_{10}$, $d_{18} = d_{15}-3d_{13}$; $d_{19} = f_2+d_{16}+(1-3d_{10})f_2-3d_{12}$;	35M+6SQ
$*_2$	$d_{20} = (t_3+t_2+t_1+t_0)(h_3+h_2+h_1+h_0+d_4+d_6+d_5-2(t_3+t_2+t_1+t_0))$; $d_{21} = (-t_3+t_2-t_1+t_0)(2(t_3-t_2+t_1-t_0)-h_3+h_2-h_1+h_0+d_4-d_6+d_5)$; $d_{22} = (8t_3+4t_2+2t_1+t_0)(8(-2t_3+h_3)+4(d_4-2t_2+h_2)+2(d_6-2t_1+h_1)+d_5-2t_0+h_0)$; $d_{23} = (-8t_3+4t_2-2t_1+t_0)(-8(-2t_3+h_3)+4(d_4-2t_2+h_2)-2(d_6-2t_1+h_1)+d_5-2t_0+h_0)$; $d_{24} = (27t_3+9t_2+3t_1+t_0)(27(-2t_3+h_3)+9(d_4-2t_2+h_2)+3(d_6-2t_1+h_1)+d_5-2t_0+h_0)$; $d_{25} = t_0(d_5-2t_0+h_0)$, $d_{26} = t_3(-2t_3+h_3)$, $d_{32} = f_2s_1$; $d_{27} = -5d_{26}+((-(d_{20}+d_{21})+(3d_{25}+(d_{23}+d_{22})/2)/2)/2)/3$; $d_{28} = 15d_{26}+(((5d_{25}-7d_{20}+(-d_{24}+7d_{22}-d_{23}-d_{21})/2)/2)/2)/3$; $d_{29} = -3d_{26}+(((d_{20}-d_{25}+(d_{21}-d_{22}+(d_{24}-d_{23})/5)/2)/2)/2)/3$; $d_{33} = (h_3+h_2+h_1+h_0)(d_{26}+d_{29}+d_{27}+d_{28}+s_1+s_0+d_{32})$; $d_{34} = (-h_3+h_2-h_1+h_0)(-d_{26}+d_{29}-d_{27}+d_{28}+s_1-s_0+d_{32})$; $d_{35} = (8h_3+4h_2+2h_1+h_0)(8d_{26}+4(d_{29}+s_1)+2(d_{27}+s_0)+d_{28}+d_{32})$; $d_{36} = (-8h_3+4h_2-2h_1+h_0)(-8d_{26}+4(d_{29}+s_1)-2(d_{27}+s_0)+d_{28}+d_{32})$; $d_{37} = (27h_3+9h_2+3h_1+h_0)(27d_{26}+9(d_{29}+s_1)+3(d_{27}+s_0)+d_{28}+d_{32})$; $d_{38} = h_0(d_{28}+d_{32})$, $d_{44} = h_3d_{26}$; $d_{42} = -5d_{44}+((-(d_{33}+d_{34})+(3d_{38}+(d_{36}+d_{35})/2)/2)/2)/3$; $d_{41} = 15d_{44}+(((5d_{38}-7d_{33}+(-d_{37}+7d_{35}-d_{36}-d_{34})/2)/2)/2)/3$; $d_{43} = -3d_{44}+(((d_{33}-d_{38}+(d_{34}-d_{35}+(d_{37}-d_{36})/5)/2)/2)/2)/3$; $d_{40} = (d_{33}+d_{34})/2-(d_{38}+d_{42}+d_{44})$; $d_{39} = d_{33}-(d_{38}+d_{40}+d_{41}+d_{42}+d_{43}+d_{44})$;	(15M)
	$d_{45} = k_1^3$, $d_{46} = d_{45}(d_{19}+d_{41})+d_3$, $d_{47} = d_{45}(d_{18}+d_{42})+d_2$; $d_{48} = d_{45}(d_{17}+d_{43})+d_1$;	
$*_3$	$d_{46} = d_{46}c_{16}$, $d_{47} = d_{47}c_{16}$, $d_{48} = d_{48}c_{16}$;	(3M)
	$d_{49} = 2u_{12}$, $d_{50} = 2u_{11}+u_{12}^2$, $d_{51} = 2u_{10}+2u_{12}u_{11}$, $u_2' = d_{48}-d_{49}$; $u_1' = d_{47}-d_{50}-d_{49}u_2'$, $u_0' = -d_{49}u_1'+d_{46}-d_{51}-d_{50}(d_{48}-d_{49})$; $u' = x^3+u_2'x^2+u_1'x+u_0'$	
2.2	compute the inverse α_1 of $t-s^2-h$ modulo u' $g_1 = t_3-h_3$, $g_0 = g_1(1+u_2'+u_1'+u_0')$, $g_2 = t_0-(d_5+h_0+g_1u_0')$; $g_3 = t_3+t_2+t_1+t_0-(h_3+h_2+h_1+h_0+d_4+d_6+d_5+g_0)$; $g_5 = (t_3+t_2+t_1+t_0-(h_3+h_2+h_1+h_0+d_4+d_6+d_5+g_0)-(-t_3+t_2-t_1+t_0+h_3-h_2+h_1-h_0-d_4+d_6-d_5-g_1(-1+u_2'-u_1'+u_0'))))/2$; $g_6 = g_3-g_5-g_2$, $g_7 = g_6u_2'-g_5$, $g_8 = g_6^2$, $g_{10} = g_8u_0'-g_7g_2$; $g_{11} = g_8(1+u_2'+u_1'+u_0')-(g_6+g_7)(g_6+g_5+g_2)-g_{10}$, $g_{12} = g_{11}g_6$; $g_{13} = g_{11}g_5-g_{10}g_6$, $g_9 = g_{11}^2$, $res_2 = g_9g_2-g_{13}g_{10}$, $\alpha_{10} = g_7g_{13}$; $\alpha_{12} = g_6g_{12}$, $\alpha_{11} = (g_6+g_7)(g_{12}+g_{13})-\alpha_{10}-\alpha_{12}$, $\alpha_{10} = \alpha_{10}+g_9$; $\alpha_1 = \alpha_{12}x^2+\alpha_{11}x+\alpha_{10}$	16M+2SQ
2.3	compute the remainder v of $\alpha_1(st-f_4)$ by u' $i_1 = d_{10}-1$, $i_2 = d_{13}-(d_{10}-1)u_2'$, $i_3 = d_{11}-f_0-i_2u_0'$; $i_4 = d_{10}+d_{13}+d_{12}+d_{14}+d_{11}-(1+f_2+f_1+f_0)-(i_2+i_1)(u_2'+u_1'+u_0'+1)$; $i_5 = (i_4-((d_{10}-d_{13}+d_{12}-d_{14}+d_{11}-1-f_2+f_1-f_0)-(i_2-i_1)(u_2'-u_1'+u_0'-1))))/2$; $i_6 = i_4-i_3-i_5$, $i_7 = (i_6+i_5+i_3)(\alpha_{12}+\alpha_{11}+\alpha_{10})$, $i_9 = i_6\alpha_{12}$, $i_{10} = i_3\alpha_{10}$; $i_8 = (i_6-i_5+i_3)(\alpha_{12}-\alpha_{11}+\alpha_{10})$, $i_{11} = (i_7+i_8)/2-(i_{10}+i_9)$; $i_{12} = (((4i_6+2i_5+i_3)(4\alpha_{12}+2\alpha_{11}+\alpha_{10})-i_7+i_8-i_{10})/2-2(4i_9+i_{11}))/3$; $i_{13} = i_7-(i_{10}+i_{11}+i_{12}+i_9)$, $i_{14} = i_9$, $i_{15} = i_{12}-i_9u_2'$, $i_{16} = i_{10}-i_{15}u_0'$; $i_{17} = (i_9+i_{12}+i_{11}+i_{13}+i_{10})-(i_{15}+i_{14})(u_2'+u_1'+u_0'+1)$; $i_{18} = (i_{17}-(i_9-i_{12}+i_{11}-i_{13}+i_{10})+(i_{15}-i_{14})(u_2'-u_1'+u_0'-1))/2$; $i_{19} = i_{17}-i_{16}-i_{18}$, $inv_2 = (res_2\cdot i_{19})^{-1}$, $i_{20} = inv_2\cdot i_{19}$; $v_0 = i_{20}i_{16}$, $v_1 = i_{20}i_{18}$, $v_2 = i_{20}i_{19}$; $v = v_2x^2+v_1x+v_0$	18M+I

Table 5. **Doubling (cont.)**

3	compute $u := u_{2D_1}$ $j_1 = inv_2 \cdot res_2^2, j_2 = j_1^3, j_3 = j_1 v_1, j_4 = j_3^2, j_5 = j_1 v_0, j_6 = j_3(j_4 + 6j_5);$	16M+3SQ
$*_4$	$j_7 = (v_2 + v_1 + v_0)(h_3 + h_2 + h_1)$, $j_8 = (v_2 - v_1 + v_0)(h_3 - h_2 + h_1)$, $j_9 = v_2 h_3$; $j_{10} = v_0 h_1, j_{11} = (j_7 + j_8)/2 - (j_{10} + j_9)$, $j_{12} = 3j_3 + j_2 j_9$, $j_{14} = j_6 + j_2 j_{11}$; $j_{13} = 3(j_5 + j_4) - j_2 + j_2((((4v_2 + 2v_1 + v_0)(4h_3 + 2h_2 + h_1) - j_7 + j_8 - j_{10})/2 - 2(4j_9 + j_{11}))/3);$	(8M)
	$u_2 = j_{12} - u_2'$, $u_1 = j_{13} - u_1' - u_2' u_2$, $u_0 = -u_2' u_1 + j_{14} - u_0' - u_1'(j_{12} - u_2')$; $u = x^3 + u_2 x^2 + u_1 x + u_0$	
total		165M+20SQ+2I

Table 6. If $h_3 = 0$ then replace $*_1, *_2, *_3$ and $*_4$ by

$*_1$	$inv_1 = c_{10}^{-1}, c_{14} = inv_1, c_{15} = inv_1 \cdot c_9, c_{16} = 1;$	(M+I)
$*_2$	$d_{27} = (t_3 + t_2 + t_1)(h_1 + h_2 + d_4 + d_6 - 2(t_3 + t_2 + t_1));$ $d_{28} = (t_3 - t_2 + t_1)(h_1 - h_2 + d_6 - d_4 - 2(t_3 - t_2 + t_1));$ $d_{29} = (4t_3 + 2t_2 + t_1)(-8t_3 + 2(d_4 - 2t_2 + h_2) + d_6 - 2t_1 + h_1);$ $d_{30} = -2t_3^2, d_{31} = t_1(d_6 - 2t_1 + h_1), d_{32} = (d_{27} + d_{28})/2 - (d_{31} + d_{30});$ $d_{33} = ((d_{29} - d_{27} + d_{28} - d_{31})/2 - 2(4d_{30} + d_{32}))/3;$ $d_{35} = (h_2 + h_1 + h_0)(d_{30} + d_{33} + d_{32} + s_0 + s_1);$ $d_{36} = (h_2 - h_1 + h_0)(d_{30} - d_{33} + d_{32} + s_0 - s_1);$ $d_{37} = (4h_2 + 2h_1 + h_0)(4d_{30} + 2(d_{33} + s_1) + d_{32} + s_0);$ $d_{43} = h_2 d_{30}, d_{39} = h_0(d_{32} + s_0), d_{41} = (d_{35} + d_{36})/2 - (d_{39} + d_{43});$ $d_{42} = ((d_{37} - d_{35} + d_{36} - d_{39})/2 - 2(4d_{43} + d_{41}))/3;$ $d_{40} = d_{35} - (d_{39} + d_{41} + d_{42} + d_{43}), d_{44} = 0;$	(9M+SQ)
$*_3$		
$*_4$	$j_{11} = v_2 h_2, j_{12} = 3j_3, j_{13} = 3(j_5 + j_4) - j_2 + j_2 j_{11};$ $j_{14} = j_6 + j_2((v_2 + v_1)(h_2 + h_1) - (v_1 h_1 + j_{11}));$	(5M)

Table 7. If $h_3, h_2 = 0$ then replace $*_1, *_2, *_3$ and $*_4$ by

$*_1$	$inv_1 = c_{10}^{-1}, c_{14} = inv_1, c_{15} = inv_1 \cdot c_9, c_{16} = 1;$	(M+I)
$*_2$	$d_{37} = -2t_3^2, d_{35} = t_2(d_4 - 2t_2), d_{38} = 0, d_{39} = 0, d_{43} = 0, d_{44} = 0;$ $d_{36} = (t_3 + t_2)(d_4 - 2(t_3 + t_2)) - (d_{35} + d_{37}), d_{42} = h_1 d_{37}, d_{40} = h_0(d_{36} + s_1), \quad d_{41} = (h_1 + h_0)(d_{37} + d_{36} + s_1) - (d_{40} + d_{42});$	(5M+SQ)
$*_3$		
$*_4$	$j_{12} = 3j_3, j_{13} = 3(j_5 + j_4) - j_2, j_{14} = j_6 + j_2(h_1 v_2);$	(2M)

Table 8. If $h_3, h_2, h_1 = 0$ then replace $*_1, *_2, *_3$ and $*_4$ by

$*_1$	$inv_1 = c_{10}^{-1}, c_{14} = inv_1, c_{15} = inv_1 \cdot c_9, c_{16} = 1;$	(M+I)
$*_2$	$d_{41} = -2h_0 t_3^2, d_{38} = 0, d_{39} = 0, d_{40} = 0, d_{42} = 0, d_{43} = 0, d_{44} = 0;$	(M+SQ)
$*_3$		
$*_4$	$j_{12} = 3j_3, j_{13} = 3(j_5 + j_4) - j_2, j_{14} = j_6;$	

Table 9. If $h_3, h_2, h_1, h_0 = 0$ then replace $*_1, *_2, *_3$ and $*_4$ by

$*_1$	$inv_1 = c_{10}^{-1}, c_{14} = inv_1, c_{15} = inv_1 \cdot c_9, c_{16} = 1;$	(M+I)
$*_2$	$d_{38} = 0, d_{39} = 0, d_{40} = 0, d_{41} = 0, d_{42} = 0, d_{43} = 0, d_{44} = 0;$	
$*_3$		
$*_4$	$j_{12} = 3j_3, j_{13} = 3(j_5 + j_4) - j_2, j_{14} = j_6;$	

References

1. S. Abhyankar. Remark on Hessians and flexes. *Nieuw Arch. Wisk.*, 11:110–117, 1963.
2. R. M. Avanzi, G. Frey, T. Lange and R. Oyono. On using expansions to the base of -2. *Inter. J. of. Comp. Math.*, 81(4):403–406, 2004.
3. E. Reinaldo Barreiro, J. Estrada Sarlabous and J. P. Cherdieu. Efficient reduction on the Jacobian variety of Picard curves. In *Coding theory, cryptography and related areas (Guanajuato, 1998)*, pages 13–28. Springer, Berlin, 2000.
4. A. Basiri, A. Enge, J-C. Faugère and N. Gürel. Implementing the Arithmetic of $C_{3,4}$ Curves. In *Algorithmic Number Theory Symposium - ANTS-VI*, volume 3076 of *LNCS*, pages 87–101. Springer, 2004.
5. R. Blache, J. Estrada Sarlabous and M. Petkova. A geometric interpretation of reduction in the Jacobian of C_{ab} curves. preprint.
6. C. Diem and E. Thomé. Index calculus in class groups of non-hyperelliptic curves of genus three. accepted at J. of Cryptology, 2007.
7. E. W. Howe, K. E. Lauter and J. Top. Pointless curves of genus three and four. In *Algebra, Geometry, and Coding Theory (AGCT 2003) (Y. Aubry and G. Lachaud, eds.)*, volume 11 of *Séminaires et Congrés*. Société Mathématique de France, Paris, 2005.
8. S. Flon and R. Oyono. Fast arithmetic on Jacobians of Picard curves. In *Public Key Cryptography - PKC 2004*, volume 2947 of *LNCS*, pages 55–68. Springer, 2004.
9. S. Flon, R. Oyono and C. Ritzenthaler. Rationality of the intersection points of a line with a plane quartic. In progress, 2007.
10. G. Frey and M. Müller. Arithmetic of modular curves and applications. In *Algorithmic Algebra and Number Theory*, pages 11–48. Ed. Matzat et al., Springer-Verlag, Berlin, 1999.
11. S. D. Galbraith. *Equations for modular curves*. PhD thesis, Oxford, 1996.
12. M. Gonda, K. Matsuo, K. Aoki, J. Chao and S. Tsujii. Improvements of addition algorithm on genus 3 hyperelliptic curves and their implementations. In *SCIS 2004*, 2004.
13. E. González-Jiménez and R. Oyono. Non-hyperelliptic modular curves of genus 3. preprint, 2007.
14. R. Harasawa and J. Suzuki. Fast Jacobian group arithmetic on C_{ab} curves. *in Algorithmic Number Theory Symposium - ANTS-IV*, 1838:359–376, 2000. Springer.
15. M. Homma. Funny plane curves in characteristic $p > 0$. *Comm. Algebra*, 15:1469–1501, 1987.
16. K. Khuri-Makdisi. Linear algebra algorithms for divisor on an algebraic curve. *Math. of Computations*, 73:333–357, 2004.
17. K. Khuri-Makdisi. asymptotically fast group operations on Jacobian of general curves. 2006. Available on `http://arxiv.org/abs/math.NT/0409209`.
18. K. Miura and H. Yoshihara. Field theory for function fields of plane quartic curves. *J. of Algebra*, 226:283–294, 2000.

19. E. Nart and C. Ritzenthaler. Non hyperelliptic curves of genus three over finite fields of characteristic two. *J. of Number Theory*, 116:443–473, 2006.
20. G. Orzech and M. Orzech. *Plane algebraic curves*, volume 61. Pure and Applied Math., New-York, 1981.
21. C. Ritzenthaler. Point counting on genus 3 non hyperelliptic curves. In *Algorithmic Number Theory Symposium - ANTS-VI*, volume 3076 of *LNCS*, pages 379–394. Springer, 2004.
22. F. Abu Salem and K. Khuri-Makdisi. Fast Jacobian group operations for $C_{3,4}$ curves over a large finite field. Available at `http://www-lb.cs.aub.edu.lb/fa21/articleOct3.pdf`.
23. J. Estrada Sarlabous, E. Reinaldo Barreiro and J. A. Piñeiro Barceló. On the Jacobian varieties of Picard curves: explicit addition law and algebraic structure. *Math. Nachr.*, 208:149–166, 1999.
24. K.-O. Stöhr and J. F. Voloch. Weierstrass points and curves over finite fields. *Proc. Lond. Math. Soc., III. Ser.*, 52:1–19, 1986.
25. F. Torres. The approach of Stöhr-Voloch to the Hasse-Weil bound with applications to optimal curves and plane arcs. Available on `http://arxiv.org/abs/math.AG/0011091`, 2000.
26. A.M. Vermeulen. *Weierstrass points of weight two on curves of genus* 3. PhD thesis, Universiteit van Amsterdam, 1983.

Computing endomorphism rings of Jacobians of genus 2 curves over finite fields

David Freeman

University of California, Berkeley
E-mail: dfreeman@math.berkeley.edu

Kristin Lauter

Microsoft Research
E-mail: klauter@microsoft.com

We present probabilistic algorithms which, given a genus 2 curve C defined over a finite field and a quartic CM field K, determine whether the endomorphism ring of the Jacobian J of C is the full ring of integers in K. In particular, we present algorithms for computing the field of definition of, and the action of Frobenius on, the subgroups $J[\ell^d]$ for prime powers ℓ^d. We use these algorithms to create the first implementation of Eisenträger and Lauter's algorithm for computing Igusa class polynomials via the Chinese Remainder Theorem [11], and we demonstrate the algorithm for a few small examples. We observe that in practice the running time of the CRT algorithm is dominated not by the endomorphism ring computation but rather by the need to compute p^3 curves for many small primes p.

1. Introduction

Many public-key cryptographic protocols are based on the difficulty of the discrete logarithm problem in groups of points on elliptic curves and Jacobians of hyperelliptic curves. For such protocols one needs to work in a subgroup of large prime order of the Jacobian of the curve, so it is useful to be able to construct curves over finite fields whose Jacobians have a specified number of points.

The problem of constructing elliptic curves over finite fields with a given number of points has been studied extensively. Current solutions rely on computing the j-invariant via the construction of the Hilbert class polynomial for a quadratic imaginary field. There are three different approaches to computing the Hilbert class polynomial: a complex-analytic algorithm [2], [12]; a Chinese Remainder Theorem algorithm [5], [1]; and a p-adic al-

gorithm [9], [4]. The best running time for these algorithms is $\tilde{O}(|d|)$, where d is the discriminant of the quadratic imaginary field [12], [4].

Analogous methods exist for constructing genus 2 curves with a given number of points on their Jacobians. In this case, the solutions rely on computing the curves' Igusa invariants via the computation of Igusa class polynomials for quartic CM fields. Again there are three different approaches: a complex-analytic algorithm [26], [28], [29], [7]; a Chinese Remainder Theorem algorithm [11]; and a p-adic algorithm [15]. These algorithms are less extensively developed than their elliptic curve analogues, and to date there is no running time analysis for any of them.

In this paper we study the implementation of Eisenträger and Lauter's Chinese Remainder Theorem algorithm [11]. The algorithm takes as input a primitive quartic CM field K, i.e. a purely imaginary quadratic extension of a real quadratic field with no proper imaginary quadratic subfields, and produces the Igusa class polynomials of K. The basic outline of the algorithm is as follows:

(1) Define S to be a set of primes with certain splitting behavior in the field K and its reflex field K^*.
(2) For each prime p in S:
   (a) For each triple $(i_1, i_2, i_3) \in \mathbb{F}_p^3$ of Igusa invariants, construct a genus 2 curve C over $\mathbb{F}_p$ corresponding to that triple.
   (b) Check the isogeny class of each curve. For each curve in the desired isogeny class, compute the endomorphism ring of the Jacobian of the curve and keep only those curves for which the endomorphism ring is the full ring of integers $\mathcal{O}_K$.
   (c) Construct the Igusa class polynomials mod p from the triples collected in Step 2b.
(3) Use the Chinese Remainder Theorem or the Explicit CRT [3] to construct the Igusa polynomials either with rational coefficients or modulo a prime of cryptographic size.

One advantage of the CRT algorithm over other algorithms for computing Igusa class polynomials is that the CRT algorithm does not require that the real quadratic subfield have class number one.

Our contribution is to provide an efficient probabilistic algorithm for computing endomorphism rings of Jacobians of genus 2 curves over small prime fields. Using this algorithm to compute endomorphism rings, we have implemented a probabilistic version of the full Eisenträger-Lauter CRT al-

gorithm (Algorithm 7.1) in MAGMA and used it to compute Igusa class polynomials for several fields K with small discriminant.

It was previously believed that computing endomorphism rings would be the bottleneck in the genus 2 CRT algorithm. Our results are surprising in the sense that we find that the time taken to compute the endomorphism rings with our probabilistic algorithms is negligible compared with the time needed to compute p^3 genus 2 curves via Mestre's algorithm for each small prime p. For example, for $K = \mathbb{Q}(i\sqrt{13+2\sqrt{13}})$ and $p = 157$, the largest prime for which endomorphism rings are computed for this K, our (unoptimized) MAGMA program takes about 52 minutes to loop through 157^3 curves and find 243 curves in the specified isogeny class. Our probabilistic algorithm (also implemented in MAGMA) applied to these 243 curves then takes 16.5 *seconds* to find the single curve whose Jacobian has endomorphism ring equal to $\mathcal{O}_K$.

The algorithm works as follows. Let C be a genus 2 curve over a finite field $\mathbb{F}_p$, and let J be its Jacobian; we assume J is ordinary. Let K be a primitive quartic CM field, which we assume is given via an embedding in $\mathbb{C}$. The first test is whether $\mathrm{End}(J)$, the endomorphism ring of J, is an order in $\mathcal{O}_K$. This computation is outlined in [11, Section 5] and described in more detail in Section 2 below. If $\mathrm{End}(J)$ is an order in $\mathcal{O}_K$, we compute a set of possible elements $\pi \in \mathcal{O}_K$ that could represent the Frobenius endomorphism of J. If π represents the Frobenius endomorphism, then its complex conjugate $\overline{\pi}$ represents the Verschiebung endomorphism.

We next determine a set $\{\alpha_i\}$ of elements of $\mathcal{O}_K$ such that $\mathbb{Z}[\pi, \overline{\pi}, \{\alpha_i\}] = \mathcal{O}_K$. It follows that $\mathrm{End}(J) = \mathcal{O}_K$ if and only if each α_i is an endomorphism of J. We show in Section 3 that we can take each α_i to have one of two forms: either $\alpha_i = \frac{\pi^k - 1}{\ell}$ for some positive integer k and prime ℓ, or $\alpha_i = \frac{h_i(\pi)}{\ell^d}$ for some cubic polynomial h_i with integer coefficients and some prime power ℓ^d. In Section 4 we show how to determine whether an element of the first form is an endomorphism; this is equivalent to determining the field of definition of the ℓ-torsion points of J. In Section 5 we show how to determine whether an element of the second form is an endomorphism; this is equivalent to computing the action of Frobenius on a basis of $J[\ell^d]$. The main results are Algorithms 4.3 and 5.1, two very efficient probabilistic algorithms which check fields of definition and compute the action of Frobenius, respectively. The running times of these algorithms depend primarily on the sizes of the fields over which the points of $J[\ell^d]$ are defined. Section 6 provides upper bounds for these sizes in terms of the prime ℓ and the size of the base field p.

A detailed statement of the Eisenträger-Lauter CRT algorithm, incorporating the algorithms of Sections 2, 4, and 5, appears in Section 7. Section 8 describes various ways in which we have modified our MAGMA implementation to improve the algorithm's performance. Finally, in Section 9 we give examples of our algorithm run on several small quartic CM fields.

Notation and assumptions

Throughout this paper, a *curve* will refer to a smooth, projective, absolutely irreducible algebraic curve C. The Jacobian of C, denoted $\mathrm{Jac}(C)$, is an abelian variety of dimension g, where g is the genus of C. We assume throughout that p is a prime, and that $\mathrm{Jac}(C)$ is an ordinary abelian variety modulo p.

A number field K is a *CM field* if it is a totally imaginary quadratic extension of a totally real field. We denote by K^* the reflex field of K, and by K_0 the real quadratic subfield of K. A CM field is *primitive* if it has no proper CM subfields. We will assume unless otherwise noted that K is a primitive quartic CM field not isomorphic to $\mathbb{Q}(\zeta_5)$. This implies that K is either Galois cyclic or non-Galois. If K is Galois cyclic, then $K^* = K$; if K is non-Galois, then K^* is another primitive quartic CM field [25, p. 64]. A curve C has *CM by* K if the endomorphism ring of $\mathrm{Jac}(C)$ is isomorphic to an order in $\mathcal{O}_K$, the ring of integers of the CM field K.

Acknowledgments

This research was conducted during the first author's internship at Microsoft Research, Redmond, during the summer of 2006. The first author thanks Microsoft for its hospitality and Denis Charles, Jean-Marc Couveignes, and Edward Schaefer for many helpful discussions. The second author thanks Pierrick Gaudry for helpful correspondence and pointers to his code. Both authors thank Reinier Bröker, David Kohel, and Christophe Ritzenthaler for their feedback on previous versions of this paper.

2. Computing zeta functions and the Frobenius element

To determine whether the Jacobian J of a given genus 2 curve C has endomorphism ring equal to $\mathcal{O}_K$, the first step is to determine whether the endomorphism ring is even an order in $\mathcal{O}_K$. This is accomplished by computing the characteristic polynomial of Frobenius, to see if the Frobenius element corresponds to an algebraic integer of K. This in turn is equivalent to determining the zeta function of C, which can be computed by

finding the number of points on the curve and its Jacobian, $n = \#C(\mathbb{F}_p)$ and $m = \#J(\mathbb{F}_p)$. For a given field K there are several possibilities for the pairs (n, m), as described in [11, Prop. 4].

In this section we give an explicit algorithm that determines whether $\mathrm{End}(J)$ is an order in $\mathcal{O}_K$ and if so, gives a set $S \subset \mathcal{O}_K$ of possibilities for the Frobenius endomorphism of J. The main point is to find the possible Frobenius elements by finding generators of certain principal ideals (Step 2) with absolute value equal to $\sqrt{p}$ (Step 4a).

Algorithm 2.1. Let K be a primitive quartic CM field and K^* the reflex of K. The following algorithm takes as input the field K, a prime p that splits completely in K and splits completely into principal ideals in K^*, and a curve C defined over the finite field $\mathbb{F}_p$. The algorithm returns `true` or `false` according to whether $\mathrm{End}(J)$ is an order in $\mathcal{O}_K$, where $J = \mathrm{Jac}(C)$. If the answer is `true`, the algorithm also outputs a set $S \subset \mathcal{O}_K$ that consists of the $\mathrm{Aut}(K/\mathbb{Q})$-orbit of the Frobenius endomorphism of J.

(1) Compute the decomposition $p = \mathfrak{p}_1\mathfrak{p}_2\mathfrak{p}_3\mathfrak{p}_4$ in $\mathcal{O}_K$, using e.g. [6, Alg. 6.2.9]. Renumber so that $\mathfrak{p}_2 = \overline{\mathfrak{p}_1}$ and $\mathfrak{p}_3 = \overline{\mathfrak{p}_4}$.
(2) Compute generators α_1 and α_2 for the principal ideals $\mathfrak{p}_1\mathfrak{p}_3$ and $\mathfrak{p}_2\mathfrak{p}_3$, respectively, using e.g. [6, Alg. 6.5.10].
(3) Compute a fundamental unit u of K_0 with $|u| > 1$, using e.g. [6, Alg. 5.7.1].
(4) For $i \leftarrow 1, 2$, do the following:
   (a) If $|\alpha_i| < \sqrt{p}$, set $\alpha_i \leftarrow \alpha_i u$ until $|\alpha_i| = \sqrt{p}$. If $|\alpha_i| > \sqrt{p}$, set $\alpha_i \leftarrow \alpha_i u^{-1}$ until $|\alpha_i| = \sqrt{p}$.
   (b) Compute the characteristic polynomial $h_i(x)$ of α_i, using e.g. [6, Prop. 4.3.4].
   (c) If K is Galois and $h_1(x) = h_2(-x)$, set $\alpha_2 \leftarrow -\alpha_2$ and $h_2(x) \leftarrow h_2(-x)$.
   (d) Set $(n_{i,+1}, m_{i,+1}) \leftarrow (p + 1 - \frac{h_i'(0)}{p}, h_i(1))$. Set $(n_{i,-1}, m_{i,-1}) \leftarrow (p + 1 + \frac{h_i'(0)}{p}, h_i(-1))$.
(5) Determine whether the Frobenius endomorphism of J has characteristic polynomial equal to $h_i(\pm x)$ for some i:
   (a) Choose a random point $P \in J(\mathbb{F}_p)$ and compute $Q_{j,\tau} = [m_{i,\tau}]P$ for $i \in \{1, 2\}$, $\tau \in \{\pm 1\}$. If none of $Q_{i,\tau}$ is the identity, return `false`. Otherwise, optionally repeat with another random point P.
   (b) If J passes a certain fixed number of trials of Step 5a, compute $\#C(\mathbb{F}_p)$. If $\#C(\mathbb{F}_p) \neq n_{i,\tau}$ for all $i \in \{1, 2\}$, $\tau \in \{\pm 1\}$, return

`false`.

(c) If $\#C(\mathbb{F}_p) = n_{i,\tau}$, compute $\#J(\mathbb{F}_p)$, using e.g. Baby Step Giant Step [6, Alg 5.4.1]. If $\#J \neq m_{i,\tau}$ for the same i, τ, return `false`.

(6) If K is Galois, output $S = \{\tau\alpha_1, \tau\overline{\alpha_1}, \tau\alpha_2, \tau\overline{\alpha_2}\}$. If K is not Galois, output $S = \{\tau\alpha_i, \tau\overline{\alpha_i}\}$, using the i determined in Step 5c.

(7) Return `true`.

Proof. The proof of [11, Prop. 4] shows that the ideals $\mathfrak{p}_1\mathfrak{p}_3$ and $\mathfrak{p}_2\mathfrak{p}_3$ are principal and the Frobenius endomorphism of J corresponds to a generator of one of these ideals or their complex conjugates. Furthermore, this generator must have complex absolute value $\sqrt{p}$. The generators determined in Step 2 are unique up to unit multiple, so Step 4a ensures that the absolute values are $\sqrt{p}$, thus making each α_i unique up to complex conjugation and sign.

If the Frobenius element corresponds to α_i or $\overline{\alpha_i}$, then $h_i(x)$ is the characteristic polynomial of Frobenius, so we can determine this case by checking whether $\#C(\mathbb{F}_p) = n_{i,+1}$ and $\#J(\mathbb{F}_p) = m_{i,+1}$. Similarly, if the Frobenius element corresponds to $-\alpha_i$ or $-\overline{\alpha_i}$, then $h_i(-x)$ is the characteristic polynomial of Frobenius, so we can determine this case by checking whether $\#C(\mathbb{F}_p) = n_{i,-1}$ and $\#J(\mathbb{F}_p) = m_{i,-1}$.

If K is Galois (with Galois group C_4), then the ideal (α_2) is equal to $(\alpha_1)^\sigma$ for some σ generating the Galois group. Since complex absolute value squared is the same as the norm from K to its real quadratic subfield K_0, $|\alpha_1| = \sqrt{p}$ implies that $|\alpha_1^\sigma| = \sqrt{p}$. Since α_1^σ and α_2 both generate (α_2) and have absolute value $\sqrt{p}$, we deduce that $\alpha_1^\sigma = \pm\alpha_2$. Step 4c ensures that this sign is positive, so α_1 and α_2 have the same characteristic polynomial $h_i(x)$, and thus the Frobenius element could be any of the elements output by Step 6. Since $\mathrm{Aut}(K/\mathbb{Q})$ is generated by σ and σ^2 is complex conjugation, we have output the $\mathrm{Aut}(K/\mathbb{Q})$-orbit of the Frobenius element.

If K is not Galois, then the Frobenius element must be either α_i or $\overline{\alpha_i}$. Since $\mathrm{Aut}(K/\mathbb{Q})$ in this case consists of only the identity and complex conjugation, Step 6 outputs the $\mathrm{Aut}(K/\mathbb{Q})$-orbit of the Frobenius element. □

3. Constructing a generating set for $\mathcal{O}_K$

Given the Jacobian J of a genus 2 curve over $\mathbb{F}_p$ and a primitive quartic CM field K, Algorithm 2.1 allows us to determine whether there is some $\pi \in \mathcal{O}_K$ that represents the Frobenius endomorphism of J. Since the complex conjugate $\overline{\pi}$ represents the Verschiebung endomorphism, if Algorithm 2.1

outputs `true` then we have

$$\mathbb{Z}[\pi, \overline{\pi}] \subseteq \mathrm{End}(J) \subseteq \mathcal{O}_K. \tag{1}$$

In this section, we assume we are given a $J/\mathbb{F}_p$ and a π such that (1) holds, and we wish to determine whether $\mathrm{End}(J) = \mathcal{O}_K$.

Let $\mathcal{B}$ be a $\mathbb{Z}$-module basis for $\mathcal{O}_K$, and consider the collection of elements $\{\alpha \in \mathcal{B} \backslash \mathbb{Z}\}$. Since this collection generates $\mathcal{O}_K$ over $\mathbb{Z}[\pi, \overline{\pi}]$, it suffices to determine whether or not each element of the collection is an endomorphism of J. Assuming K satisfies some mild hypotheses, Eisenträger and Lauter give one example of a basis $\mathcal{B}$ that suffices to determine the endomorphism ring [11, Lemma 6]. However, the method given in [11] lacks an efficient procedure for testing whether a given $\alpha \in \mathcal{B}$ is an endomorphism of J.

In this section, we derive from an arbitrary basis $\mathcal{B}$ a set of generators for $\mathcal{O}_K$ over $\mathbb{Z}[\pi, \overline{\pi}]$ that is convenient in the sense that there is an efficient probabilistic algorithm (Algorithm 4.3 or Algorithm 5.1) for determining whether an element of the set is an endomorphism of J. Our findings are summarized in Proposition 3.8.

We begin by observing that since $K = \mathbb{Q}(\pi)$, any $\alpha \in \mathcal{O}_K$ can be expressed as a polynomial $f \in \mathbb{Q}[\pi]$. Since π satisfies a polynomial of degree 4 (the characteristic polynomial of Frobenius), f can be taken to have degree 3. We may thus write

$$\alpha = \frac{a_0 + a_1\pi + a_2\pi^2 + a_3\pi^3}{n} \tag{2}$$

for some integers a_0, a_1, a_2, a_3, n. We assume that a_0, a_1, a_2, a_3 have no common factor with n, so that n is the smallest integer such that $n\alpha \in \mathbb{Z}[\pi]$.

Remark 3.1. The LLL lattice reduction algorithm [19], as implemented by the MAGMA command `LinearRelation`, finds an expression of the form (2) for any $\alpha \in \mathcal{O}_K$. Given as input the sequence $[1, \pi, \pi^2, \pi^3, -\alpha]$, the algorithm outputs a sequence $[a_0, a_1, a_2, a_3, n]$ satisfying the relation (2).

The following lemma shows that each $\alpha \in \mathcal{B} \setminus \mathbb{Z}$ can be replaced with a collection of elements that generate the same ring, each with a power of a single prime in the denominator of the expression (2).

Lemma 3.2. *Let $A \subset B$ be commutative rings with 1, with $[B : A]$ finite. Suppose $\alpha \in B$, and let n be the smallest integer such that $n\alpha \in A$. Suppose*

n factors into primes as $\ell_1^{d_1} \cdots \ell_r^{d_r}$. Then

$$A[\alpha] = A\Big[\tfrac{n}{\ell_1^{d_1}}\alpha, \ldots, \tfrac{n}{\ell_r^{d_r}}\alpha\Big].$$

Proof. Clearly the ring on the right is contained in the ring on the left, so we must show that α is contained in the ring on the right. It suffices to show that there are integers c_i such that

$$c_1 \frac{n}{\ell_1^{d_1}} + \cdots + c_r \frac{n}{\ell_r^{d_r}} = 1, \tag{3}$$

for then we can multiply this identity by α to get our result. We use the extended Euclidean algorithm and induct on r, the number of distinct primes dividing n. If $r = 1$ the result is trivial, for in this case $n/\ell_1^{d_1} = 1$. Now suppose (3) holds for any n that is divisible by r distinct primes. If n' is divisible by $r+1$ distinct primes, we can write $n' = n\ell_{r+1}^{d_{r+1}}$ for some n divisible by r distinct primes. Since ℓ_{r+1} is relatively prime to n, we can use the extended Euclidean algorithm to write $a\ell_{r+1}^{d_{r+1}} + bn = 1$ for some integers a, b. We can then multiply the first term by the left-hand side of (3) (which is equal to 1) to get

$$ac_1 \frac{n\ell_{r+1}^{d_{r+1}}}{\ell_1^{d_1}} + \cdots + ac_r \frac{n\ell_{r+1}^{d_{r+1}}}{\ell_r^{d_r}} + bn = ac_1 \frac{n'}{\ell_1^{d_1}} + \cdots + ac_r \frac{n'}{\ell_r^{d_r}} + b\frac{n'}{\ell_{r+1}^{d_{r+1}}} = 1.$$

This is an equation of the form (3) for n', which completes the proof. □

The next lemma shows that only primes dividing the index $[\mathcal{O}_K : \mathbb{Z}[\pi]]$ appear in the denominators.

Lemma 3.3. *Let α be an element of $\mathcal{O}_K$, and suppose n is the smallest integer such that $n\alpha \in \mathbb{Z}[\pi]$. Then n divides the index $[\mathcal{O}_K : \mathbb{Z}[\pi]]$.*

Proof. Let $N = [\mathcal{O}_K : \mathbb{Z}[\pi]]$. By definition, N is the size of the abelian group $\mathcal{O}_K/\mathbb{Z}[\pi]$. Thus we can write any $\alpha \in \mathcal{O}_K$ as $\alpha = a + b$ with $b \in \mathbb{Z}[\pi]$ and $N \cdot a \in \mathbb{Z}[\pi]$. This shows that $\mathcal{O}_K$ is contained in $\frac{1}{N}\mathbb{Z}[\pi]$. We may thus write $\alpha = f(\pi)/N$ for a unique polynomial f with integer coefficients and degree at most 3. Furthermore, since $n\alpha$ is the smallest multiple of α in $\mathbb{Z}[\pi]$, we may write $\alpha = g(\pi)/n$ for a unique polynomial g with integer coefficients and degree at most 3, such that n has no factor in common with all the coefficients of g. We thus have $n \cdot f(\pi) = N \cdot g(\pi)$. If we let d be the gcd of the coefficients of f and e be the gcd of the coefficients of g, then we have $n \cdot d = N \cdot e$.

Let ℓ be a prime dividing e. Since $\gcd(n, e) = 1$, ℓ must divide d, so we can cancel ℓ from both sides and get $n \cdot d' = N \cdot e'$ with $e' < e$. Proceeding in this manner until $e' = 1$, we conclude that n divides N. □

We now know that each $\alpha \in \mathcal{B} \setminus \mathbb{Z}$ can be replaced with a collection of elements $\{\frac{n}{\ell_i^{d_i}}\alpha\}$, and the only ℓ_i appearing are divisors of the index the index $[\mathcal{O}_K : \mathbb{Z}[\pi]]$. The following lemma and corollary show that for any ℓ which divides $[\mathcal{O}_K : \mathbb{Z}[\pi]]$ exactly (i.e. $\ell \mid [\mathcal{O}_K : \mathbb{Z}[\pi]]$ and $\ell^2 \nmid [\mathcal{O}_K : \mathbb{Z}[\pi]]$), the element $\frac{n}{\ell}\alpha$ can be replaced by an element of the form $\frac{\pi^k - 1}{\ell}$. This replacement is useful since by [11, Fact 10], determining whether an element of the form $\frac{\pi^k - 1}{\ell}$ is an endomorphism is equivalent to testing the field of definition of the ℓ-torsion.

Lemma 3.4. *Let $A \subset B \subset C$ be abelian groups, with $[C : A]$ finite. Let ℓ be a prime, and suppose ℓ divides $[C : A]$ and ℓ^2 does not divide $[C : A]$. Suppose there is some $\beta \in B$ such that $\beta \notin A$ and $\ell\beta \in A$. Then for any $\alpha \in C$ such that $\ell\alpha \in A$, $\alpha \in B$.*

Proof. The hypotheses on $[C : A]$ imply that the ℓ-primary part of C/A (denoted $(C/A)_\ell$) is isomorphic to $\mathbb{Z}/\ell\mathbb{Z}$, so $(B/A)_\ell$ is either trivial or $\mathbb{Z}/\ell\mathbb{Z}$. The conditions on β imply that β has order ℓ in B/A, so $(B/A)_\ell \cong \mathbb{Z}/\ell\mathbb{Z} \cong (C/A)_\ell$, with the isomorphism induced by the inclusion map $B \hookrightarrow C$. Since α is in the ℓ-primary part of C/A, α must also be in the ℓ-primary part of B/A, so $\alpha \in B$. □

Corollary 3.5. *Suppose ℓ divides $[\mathcal{O}_K : \mathbb{Z}[\pi]]$ exactly and $\beta = \frac{\pi^k - 1}{\ell} \notin \mathbb{Z}[\pi]$. Then $\frac{\pi^k - 1}{\ell}$ is an endomorphism of J if and only if any $\alpha \in \mathcal{O}_K \setminus \mathbb{Z}[\pi]$ with $\ell\alpha \in \mathbb{Z}[\pi]$ is also an endomorphism.*

Proof. The result follows directly from Lemma 3.4, with $A = \mathbb{Z}[\pi]$, $B = \mathrm{End}(J)$, and $C = \mathcal{O}_K$. □

Furthermore, if $p \nmid [\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$, then any element α_i with denominator $\ell_i = p$ may be ignored due to the following corollary.

Corollary 3.6. *Suppose $p \nmid [\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$. Then for any $\alpha \in \mathcal{O}_K$ such that $p\alpha \in \mathbb{Z}[\pi]$, $\alpha \in \mathbb{Z}[\pi, \overline{\pi}]$.*

Proof. Since π is the Frobenius element, it satisfies a characteristic polynomial of the form

$$\pi^4 + s_1\pi^3 + s_2\pi^2 + s_1 p\pi + p^2 = 0.$$

Using $\pi\overline{\pi} = p$ and dividing this equation by π gives

$$\pi^3 + s_1\pi^2 + s_2\pi + s_1 p + p\overline{\pi} = 0. \tag{4}$$

From this equation we see that $p\overline{\pi} \in \mathbb{Z}[\pi]$, so either $[\mathbb{Z}[\pi, \overline{\pi}] : \mathbb{Z}[\pi]] = p$ or $\overline{\pi} \in \mathbb{Z}[\pi]$. If $\overline{\pi} \in \mathbb{Z}[\pi]$ then p divides the coefficients of all the terms on the left hand side of (4), which it does not, so we deduce that $\overline{\pi} \notin \mathbb{Z}[\pi]$ and $[\mathbb{Z}[\pi, \overline{\pi}] : \mathbb{Z}[\pi]] = p$. The hypothesis $p \nmid [\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ thus implies that p divides $[\mathcal{O}_K : \mathbb{Z}[\pi]]$ exactly, so we may apply Lemma 3.4 with $\ell = p$, $A = \mathbb{Z}[\pi]$, $B = \mathbb{Z}[\pi, \overline{\pi}]$, $C = \mathcal{O}_K$, and $\beta = \overline{\pi}$. □

Thus any α satisfying the conditions of the corollary is automatically an endomorphism. We now show that the condition $p \nmid [\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ is automatically satisfied for all primes p except possibly 2 and 3.

Proposition 3.7. *Suppose $p > 3$ and that $\pi \in \mathcal{O}_K$ corresponds to the Frobenius endomorphism of an ordinary abelian surface A over $\mathbb{F}_p$. Then $p \nmid [\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$.*

Proof. Let $\Delta(R)$ denote the discriminant of a $\mathbb{Z}$-module R. This proposition follows from [18, Proposition 9.4], which shows that

$$\Delta(\mathbb{Z}[\pi, \overline{\pi}]) = \pm \mathrm{Norm}_{K/\mathbb{Q}}(\pi - \overline{\pi})\, \Delta(\mathbb{Z}[\pi + \overline{\pi}]).$$

Alternatively, it is shown in [20, Proposition 7.4] that any prime that divides the index $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ must divide either $[\mathcal{O}_{K_0} : \mathbb{Z}[\pi + \overline{\pi}]]$ or $\frac{\Delta(\mathcal{O}_{K_0}[\pi])}{\Delta(\mathcal{O}_K)}$. Using [18, Theorem 1.3], the second quantity is prime to p if the abelian surface is ordinary. The same proposition also shows that $\Delta(\mathbb{Z}[\pi+\overline{\pi}]) < 16p$, and since

$$\frac{\Delta(\mathbb{Z}[\pi + \overline{\pi}])}{\Delta(\mathcal{O}_{K_0})} = [\mathcal{O}_{K_0} : \mathbb{Z}[\pi + \overline{\pi}]]^2,$$

we conclude that if p divides $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ then p^2 divides $[\mathcal{O}_{K_0} : \mathbb{Z}[\pi+\overline{\pi}]]^2$, and thus

$$p^2 \leq \frac{\Delta(\mathbb{Z}[\pi + \overline{\pi}])}{\Delta(\mathcal{O}_{K_0})} < \frac{16p}{5}$$

(since a real quadratic field has discriminant at least 5), which implies $p \leq 3$. □

The following proposition summarizes the results of this section.

Proposition 3.8. *Suppose $\{\alpha_i\}$ generates $\mathcal{O}_K$ as a $\mathbb{Z}$-algebra. Let n_i be the smallest integer such that $n_i\alpha_i \in \mathbb{Z}[\pi]$, and write the prime factorization of n_i as $n_i = \prod_j \ell_{ij}^{d_{ij}}$. For each (i,j) with $\ell_{ij} \neq p$, let k_{ij} be an integer such that $\pi^{k_{ij}} - 1 \in \ell_{ij}\mathcal{O}_K$. Suppose $p > 3$. Then the following set generates $\mathcal{O}_K$ over $\mathbb{Z}[\pi, \overline{\pi}]$:*

$$\left\{ \frac{n_i}{\ell_{ij}^{d_{ij}}}\alpha_i : \ell_{ij}^2 \mid [\mathcal{O}_K : \mathbb{Z}[\pi]] \right\} \cup \left\{ \frac{\pi^{k_{ij}} - 1}{\ell_{ij}} : \ell_{ij}^2 \nmid [\mathcal{O}_K : \mathbb{Z}[\pi]], \ell_{ij} \neq p \right\}.$$

Remark 3.9. Proposition 3.8 shows that if $p > 3$ and the index $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ is square-free, then $\mathcal{O}_K$ can be generated over $\mathbb{Z}[\pi, \overline{\pi}]$ by a collection of elements of the form $\frac{\pi^k - 1}{\ell}$. This answers a question raised by Eisenträger and Lauter [11, Remark 5].

In our application, $\pi \in \mathcal{O}_K$ is only determined up to an automorphism of K, but Proposition 3.8 can still be used to determine a generating set for $\mathcal{O}_K$.

Corollary 3.10. *Let $\mathcal{S} \subset \mathcal{O}_K$ be the set given in Proposition 3.8. Let σ be an element of* $\mathrm{Aut}(K/\mathbb{Q})$. *Then the set $\{\beta^\sigma : \beta \in \mathcal{S}\}$ generates $\mathcal{O}_K$ over $\mathbb{Z}[\pi^\sigma, \overline{\pi}^\sigma]$.*

Proof. By Proposition 3.8, the set $\{\pi, \overline{\pi}\} \cup \mathcal{S}$ generates $\mathcal{O}_K$ as a $\mathbb{Z}$-algebra. Since $\mathcal{O}_K$ is mapped to itself by $\mathrm{Aut}(K/\mathbb{Q})$, the set $\{\pi^\sigma, \overline{\pi}^\sigma\} \cup \{\beta^\sigma : \beta \in \mathcal{S}\}$ also generates $\mathcal{O}_K$ as a $\mathbb{Z}$-algebra. The statement follows immediately. □

4. Determining fields of definition

In this section, we consider the problem of determining the field of definition of the n-torsion points of the Jacobian J of a genus 2 curve over $\mathbb{F}_p$. By [11, Fact 10], the n-torsion points of J are defined over $\mathbb{F}_{p^k}$ if and only if $(\pi^k - 1)/n$ is an endomorphism of J, where π is the Frobenius endomorphism of J. Thus determining the field of definition of the ℓ-torsion points allows us to determine whether some of the elements given by Proposition 3.8 are endomorphisms.

Algorithm 4.1. The following algorithm takes as input a primitive quartic CM field K, an element $\pi \in \mathcal{O}_K$ with $\pi\overline{\pi} = p$, and an integer n with $\gcd(n, p) = 1$, and outputs the smallest integer k such that $\pi^k - 1 \in n\mathcal{O}_K$. If J is the Jacobian of a genus 2 curve over $\mathbb{F}_p$ with Frobenius π^σ for some $\sigma \in \mathrm{Aut}(K/\mathbb{Q})$ and $\mathrm{End}(J) = \mathcal{O}_K$, this integer k is such that the n-torsion points of J are defined over $\mathbb{F}_{p^k}$.

(1) Compute a $\mathbb{Z}$-basis $\mathcal{B} = (1, \delta, \gamma, \kappa)$ of $\mathcal{O}_K$, using [27] or [6, Algorithm 6.1.8], and write $\pi = (a, b, c, d)$ in this basis. Set $k \leftarrow 1$.
(2) Let $\bar{\mathcal{B}}$ be the reduction of the elements of $\mathcal{B}$ modulo n. Let $(a_1, b_1, c_1, d_1) = (a, b, c, d) \pmod{n}$.
(3) Compute $\pi^k \equiv (a_k, b_k, c_k, d_k) \pmod{n}$ with respect to $\bar{\mathcal{B}}$.
(4) If $(a_k, b_k, c_k, d_k) \equiv (1, 0, 0, 0) \pmod{n}$, output k. Otherwise set $k \leftarrow k + 1$ and go to Step 3.

Proof. The set $\bar{\mathcal{B}}$ is a $\mathbb{Z}/n\mathbb{Z}$-basis of $\mathcal{O}_K/n\mathcal{O}_K$, so if $\pi^k \equiv (1, 0, 0, 0) \pmod{n}$, then $\pi^k - 1 \in n\mathcal{O}_K$ (since the first element of $\bar{\mathcal{B}}$ is 1). Since $n\mathcal{O}_K$ is mapped to itself by $\mathrm{Aut}(K/\mathbb{Q})$, we have $(\pi^\sigma)^k - 1 \in n\mathcal{O}_K$. If $\mathrm{End}(J) = \mathcal{O}_K$, then $\frac{(\pi^\sigma)^k - 1}{n} \in \mathcal{O}_K = \mathrm{End}(J)$, so by [11, Fact 10], $J[n] \subset J(\mathbb{F}_{p^k})$. □

Remark 4.2. Since $J[n] = \bigoplus J[\ell^d]$ for prime powers ℓ^d dividing n, we may speed up Algorithm 4.1 by factoring n and computing $k(\ell^d)$ for each prime power factor ℓ^d; then $k(n) = \mathrm{lcm}(k(\ell^d))$. Furthermore, we will see in Propositions 6.2 and 6.3 below that for a fixed ℓ^d, the possible values of k are very limited. Thus we may speed up the algorithm even further by precomputing these possible values and testing each one, rather than increasing the value of k by 1 until the correct value is found.

Eisenträger and Lauter [11] computed endomorphism rings in several examples by determining the group structure of $J(\mathbb{F}_{p^k})$ to decide whether $J[n] \subset J(\mathbb{F}_{p^k})$. This is an exponential-time algorithm that is efficient only for very small k. Eisenträger and Lauter also suggested that the algorithm of Gaudry-Harley [14] could be used to determine the field of definition of the n-torsion points. One of the primary purposes of this article is to present an efficient probabilistic algorithm to test the field of definition of $J[n]$. Below we describe the various methods of testing the field of definition of the n-torsion of J. Since $J[n] = \bigoplus J[\ell^d]$ as ℓ^d ranges over maximal prime-power divisors of n, it suffices to consider each prime-power factor separately. We thus assume in what follows that $n = \ell^d$ is a prime power.

4.1. *The brute force method*

The simplest method of determining the field of definition of the n-torsion is to compute the abelian group structure of $J(\mathbb{F}_{p^k})$. The MAGMA syntax for this computation is straightforward, and the program returns a group structure of the form

$$J(\mathbb{F}_{p^k}) \cong \frac{\mathbb{Z}}{a_1\mathbb{Z}} \times \cdots \times \frac{\mathbb{Z}}{a_j\mathbb{Z}},$$

with $a_1 \mid \cdots \mid a_j$. The n-torsion of J is contained in $J(\mathbb{F}_{p^k})$ if and only if $j = 4$ and n divides a_1.

While this method is easy to implement, if k is too large it may take too long to compute the group structure (via Baby-Step/Giant-Step or similar algorithms), or even worse we may not even be able to factor $\#\operatorname{Jac}(C)(\mathbb{F}_{p^k})$. In practice, computing group structure in MAGMA seems to be feasible for group sizes up to roughly 2^{200}, which means p^k should be no more than roughly 2^{100}, and thus k will have to be very small. Thus the brute force method is very limited in scope; however, it has the advantage that in the small cases it can handle it runs fairly quickly and always outputs the right answer.

4.2. *The Gaudry-Harley-Schost method*

Gaudry and Harley [14] define a Schoof-Pila-like algorithm for counting points on genus 2 curves. The curves input to this algorithm are assumed to have a degree 5 model over $\mathbb{F}_q$, so we can write elements of the Jacobian as pairs of affine points minus twice the Weierstrass point at infinity. An intermediate step in the algorithm is to construct a polynomial $R(x) \in \mathbb{F}_q[x]$ with the following property: if P_1 and P_2 are points on C such that $D = [P_1] + [P_2] - 2[\infty]$ is an n-torsion point of J, then the x-coordinates of P_1 and P_2 are roots of R. The field of definition of the x-coordinates is at most a degree-two extension of the field of definition of D. Thus in many cases the field of definition of the n-torsion points can be determined from the factorization of $R(x)$.

Gaudry has implemented the algorithm in MAGMA and NTL; the algorithm involves taking two resultants of pairs of two-variable polynomials of degree roughly n^2. The algorithm uses the clever trick of computing a two-variable resultant by computing many single-variable resultants and interpolating the result. The interpolation only works if the field of definition of J has at least $4n^2 - 8n + 4$ elements, so we must base extend J until the field of definition is large enough. Since $R(x)$ has coefficients in $\mathbb{F}_p$, this base extension has no effect on the result of the computation.

Gaudry and Harley's analysis of the algorithm gives a running time of $\tilde{O}(n^6)$ field multiplications if fast polynomial arithmetic is used, and $O(n^8)$ otherwise. Due to its large space requirements, the algorithm has only succeeded at handling inputs of size $n \leq 19$ [16].

4.3. *A probabilistic method*

As usual, we let J be the Jacobian of a genus 2 curve over $\mathbb{F}_{p^k}$, and $\ell \neq p$ be a prime. Let H be the ℓ-primary part of $J(\mathbb{F}_{p^k})$. Then H has the structure

$$H = \frac{\mathbb{Z}}{\ell^{\alpha_1}\mathbb{Z}} \times \frac{\mathbb{Z}}{\ell^{\alpha_2}\mathbb{Z}} \times \frac{\mathbb{Z}}{\ell^{\alpha_3}\mathbb{Z}} \times \frac{\mathbb{Z}}{\ell^{\alpha_4}\mathbb{Z}},$$

with $\alpha_1 \leq \alpha_2 \leq \alpha_3 \leq \alpha_4$. Our test rests on the following observations:

- If the ℓ^d-torsion points of J are defined over $\mathbb{F}_{p^k}$, then $\alpha_1 \geq d$, and the number of ℓ^d-torsion points in H is ℓ^{4d}.
- If the ℓ^d-torsion points of J are not defined over $\mathbb{F}_{p^k}$, then $\alpha_1 < d$, and the number of ℓ^d-torsion points in H is at most ℓ^{4d-1}.

We thus make the following calculation: write $\#J(\mathbb{F}_{p^k}) = \ell^s m$ with $\ell \nmid m$. Choose a random point $P \in J$. Then $[m]P \in H$, and we test whether $[\ell^d m]P = O$ in J. If the ℓ^d-torsion points of J are defined over $\mathbb{F}_{p^k}$, then $[\ell^d m]P = O$ with probability $\rho = \ell^{4d-s}$, while if the ℓ^d-torsion points of J are not defined over $\mathbb{F}_{p^k}$ then $[\ell^d m]P = O$ with probability at most ρ/ℓ. If we perform the test enough times, we can determine which probability distribution we are observing and thus conclude, with a high degree of certainty, whether the ℓ^d-torsion points are defined over $\mathbb{F}_{p^k}$.

This method is very effective in practice, and can be implemented for large k: while computing the group structure of $J(\mathbb{F}_{p^k})$ for large k may be infeasible, it is much easier to compute *points* on $J(\mathbb{F}_{p^k})$ and to do arithmetic on those points. We now give a formal description of the algorithm and determine its probability of success.

Algorithm 4.3. The following algorithm takes as input the Jacobian J of a genus 2 curve defined over a finite field $\mathbb{F}_q$, a prime power ℓ^d with $\gcd(\ell, q) = 1$, and a real number $\epsilon \in (0, 1)$. If $J[\ell^d] \subset J(\mathbb{F}_q)$, then the algorithm outputs `true` with probability at least $1 - \epsilon$. If $J[\ell^d] \not\subset J(\mathbb{F}_q)$, then the algorithm outputs `false` with probability at least $1 - \epsilon$.

(1) Compute $\#J(\mathbb{F}_q) = \ell^s m$, where $\ell \nmid m$. If $s < 4d$ output `false`.
(2) Set $\rho \leftarrow \ell^{4d-s}$, $N \leftarrow \lceil \frac{\sqrt{-2\log\epsilon}}{\rho}(\frac{2\ell}{\ell-1})\rceil$, $B \leftarrow \rho N \left(\frac{\ell+1}{2\ell}\right)$.
(3) Repeat N times:
    (a) Choose a random point $P_i \in J(\mathbb{F}_q)$.
    (b) Compute $Q_i \leftarrow [\ell^d m]P_i$
(4) If at least B of the Q_i are the identity element O of J, output `true`; otherwise output `false`.

Proof. As observed above, if $J[\ell^d] \subset J(\mathbb{F}_q)$, then $Q_i = O$ with probability ρ, while if $J[\ell^d] \not\subset J(\mathbb{F}_q)$, then $Q_i = O$ with probability at most ρ/ℓ. Thus all we have to do is compute enough Q_i to distinguish the two probability distributions. To figure out how many "enough" is, we use the Chernoff bound [23, Ch. 8, Prop. 5.3]. The version of the bound we use is as follows: If N weighted coins are flipped and μ is the expected number of heads, then for any $\delta \in (0, 1]$ we have

$$\Pr[\#\text{heads} < (1-\delta)\mu] < e^{-\mu^2\delta^2/2}$$
$$\Pr[\#\text{heads} > (1+\delta)\mu] < e^{-\mu^2\delta^2/2}.$$

In our case we are given two different probability distributions for the coin flip and wish to tell them apart. If the ℓ^d-torsion points of J are defined over $\mathbb{F}_q$, then the probability that $Q_i = O$ is $\rho = \ell^{4d}/\ell^s$. Thus the expected number of Q_i equal to O is $\mu_1 = \rho N$. If the ℓ^d-torsion points are not defined over $\mathbb{F}_q$, then the expected number of Q_i equal to O is at most $\mu_2 = \rho N/\ell$. Thus if we set $B = \rho N(\frac{\ell+1}{2\ell})$ to be the midpoint of $[\mu_2, \mu_1]$, we will deduce that $J[\ell^d] \subset J(\mathbb{F}_q)$ if the number of Q_i equal to O is at least B, and $J[\ell^d] \not\subset J(\mathbb{F}_q)$ otherwise.

We thus wish to find an N such that this deduction is correct with probability at least $1 - \epsilon$, i.e. an N such that

$$\Pr[\#\{Q_i : Q_i = O\} < B] < \epsilon \quad \text{if } J[\ell^d] \subset J(\mathbb{F}_q),$$
$$\Pr[\#\{Q_i : Q_i = O\} > B] < \epsilon \quad \text{if } J[\ell^d] \not\subset J(\mathbb{F}_q).$$

Substituting our choice of B into the Chernoff bound (4.3) gives

$$\Pr[\#\{Q_i : Q_i = O\} < B] < e^{-2\mu_1^2(\frac{\ell-1}{4\ell})^2} \quad \text{if } J[\ell^d] \subset J(\mathbb{F}_q),$$
$$\Pr[\#\{Q_i : Q_i = O\} > B] < e^{-2\mu_2^2(\frac{\ell-1}{4})^2} \quad \text{if } J[\ell^d] \not\subset J(\mathbb{F}_q).$$

From these equations, we see that we wish to have $2\mu_1^2(\frac{\ell-1}{4\ell})^2 > -\log\epsilon$ and $2\mu_2^2(\frac{\ell-1}{4})^2 > -\log\epsilon$. The two left sides are equal since $\mu_2 = \mu_1/\ell$. We thus substitute $\mu_1 = \rho N$ into the relation $2\mu_1^2(\frac{\ell-1}{4\ell})^2 > -\log\epsilon$, and find that

$$N > \frac{\sqrt{-2\log\epsilon}}{\rho} \cdot \frac{2\ell}{\ell-1}.$$

Thus this value of N suffices to give the desired success probabilities. □

Remark 4.4. If $s = 4d$, then the algorithm can be simplified considerably. In this case, if $J[\ell^d] \subset J(\mathbb{F}_q)$ then the ℓ-primary part H of $J(\mathbb{F}_q)$ is isomorphic to $(\mathbb{Z}/\ell^d\mathbb{Z})^4$, and if not then it contains a point of order greater than ℓ^d. Thus if $J[\ell^d] \subset J(F)$ then Q_i will always be the identity, and the

algorithm will always return `true`. On the other hand, if $J[\ell^d] \not\subset J(\mathbb{F}_q)$, we may abort the algorithm and return `false` as soon as we find a point $Q_i \neq O$, for in this case we have found a point in H of too large order, and thus the ℓ^d-torsion points are not defined over $\mathbb{F}_q$. If $J[\ell^d] \not\subset J(\mathbb{F}_q)$, then the probability that a random point in H has order $\leq \ell^d$ is at most $1/\ell$, so we must conduct at least $N = \lceil \frac{-\log \epsilon}{\log \ell} \rceil$ trials to ensure a success probability of at least $1 - \epsilon$. Thus in this case the method may require many fewer trials.

Remark 4.5. Note that while $\#J(\mathbb{F}_q)$ may be very large, in our application where J is defined over a small prime field it is easy to compute $\#J(\mathbb{F}_q)$ from the zeta function of the curve of which J is the Jacobian. Furthermore, while it is probably impossible to factor $\#J(\mathbb{F}_q)$ completely in a reasonable amount of time, it is easy to determine the highest power of ℓ that divides $\#J(\mathbb{F}_q)$.

Proposition 4.6. *Let J be the Jacobian of a genus 2 curve over $\mathbb{F}_p$. Assume that the zeta function of $J/\mathbb{F}_p$ is known, so that the cost to compute $\#J(\mathbb{F}_{p^k}) = \ell^s m$ is negligible. Then the expected number of operations in $\mathbb{F}_p$ necessary to execute Algorithm 4.3 on $J/\mathbb{F}_{p^k}$ (ignoring $\log\log p$ factors) is*

$$O(k^2 \log k (\log^2 p) \ell^{s-4d} (-\log \epsilon)^{1/2})$$

Proof. We must compare the cost of the two actions of Step 3, repeated N times. Choosing a random point on $J(\mathbb{F}_q)$ is equivalent to computing a constant number of square roots in $\mathbb{F}_q$, and taking a square root requires $O(\log q)$ field operations in $\mathbb{F}_q$ (see [13, Algorithm 14.15 and Corollary 14.16]). The order of $J(\mathbb{F}_q)$ is roughly q^2, so multiplying a point on $J(\mathbb{F}_q)$ by an integer using a binary expansion takes $O(\log q)$ point additions on $J(\mathbb{F}_q)$. Each point addition takes a constant number of field operations in $\mathbb{F}_q$, so we see that the time of each trial is $O(\log q) = O(k \log p)$. If fast multiplication techniques are used, then the number of field operations in $\mathbb{F}_p$ needed to perform one field operation in $\mathbb{F}_q$ is $O(\log q \log\log q) = O(k \log k \log p)$ (ignoring $\log\log p$ factors), so each trial takes $O(k^2 \log k \log^2 p)$ field operations in $\mathbb{F}_p$. The number of trials is $O(\ell^{s-4d}\sqrt{-\log \epsilon})$, which gives a total of $O(k^2 \log k (\log^2 p) \ell^{s-4d} (-\log \epsilon)^{1/2})$ field operations in $\mathbb{F}_p$. □

5. Computing the action of Frobenius

As in the previous section, we consider a genus 2 curve C over $\mathbb{F}_p$ with Jacobian J, and assume that the endomorphism ring of J is an order in the ring of integers $\mathcal{O}_K$ of a primitive quartic CM field K. We let π represent

the Frobenius endomorphism, and we look at elements $\alpha \in \mathcal{O}_K$ such that $\ell^d\alpha \in \mathbb{Z}[\pi]$ for some prime power ℓ^d. We wish to devise a test that, given such an α, determines whether α is an endomorphism of J.

Since π satisfies a quartic polynomial with integer coefficients, we can write α as

$$\alpha = \frac{a_0 + a_1\pi + a_2\pi^2 + a_3\pi^3}{\ell^d} \tag{5}$$

for some integers a_0, a_1, a_2, a_3. Expressing α in this form is useful because of the following fact proved by Eisenträger and Lauter [11, Corollary 9]: α is an endomorphism if and only if $T = a_0 + a_1\pi + a_2\pi^2 + a_3\pi^3$ acts as zero on the ℓ^d-torsion. Thus we need a method for determining whether T acts as zero on the ℓ^d-torsion. Since T is a linear operator, it suffices to check whether $T(Q_i)$ is zero for each Q_i in some set whose points span the full ℓ^d-torsion. Below we describe three different ways to compute such a spanning set.

5.1. *The brute force method*

The most straightforward way to compute a spanning set for the ℓ^d-torsion is to use group structure algorithms to compute a basis of $J[\ell^d]$. This method was used in [11] to compute the class polynomials in one example. The methods of Section 4 determine a k for which $J[\ell^d] \subset J(\mathbb{F}_{p^k})$. The computation of the group structure of $J(\mathbb{F}_{p^k})$ gives generators for the group; multiplying these generators by appropriate integers gives generators for the ℓ^d-torsion. It is then straightforward to compute the action of T on each generator g_i for $1 \le i \le 4$. If $T(g_i) = O$ for all i, then α is an endomorphism; otherwise α is not an endomorphism.

This method of computing a spanning set has the same drawback as the brute-force method of computing fields of definition: since the best algorithm for computing group structure runs in time exponential in $k \log p$, the method becomes prohibitively slow as k increases. Thus the method is only effective when ℓ^d is very small.

5.2. *A probabilistic method*

The method of Section 5.1 for computing generators of $J[\ell^d]$ becomes prohibitively slow as the field of definition of the ℓ^d-torsion points becomes large. However, we can get around this obstacle by randomly choosing many points Q_i of exact order ℓ^d, so that it is highly probable that the set $\{Q_i\}$ spans $J[\ell^d]$.

Recall that we wish to test whether the operator $T = a_0 + a_1\pi + a_2\pi^2 + a_3\pi^3$ acts as zero on the ℓ^d-torsion. To perform the test, we determine the field $\mathbb{F}_{p^k}$ over which we expect the ℓ^d-torsion to be defined. (See Section 4.) We pick a random point $P \in J(\mathbb{F}_{p^k})$ and multiply P by an appropriate integer to get a point Q whose order is a power of ℓ. If Q has order ℓ^d, we act on Q by the operator T and test whether we get the identity of J; otherwise we try again with a new P. (See Section 5.3 for another method of randomly choosing ℓ^d-torsion points.) We repeat the test until it is overwhelmingly likely that the points Q span the ℓ^d-torsion. If the set of Q spans the ℓ^d-torsion, then α is an endomorphism if and only if T acts as zero on all the Q.

Algorithm 5.1. The following algorithm takes as input the Jacobian J of a genus 2 curve over $\mathbb{F}_q$ with CM by K, a prime power ℓ^d with $\gcd(\ell, q) = 1$, the element $\pi \in \mathcal{O}_K$ corresponding to the Frobenius endomorphism of J, an element $\alpha \in \mathcal{O}_K$ such that $\ell^d\alpha \in \mathbb{Z}[\pi]$, and a real number $\epsilon > 0$. The algorithm outputs `true` or `false`.

Suppose $J[\ell^d] \subset J(\mathbb{F}_q)$. If α is an endomorphism of J, then the algorithm outputs `true`. If α is not an endomorphism of J, then the algorithm outputs `false` with probability at least $1 - \epsilon$.

(1) Compute a_0, a_1, a_2, a_3 such that α satisfies equation (5).
(2) Set N to be

$$N \leftarrow \begin{cases} \lceil \frac{1}{d - \log_\ell 2} \left(-\log_\ell \epsilon + 3d\right) \rceil & \text{if } \ell^d > 2 \\ \max\{\lceil -2\log_2 \epsilon \rceil + 6, 16\} & \text{if } \ell^d = 2. \end{cases}$$

(3) Compute $\#J(\mathbb{F}_q) = \ell^s m$, where $\ell \nmid m$.
(4) Set $i \leftarrow 1$.
(5) Choose a random point $P_i \in J(\mathbb{F}_q)$. Set $Q_i \leftarrow [m]P_i$. Repeat until $[\ell^d]Q_i = O$ and $[\ell^{d-1}]Q_i \neq O$.
(6) Compute

$$[a_0]Q_i + [a_1]\,\mathrm{Frob}_p(Q_i) + [a_2]\,\mathrm{Frob}_{p^2}(Q_i) + [a_3]\,\mathrm{Frob}_{p^3}(Q_i) \qquad (6)$$

in $J(\mathbb{F}_q)$. If the result is nonzero output `false`.
(7) If $i < N$, set $i \leftarrow i + 1$ and go to Step 5.
(8) Output `true`.

Proof. By [11, Corollary 9], α is an endomorphism of J if and only if the expression (6) is O for all ℓ^d-torsion points Q. Furthermore, it suffices to check the the expression only on a basis of the ℓ^d-torsion. Step 5 repeats

until we find a point Q_i of exact order ℓ^d; the assumption $J[\ell^d] \subset J(\mathbb{F}_q)$ guarantees that we can find such a point. The algorithm computes a total of N such points Q_i. Thus if the set of Q_i span $J[\ell^d]$, then the algorithm will output `true` or `false` correctly, according to whether $\alpha \in \mathrm{End}(J)$. We must therefore compute a lower bound for the probability that the set of Q_i computed span $J[\ell^d]$.

To compute this bound, we will compute an upper bound for the probability that N points of exact order ℓ^d do not span $J[\ell^d]$. We will make repeated use of the following inequality, which can be proved easily with simple algebra: if ℓ, d, n, and m are positive integers with $\ell > 1$ and $n > m$, then

$$\frac{\ell^{md} - \ell^{m(d-1)}}{\ell^{nd} - \ell^{n(d-1)}} < \frac{1}{\ell^{(n-m)d}}.$$

Next we observe that in any group of the form $(\mathbb{Z}/\ell^d\mathbb{Z})^r$, there are $\ell^{rd} - \ell^{r(d-1)}$ elements of exact order ℓ^d. The probability that a set of N elements does not span a 4-dimensional space is the sum of the probabilities that all the elements span a j-dimensional subspace, for $j = 1, 2, 3$. We consider each case:

- $j = 1$: All of the Q_i are in the space spanned by Q_1, and Q_1 can be any element. The probability of this happening is

$$\left(\frac{\ell^d - \ell^{d-1}}{\ell^{4d} - \ell^{4(d-1)}}\right)^{N-1} < \left(\frac{1}{\ell^{3d}}\right)^{N-1}.$$

- $j = 2$: Q_1 can be any element, one of the Q_i must be independent of Q_1, and the remaining $N - 2$ elements must be in the same 2-dimensional subspace. There are $N - 1$ ways to choose the second element, so the total probability is

$$(N-1)\left(1 - \frac{\ell^d - \ell^{d-1}}{\ell^{4d} - \ell^{4(d-1)}}\right)\left(\frac{\ell^{2d} - \ell^{2(d-1)}}{\ell^{4d} - \ell^{4(d-1)}}\right)^{N-2} < N\left(\frac{1}{\ell^{2d}}\right)^{N-2}.$$

- $j = 3$: Q_1 can be any element, and there must be two more linearly independent elements; there are $\binom{N-1}{2}$ ways of choosing these elements. The remaining $N - 3$ elements must all be in the same 3-dimensional subspace, so the total probability is

$$\frac{(N-1)(N-2)}{2}\left(1 - \frac{\ell^d - \ell^{d-1}}{\ell^{4d} - \ell^{4(d-1)}}\right)\left(1 - \frac{\ell^{2d} - \ell^{2(d-1)}}{\ell^{4d} - \ell^{4(d-1)}}\right)\left(\frac{\ell^{3d} - \ell^{3(d-1)}}{\ell^{4d} - \ell^{4(d-1)}}\right)^{N-3} < \frac{N^2}{2}\left(\frac{1}{\ell^d}\right)^{N-3}.$$

Summing these three cases, we see that the total probability that the Q_i do not span $J[\ell^d]$ is bounded above by

$$N^2 \left(\frac{1}{\ell^d}\right)^{N-3}.$$

Since $2^N \geq N^2$ for $N \geq 4$, we have

$$N^2 \left(\frac{1}{\ell^d}\right)^{N-3} \leq \ell^{-dN+3d+N\log_\ell 2}.$$

(Note that $N \geq 4$ must always hold if we want to have a spanning set of $J[\ell]$.) Setting this last expression less than ϵ and taking logs, we find

$$N \geq \frac{1}{d - \log_\ell 2}\left(-\log_\ell \epsilon + 3d\right). \tag{7}$$

Thus if the number of trials N is greater than or equal to the right hand side of (7), then the probability of success is at least $1 - \epsilon$.

The right hand side of expression (7) is undefined if $\ell = 2$, $d = 1$, so we must make a different estimate. Since $2^{N/2} \geq N^2$ for $N \geq 16$, the estimate (5.2) bounds the probability of Q_i not spanning $J[\ell^d]$ by

$$\frac{N^2}{2^{N-3}} \leq \frac{1}{2^{N/2-3}}.$$

Setting the right hand side less than ϵ and taking logs gives

$$N \geq -2\log_2 \epsilon + 6. \tag{8}$$

Thus if the number of trials N is greater than or equal to the maximum of 16 and the right hand side of (8), then the probability of success is at least $1 - \epsilon$. □

Corollary 5.2. *Let J, ℓ^d, α, and ϵ be as in Algorithm 5.1. Suppose $\pi \in \mathcal{O}_K$ is such that π^σ corresponds to the Frobenius endomorphism of J for some $\sigma \in \mathrm{Aut}(K/\mathbb{Q})$. Suppose $J[\ell^d] \subset J(\mathbb{F}_q)$, and suppose Algorithm 5.1 is run with inputs J, $\mathbb{F}_q$, π, α, ϵ. If α^σ is an endomorphism of J, then the algorithm outputs* `true`*. If α^σ is not an endomorphism of J, then the algorithm outputs* `false` *with probability at least $1 - \epsilon$.*

Proof. If we write α in the form (5), then we have

$$\alpha^\sigma = \frac{a_0 + a_1\pi^\sigma + a_2(\pi^\sigma)^2 + a_3(\pi^\sigma)^3}{\ell^d}.$$

Step 6 of the algorithm determines whether the numerator of this expression acts as zero on ℓ^d-torsion points. By [11, Corollary 9], this action is

identically zero if and only if α^σ is an endomorphism of J. The statement now follows from the correctness of Algorithm 5.1. □

Remark 5.3. Since Q_i is an ℓ^d-torsion point in Step 6, we may speed up the computation of the expression (6) by replacing each a_j with a small representative of a_j modulo ℓ^d. We may also rewrite the expression (6) as

$$[a_0]Q_i + \mathrm{Frob}_p([a_1]Q_i + \mathrm{Frob}_p([a_2]Q_i + \mathrm{Frob}_p([a_3]Q_i)))$$

to reduce the number of Frob_p operations from 6 to 3.

Remark 5.4. Algorithm 5.1 assumes that the ℓ^d-torsion points of J are defined over $\mathbb{F}_q$, so with enough trials we are almost certain to get a spanning set of points Q_i. However, if the ℓ^d-torsion points are not defined over $\mathbb{F}_q$, then the points Q_i will span a proper subspace of $J[\ell^d]$. If α is an endomorphism then T will act as zero on all of the Q_i and Algorithm 5.1 will output `true`. However, if α is not an endomorphism then T may still act as zero on all of the Q_i (in which case it must have nonzero action on the ℓ^d-torsion points that are not defined over $\mathbb{F}_q$), and the algorithm will incorrectly output `true`. Thus to test whether α is an endomorphism, we must combine Algorithm 5.1 with a method of checking the field of definition of the ℓ^d-torsion points, via the probabilistic method of Algorithm 4.3 or one of the other methods.

Proposition 5.5. *Let J be the Jacobian of a genus 2 curve over $\mathbb{F}_p$. Assume that the zeta function of $J/\mathbb{F}_p$ is known, so that the cost to compute $\#J(\mathbb{F}_{p^k}) = \ell^s m$ is negligible. Then the expected number of operations in $\mathbb{F}_p$ necessary to execute Algorithm 5.1 on $J/\mathbb{F}_{p^k}$ (ignoring $\log\log p$ factors) is*

$$O(k^2 \log k (\log^2 p)\ell^{s-4d}(-\log \epsilon))$$

Proof. Let $q = p^k$. In the proof of Proposition 4.6, we computed that the cost of computing a random point on $J(\mathbb{F}_q)$ is $O(\log q)$ operations in $\mathbb{F}_q$, and the cost of a point multiplication on $J(\mathbb{F}_q)$ is $O(\log q)$ operations in $\mathbb{F}_q$. The chance that a random point in the ℓ-primary part of $J(\mathbb{F}_q)$ has exact order ℓ is $\frac{\ell^{4d}-\ell^{4d-4}}{\ell^s}$, so the expected number of random points necessary to find one point of exact order ℓ^d is $O(\ell^{s-4d})$. The cost of computing the Frobenius action is proportional to the cost of raising an element of $\mathbb{F}_q$ to the pth power, which is $O(\log p)$ $\mathbb{F}_q$-operations.

We conclude that the expected cost of a single trial with a random point is

$$O(\log q + \log q + \log p)\ell^{s-4d} M(q)$$

operations in $\mathbb{F}_p$, where $M(q)$ is the number of field operations in $\mathbb{F}_p$ needed to perform one field operation in $\mathbb{F}_q$. If fast multiplication techniques are used, then $M(q) = O(\log q \log\log q) = O(k \log k \log p)$ (ignoring $\log\log p$ factors), so each trial takes $O(k^2 \log k (\log^2 p)\ell^{s-4d})$ field operations in $\mathbb{F}_p$. The number of points of exact order ℓ^d computed is $O(-\log \epsilon)$. Putting this all together gives a total of $O(k^2 \log k (\log^2 p)\ell^{s-4d}(-\log \epsilon))$ field operations in $\mathbb{F}_p$. □

5.3. *The Couveignes method*

Recall that to test whether an element $\alpha \in \mathcal{O}_K$ of the form (5) is an endomorphism of J, we determine whether the operator $T = a_0 + a_1\pi + a_2\pi^2 + a_3\pi^3$ acts as zero on all elements of a set $\{Q_i\}$ that spans $J[\ell^d]$. Algorithm 5.1 computes the spanning set by choosing random points P_i in $J(\mathbb{F}_{p^k})$, multiplying by an appropriate m to get points Q_i in the ℓ-primary part of $J(\mathbb{F}_{p^k})$ (denoted $J(\mathbb{F}_{p^k})_\ell$), and keeping only those Q_i whose order is exactly ℓ^d. If $J(\mathbb{F}_{p^k})_\ell$ is much larger than $J[\ell]$, the orders of most of the Q_i will be too large, and it will take many trials to find the required number of points of order exactly ℓ^d. To reduce the number of trials required, we would like to find a function from $J(\mathbb{F}_{p^k})_\ell$ to $J[\ell^d]$ that sends most of the Q_i to points of exact order ℓ^d.

One way to compute such a function is as follows: compute the order ℓ^{t_i} of each Q_i; if $t_i \geq d$ send $Q_i \mapsto [\ell^{t-d}]Q_i$, otherwise send $Q_i \mapsto O$. In most cases the image has order ℓ^d. However, since the multiplier ℓ^{t_i-d} will be different for each Q_i, this function does not define a group homomorphism, and thus the image of a set of points uniformly distributed in $J(\mathbb{F}_{p^k})_\ell$ will not be uniformly distributed in $J[\ell^d]$.

Couveignes [8] has described a map that has the properties we want and is a group homomorphism. The idea is the following: if $\pi^k - 1 \in \ell^d \operatorname{End}(J)$, then there is an endomorphism ϕ such that $\ell^d\phi = \pi^k - 1$. Since $\pi^k - 1$ acts as zero on $J(\mathbb{F}_{p^k})$, the image of ϕ on $J(\mathbb{F}_{p^k})$ must consist of ℓ^d-torsion points. Furthermore, the kernel of ϕ contains $\ell^d J(\mathbb{F}_{p^k})$, since $\phi(\ell^d P) = (\pi^k - 1)(P) = 0$ if P is defined over $\mathbb{F}_{p^k}$. Thus we have a map

$$\phi : J(\mathbb{F}_{p^k})/\ell^d J(\mathbb{F}_{p^k}) \to J[\ell^d].$$

Couveignes then uses the non-degeneracy of the Frey-Rück pairing (see [24]) to show that ϕ is a bijection. Thus for any Q_i not in $\ell J(\mathbb{F}_{p^k})$, $\phi(Q_i)$ has order exactly ℓ^d. Since ϕ is a surjective group homomorphism, the image of a set of points uniformly distributed in $J(\mathbb{F}_{p^k})$ will be uniformly distributed

in $J[\ell^d]$. The chance that $Q_i \in \ell J(\mathbb{F}_{p^k})$ is $1/\ell^4$, so applying ϕ to the Q_i will very quickly give a spanning set of $J[\ell^d]$.

However, there is one important caveat: we may not be able to compute ϕ. The only endomorphisms we can compute are those involving the action of Frobenius and scalar multiplication; namely, endomorphisms in $\mathbb{Z}[\pi]$. Thus we need to take k to be the smallest integer such that $\pi^k - 1 \in \ell^d\mathbb{Z}[\pi]$. We can then use the characteristic polynomial of Frobenius to write $\phi = \frac{\pi^k-1}{\ell^d} = M(\pi)$, where M is a polynomial of degree 3. Furthermore, since we are applying ϕ only to points $Q_i \in J(\mathbb{F}_{p^k})_\ell$, we may reduce the coefficients of M modulo ℓ^s and get the same action on the Q_i.

We have implemented the map ϕ in Magma and tested it on the examples that appear in Section 9. In our examples, the smallest k for which $\pi^k - 1 \in \ell^d\mathbb{Z}[\pi]$ is usually equal to ℓk_0, where k_0 is the integer output by Algorithm 4.1. We found that the cost of choosing random points over a field of degree ℓ times as large far outweighs the benefit of having to reject fewer of the points Q_i, so this technique does not help to speed up Algorithm 5.1.

6. Bounding the field of definition of the ℓ^d-torsion points

The running times of Algorithms 4.3 and 5.1 depend primarily on the size of the field $\mathbb{F}_{p^k}$ over which the ℓ^d-torsion points of J are defined. In this section, we bound the size of k in terms of ℓ^d and p. We also show that to determine the field of definition of the ℓ^d-torsion points of J for $d > 1$, it suffices to determine the field of definition of the ℓ-torsion points of J. This result allows us to work over much smaller fields in Algorithm 4.3, thus saving us a great deal of computation.

By Lemma 3.3, the prime powers ℓ^d input to Algorithms 4.3 and 5.1 divide the index $\mathbb{Z}[\pi, \overline{\pi}]$. Thus a bound on this index gives a bound on the ℓ^d that appear.

Proposition 6.1. *Let K be a primitive quartic CM field with discriminant $\Delta = \Delta(\mathcal{O}_K)$. Suppose $\pi \in \mathcal{O}_K$ corresponds to the Frobenius endomorphism of the Jacobian of a genus 2 curve defined over $\mathbb{F}_p$. Then*

$$[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]] \leq \frac{16p^2}{\sqrt{\Delta}}.$$

Proof. We showed in the proof of Corollary 3.6 that $[\mathbb{Z}[\pi, \overline{\pi}] : \mathbb{Z}[\pi]] = p$. Combining this result with the formula

$$[\mathcal{O}_K : \mathbb{Z}[\pi]] = [\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]][\mathbb{Z}[\pi, \overline{\pi}] : \mathbb{Z}[\pi]],$$

we see that it suffices to show that $[\mathcal{O}_K : \mathbb{Z}[\pi]] \leq 16p^3/\sqrt{\Delta}$. (Note that $\Delta > 0$ by [18, Proposition 9.4].) Next, recall that

$$[\mathcal{O}_K : \mathbb{Z}[\pi]] = \sqrt{\frac{\Delta(\mathbb{Z}[\pi])}{\Delta(\mathcal{O}_K)}}.$$

It thus suffices to show that $\sqrt{\Delta(\mathbb{Z}[\pi])} \leq 16p^3$. By definition,

$$\sqrt{\Delta(\mathbb{Z}[\pi])} = \prod_{i<j} |\alpha_i - \alpha_j|, \tag{9}$$

where α_i are the possible embeddings of π into $\mathbb{C}$. Since π represents an action of Frobenius, all of the α_i lie on the circle $|z| = \sqrt{p}$. The product (9) takes its maximum value subject to this constraint when the α_i are equally spaced around the circle, which happens when the α_i are $\sqrt{p}$ times primitive eighth roots of unity. The maximum product is thus $p^3\sqrt{\Delta(\mathbb{Q}(\zeta_8))} = 16p^3$. □

Proposition 6.1 also follows directly from [20, Proposition 7.4], where it is proved in a different manner that $\sqrt{\Delta(\mathbb{Z}[\pi, \overline{\pi}])} \leq 16p^2$.

The next two propositions give tight bounds on the degree k of the extension field of $\mathbb{F}_p$ over which the ℓ^d-torsion points of J are defined. The first considers the case $d = 1$, and the second shows that as d increases, k grows by a factor of ℓ^{d-1}.

Proposition 6.2. *Let J be the Jacobian of a genus 2 curve over $\mathbb{F}_p$, and suppose that* $\mathrm{End}(J)$ *is isomorphic to the ring of integers $\mathcal{O}_K$ of the primitive quartic CM field K. Let $\ell \neq p$ be a prime number, and suppose $\mathbb{F}_{p^k}$ is the smallest field over which the points of $J[\ell]$ are defined. If ℓ is unramified in K, then k divides one of the following:*

- *$\ell - 1$, if ℓ splits completely in K;*
- *$\ell^2 - 1$, if ℓ splits into two or three prime ideals in K;*
- *$\ell^3 - \ell^2 + \ell - 1$, if ℓ is inert in K.*

If ℓ ramifies in K, then k divides one of the following:

- *$\ell^3 - \ell^2$, if there is a prime over ℓ of ramification degree 3, or if ℓ is totally ramified in K and $\ell \leq 3$;*
- *$\ell^2 - \ell$, in all other cases where ℓ factors into four prime ideals in K (counting multiplicities);*
- *$\ell^3 - \ell$, if ℓ factors into two or three prime ideals in K (counting multiplicities).*

Proof. Let $\pi \in \mathcal{O}_K$ correspond to the Frobenius endomorphism. By [11, Fact 10], the ℓ-torsion points of J are defined over $\mathbb{F}_{p^k}$ if and only if $\pi^k - 1 \in \ell\mathcal{O}_K$. We observe that by the Chinese Remainder Theorem, this condition is satisfied if and only if $\pi^k \equiv 1 \pmod{\mathfrak{p}_i^{e_i}}$ for all primes $\mathfrak{p}_i \mid \ell\mathcal{O}_K$, where e_i is the ramification degree of $\mathfrak{p}_i$. Next, we note that the condition $\ell \neq p$ implies that $\pi \notin \mathfrak{p}_i$ for all i. To see why this is true, suppose the contrary: $\pi \in \mathfrak{p}_i$. Since $\pi\overline{\pi} = p$, we have $p \in \mathfrak{p}_i$, contradicting the fact that $\mathfrak{p}_i$ is a prime over $\ell \neq p$.

From these observations we deduce that k is the least common multiple of the multiplicative orders of π mod each $\mathfrak{p}_i^{e_i}$, and thus k must divide the least common multiple of

$$\#(\mathcal{O}_K/\mathfrak{p}_i^{e_i}\mathcal{O}_K)^\times = \ell^{f_i(e_i-1)}(\ell^{f_i} - 1),$$

where f_i is the inertia degree of $\mathfrak{p}_i$. We now consider the various possibilities for the splitting of ℓ in $\mathcal{O}_K$.

First, suppose ℓ is unramified, so $e_i = 1$ for all i.

- If ℓ splits completely, then the inertia degrees of all the $\mathfrak{p}_i$ are 1, so $k \mid \ell - 1$.
- If ℓ splits into two or three ideals, then at least one $\mathfrak{p}_i$ has $f_i = 2$ and all have $f_i \leq 2$, so $k \mid \ell^2 - 1$.
- If ℓ is inert, then there is a single $\mathfrak{p}_i$ with $f_i = 4$, and k divides $\ell^4 - 1$. We will return to this case below to get a better bound.

Now suppose ℓ ramifies; there are five possibilities for the splitting of ℓ in $\mathcal{O}_K$.

- If $\ell\mathcal{O}_K = \mathfrak{p}^3\mathfrak{q}$, then $\mathfrak{p}$ and $\mathfrak{q}$ have inertia degree 1, so k divides $\ell^2(\ell-1)$.
- If $\ell\mathcal{O}_K = \mathfrak{p}^4$, then $\mathcal{O}_K/\mathfrak{p} \cong \mathbb{F}_\ell$, and thus we have $\pi^{\ell-1} = 1 + \tau$ for some $\tau \in \mathfrak{p}$. There are now two subcases:
  - If $\ell \geq 5$, then $(1+\tau)^\ell \in 1 + \mathfrak{p}^4$, so $\pi^{\ell(\ell-1)} \equiv 1 \pmod{\mathfrak{p}^4}$. Thus k divides $\ell(\ell-1)$.
  - If $\ell = 2$ or 3, then $(1+\tau)^\ell \equiv 1 + \tau^\ell \pmod{\mathfrak{p}^4}$, so we must raise the expression to the ℓth power again to get rid of the τ^ℓ term. Thus $\pi^{\ell^2(\ell-1)} \equiv 1 \pmod{\mathfrak{p}^4}$, and k divides $\ell^2(\ell-1)$.
- If $\ell\mathcal{O}_K = \mathfrak{p}^2\mathfrak{q}^2$ or $\mathfrak{p}^2\mathfrak{q}\mathfrak{r}$, then all of the primes in question have inertia degree 1, so k divides $\ell(\ell-1)$.
- If $\ell\mathcal{O}_K = \mathfrak{p}^2\mathfrak{q}$, then $\mathfrak{p}$ has inertia degree 1 and $\mathfrak{q}$ has inertia degree 2, so k divides $\mathrm{lcm}(\ell(\ell-1), \ell^2-1) = \ell(\ell^2-1)$.

- If $\ell\mathcal{O}_K = \mathfrak{p}^2$, then $\mathcal{O}_K/\mathfrak{p} \cong \mathbb{F}_{\ell^2}$, and thus we have $\pi^{\ell^2-1} = 1 + \tau$ for some $\tau \in \mathfrak{p}$. Then $(1+\tau)^\ell \in 1 + \mathfrak{p}^2$, so $\pi^{\ell(\ell^2-1)} \equiv 1 \pmod{\mathfrak{p}^2}$. Thus k divides $\ell(\ell^2 - 1)$.

Thus far we have used only the fact that π is an algebraic integer, and we have not used the property that it represents the action of Frobenius. To get a better bound in the case where ℓ is inert in K, we recall that since π is the Frobenius endomorphism, we have $\pi\overline{\pi} = p$, and $K = \mathbb{Q}(\pi)$. Since ℓ is inert, reduction modulo ℓ gives an injective group homomorphism

$$\phi\colon \operatorname{Aut}\left({}^{K}/_{\mathbb{Q}}\right) \to \operatorname{Aut}\left({}^{(\mathcal{O}_K/\ell\mathcal{O}_K)}/_{(\mathbb{Z}/\ell\mathbb{Z})}\right).$$

Furthermore, the target group is isomorphic to $\mathrm{Gal}(\mathbb{F}_{\ell^4}/\mathbb{F}_\ell)$. This group is cyclic of order 4 and is generated by the ℓth-power Frobenius automorphism. Since complex conjugation has order 2 in $\mathrm{Aut}(K/\mathbb{Q})$, its image under ϕ must be the map $\alpha \mapsto \alpha^{\ell^2}$. Thus $\overline{\pi} \equiv \pi^{\ell^2} \pmod{\ell}$, and $\pi^{\ell^2+1} \equiv p \pmod{\ell}$. Since p must reduce to an element of $\mathbb{F}_\ell^\times$, p has order dividing $\ell - 1$, so π must have order dividing $(\ell^2+1)(\ell-1)$. □

The following proposition shows that in the cases we need for our application, the field of definition of the ℓ^d-torsion points is determined completely by the field of definition of the ℓ-torsion points.

Proposition 6.3. *Let A be an ordinary abelian variety defined over a finite field F, and let ℓ be a prime number not equal to the characteristic of F. Let d be a positive integer, and let F' be the extension field of F of degree ℓ^{d-1}. If the ℓ-torsion points of A are defined over F, then the ℓ^d-torsion points of A are defined over F'. If* $\mathrm{End}(A)$ *is integrally closed, then the converse also holds.*

Proof. Let $R = \mathrm{End}(A)$, and let $\pi \in R$ be the Frobenius endomorphism of F. By [11, Fact 10], for any positive integers t and k, the ℓ^t-torsion points of A are defined over the degree-k extension of F if and only if $\frac{\pi^k - 1}{\ell^t} \in R$, i.e. $\pi^k \equiv 1 \pmod{\ell^t R}$. To prove the proposition, it suffices to show that

$$\pi \equiv 1 \pmod{\ell R} \Leftrightarrow \pi^{\ell^{d-1}} \equiv 1 \pmod{\ell^d R},$$

with ($\Leftarrow$) holding when R is integrally closed.

First suppose that $\pi^k \equiv 1 \pmod{\ell^t R}$, with $t \geq 1$. Then we can write $\pi^k = 1 + \ell^t y$ for some $y \in R$. Then

$$\pi^{k\ell} = 1 + \ell(\ell^t y) + \binom{\ell}{2}(\ell^t y)^2 + \cdots + (\ell^t y)^\ell,$$

so $\pi^{k\ell} \equiv 1 \pmod{\ell^{t+1}R}$. We conclude that if the points of $A[\ell^t]$ are defined over the degree-k extension of F, then the points of $A[\ell^{t+1}]$ are defined over the degree-$k\ell$ extension of F. Thus if $A[\ell] \subset A(F)$, then by induction $A[\ell^d] \subset A(F')$.

Now suppose that $\pi^{k\ell} \equiv 1 \pmod{\ell^t R}$, with $t \geq 2$. Since A is ordinary, R is an order in a number ring. Thus if R is integrally closed then it is a Dedekind domain, and we may write $\ell R = \prod \mathfrak{p}_i^{e_i}$ uniquely for prime ideals $\mathfrak{p}_i \subset R$. By the Chinese Remainder Theorem, $\pi^k \equiv 1 \pmod{\ell^t R}$ if and only if $\pi^k \equiv 1 \pmod{\mathfrak{p}_i^{e_i t}}$ for each i, so we may consider the problem locally at each $\mathfrak{p}_i$. Localizing and completing the ring R at the prime $\mathfrak{p}_i$ gives a complete local ring R_v with maximal ideal $\mathfrak{p}_i$ and valuation v satisfying $v(\ell) = e_i$.

By hypothesis, we may write $\pi^{k\ell} = 1 + y$ for some $y \in \mathfrak{p}_i^{e_i t}$. We can define the ℓth-root function on R_v to be

$$(1+y)^{1/\ell} = \exp\left(\frac{1}{\ell}\log(1+y)\right).$$

By [22, Proposition II.5.5], if $y \in \mathfrak{p}_i^{e_i t}$ then $\log(1+y) \in \mathfrak{p}_i^{e_i t}$. Since $v(\ell) = e_i$, we have $v(\frac{1}{\ell}\log(1+y)) \geq e_i(t-1)$, so by the same Proposition $(1+y)^{1/\ell}$ converges and is in $1 + \mathfrak{p}_i^{e_i(t-1)}$ whenever $(t-1)(\ell-1) > 1$. Thus if $(t-1)(\ell-1) > 1$ then $\pi^k \equiv 1 \pmod{\mathfrak{p}_i^{e_i(t-1)}}$. We conclude that if $t > 2$ or $\ell > 2$ and the points of $A[\ell^t]$ are defined over the degree-$k\ell$ extension of F, then the points of $A[\ell^{t-1}]$ are defined over the degree-k extension of F. If $A[\ell^d] \subset A(F')$, then by descending induction $A[\ell] \subset A(F)$ if ℓ is odd, and $A[4] \subset A(F_2)$ if $\ell = 2$, where F_2 is the quadratic extension of F.

It remains to show that if $A[4] \subset A(F_2)$, then $A[2] \subset A(F)$. This is equivalent to showing that if $\pi^2 - 1 \in 4R$ then $\pi - 1 \in 2R$. We prove the contrapositive: suppose $\pi - 1 \notin 2R$. Then there is some prime $\mathfrak{p}$ over 2 such that $v_\mathfrak{p}(\pi - 1) < v_\mathfrak{p}(2)$. Since $\pi + 1 = (\pi - 1) + 2$ and $v_\mathfrak{p}(\pi - 1) < v_\mathfrak{p}(2)$, we must also have $v_\mathfrak{p}(\pi + 1) < v_\mathfrak{p}(2)$. Multiplying the two expressions gives $v_\mathfrak{p}(\pi^2 - 1) < v_\mathfrak{p}(4)$, so $\pi^2 - 1$ cannot be contained in $4R$. We conclude that $\pi^2 - 1 \in 4R$ implies $\pi - 1 \in 2R$. □

Corollary 6.4. *Let J be the Jacobian of a genus 2 curve over $\mathbb{F}_p$, and suppose that* $\mathrm{End}(J)$ *is isomorphic to the ring of integers $\mathcal{O}_K$ of the primitive quartic CM field K. Let ℓ^d be a prime power with $\ell \neq p$, and suppose $\mathbb{F}_{p^k}$ is the smallest field over which the points of $J[\ell^d]$ are defined. Then $k < 3p^6$.*

Proof. By Proposition 6.2, the points of $J[\ell]$ are defined over a field F of degree less than ℓ^3 over $\mathbb{F}_p$. By Proposition 6.3, the points of $J[\ell^d]$ are

defined over a field L of degree ℓ^{d-1} over F. Since degrees of extensions multiply, we get

$$k = [L : \mathbb{F}_p] < \ell^{d+2} \leq \ell^{3d}.$$

By Proposition 6.1, $\ell^d \leq \frac{16}{\sqrt{\Delta}}p^2$, where Δ is the discriminant of the quartic CM field K. Lemma 6.5 below shows that any primitive quartic CM field has $\Delta \geq 125$, so $\ell^d \leq \frac{16}{\sqrt{125}}p^2$. Since $k < \ell^{3d}$, we conclude that $k < 3p^6$. □

Lemma 6.5. *Suppose K is a primitive quartic CM field. Then $\Delta(K) \geq 125$.*

Proof. Since $\Delta(\mathbb{Q}(\zeta_5)) = 125$, it suffices to show that no smaller discriminant can occur. The fact that $\Delta(K) > 0$ follows from [18, Proposition 9.4]. Now suppose $\Delta(K) < 125$. Since $\Delta(K_0)^2 \mid \Delta(K)$, we must have $K_0 = \mathbb{Q}(\sqrt{2})$ or $\mathbb{Q}(\sqrt{5})$, as these are the only two real quadratic fields with discriminant less than 12. Since $\mathbb{Q}(\sqrt{2})$ has class number 1, by [22, Proposition VI.6.9], $\mathbb{Q}(\sqrt{2})$ has no unramified quadratic extensions, so $\Delta(K)$ is strictly greater than $\Delta(K_0)^2$. Thus if $K_0 = \mathbb{Q}(\sqrt{2})$ then $\Delta(K) \geq 128$.

We deduce that $K_0 = \mathbb{Q}(\sqrt{5})$ and K must be of the form $\mathbb{Q}(i\sqrt{a + b\sqrt{5}})$, with a, b, and $a^2 - 5b^2$ positive integers. Since K is primitive, $a^2 - 5b^2$ is not a square in $\mathbb{Q}$ and its square-free part divides $\Delta(K)/\Delta(K_0)^2$. It thus suffices to show that the square-free part of $a^2 - 5b^2$ is at least 5; this follows from the fact that 2 and 3 are inert in $\mathbb{Q}(\sqrt{5})$, so there are no integer solutions to $a^2 - 5b^2 = 2$ or 3. □

7. Computing Igusa class polynomials

This section combines the results of all of the previous sections into a full-fledged probabilistic version of Eisenträger and Lauter's CRT algorithm to compute Igusa class polynomials for primitive quartic CM fields [11, Theorem 1].

Algorithm 7.1. The following algorithm takes as input a primitive quartic CM field K, three integers $\lambda_1, \lambda_2, \lambda_3$ which are multiples of the denominators of the three Igusa class polynomials, and a real number $\epsilon > 0$, and outputs three polynomials $H_1, H_2, H_3 \in \mathbb{Q}[x]$. With high probability, the polynomials $H_i(x)$ output by the algorithm are the Igusa class polynomials for K.

(1) (Initialization.)

(a) Let D be the degree of the Igusa class polynomials for K, computed via class number algorithms, e.g. [6, Algorithm 6.5.9].
(b) Compute an integral basis $\mathcal{B}$ for $\mathcal{O}_K$, using e.g. [6, Algorithm 6.1.8].
(c) Set $p \leftarrow 3$, $B \leftarrow 1$, $H_1, H_2, H_3 \leftarrow 0$, $F_1, F_2, F_3 \leftarrow 0$.

(2) Set $p \leftarrow \texttt{NextPrime}(p)$ until p splits completely in K and p splits into principal ideals in K^* (the reflex field of K).
(3) (Finding the curves.) Set $T_1, T_2, T_3 \leftarrow \{\}$. For each $(i_1, i_2, i_3) \in \mathbb{F}_p^3$, do the following:
   (a) Compute a curve $C/\mathbb{F}_p$ with Igusa invariants (i_1, i_2, i_3), using the algorithms of Mestre [21] and Cardona-Quer [10].
   (b) Run Algorithm 2.1 with inputs K, p, C.
      i. If the algorithm outputs `false`, go to the next triple (i_1, i_2, i_3).
      ii. If the algorithm outputs `true`, let π be one of the possible Frobenius elements it outputs.
   (c) For each prime ℓ dividing $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$, do the following:
      i. Run Algorithm 4.1 with inputs K, ℓ, π. Let the output be k.
      ii. Run Algorithm 4.3 with inputs $\mathrm{Jac}(C)$, $\mathbb{F}_{p^k}$, ℓ, and ϵ. If the output is `false`, go to the next triple (i_1, i_2, i_3).
      iii. If ℓ^2 divides $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$, then for each $\alpha \in \mathcal{B} \setminus \mathbb{Z}$ written in the form (2) with denominator n, do the following:
         A. Let d be the largest integer such that $\ell^d \mid n$. If $d = 0$, go to the next α.
         B. Set $k' \leftarrow k\ell^{d-1}$.
         C. Run Algorithm 5.1 with inputs $\mathrm{Jac}(C), \mathbb{F}_{p^{k'}}, \ell^d, \pi, \frac{n}{\ell^d}\alpha, \epsilon$.
         D. If Algorithm 5.1 outputs `false`, go to the next triple (i_1, i_2, i_3). Otherwise go to the next α.
   (d) Adjoin i_1, i_2, i_3 to the sets T_1, T_2, T_3, respectively (counting multiplicities).
(4) If the size of each set T_1, T_2, T_3 is not equal to D, go to Step 2.
(5) (Computing the Igusa class polynomials.) For $i \in \{1, 2, 3\}$, do the following:
   (a) Compute $F_{i,p}(x) = \lambda_i \prod_{j \in T_i}(x - j)$ in $\mathbb{F}_p[x]$.
   (b) Use the Chinese Remainder Theorem to compute $F_i'(x) \in \mathbb{Z}[x]$ such that $F_i'(x) \equiv F_i(x) \pmod{B}$, $F_i'(x) \equiv F_{i,p}(x) \pmod{p}$, and the coefficients of $F_i'(x)$ are in the interval $[-pB/2, pB/2]$.
   (c) If $F_i'(x) = F_i(x)$, output $H_i(x) = \lambda_i^{-1} F_i(x)$. If $H_i(x)$ has been output for all i, terminate the algorithm.

(d) Set $F_i(x) \leftarrow F_i'(x)$.

(6) Set $B \leftarrow pB$, and return to Step 2.

Proof. In view of [11, Theorem 1], it suffices to prove that Step 3c correctly determines the set of curves with $\mathrm{End}(\mathrm{Jac}(C)) = \mathcal{O}_K$. It follows from Section 3 that $\mathrm{End}(\mathrm{Jac}(C)) = \mathcal{O}_K$ if and only if each of the elements of the generating set listed in Proposition 3.8 is an endomorphism.

By Algorithm 2.1, the π computed in Step 3b is such that π^σ is the Frobenius element of $\mathrm{Jac}(C)$ for some $\sigma \in \mathrm{Aut}(K/\mathbb{Q})$. By Corollary 3.10, $\mathrm{End}(\mathrm{Jac}(C)) = \mathcal{O}_K$ if and only if β^σ is an endomorphism for each β in the generating set of Proposition 3.8. Since elements of $\mathrm{Aut}(K/\mathbb{Q})$ preserve $\mathcal{O}_K$ as a set, $[\mathcal{O}_K : \mathbb{Z}[\pi^\sigma, \overline{\pi}^\sigma]] = [\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$.

For each ℓ dividing $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$, Steps 3(c)i and 3(c)ii test probabilistically whether $\frac{(\pi^\sigma)^k - 1}{\ell}$ is an endomorphism for an appropriate k. By Corollary 3.5, for any such ℓ dividing $[\mathcal{O}_K : \mathbb{Z}[\overline{\pi}]]$ exactly, this suffices to determine whether $\frac{n}{\ell}\alpha^\sigma$ is an endomorphism for each $\alpha \in \mathcal{B} \setminus \mathbb{Z}$.

By Corollary 5.2, if ℓ^2 divides $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ then Step 3(c)iii tests probabilistically whether $\frac{n}{\ell^d}\alpha^\sigma$ is an endomorphism. The input uses the field $\mathbb{F}_{p^{k'}}$ because Proposition 6.3 implies that if the ℓ-torsion points are defined over $\mathbb{F}_{p^k}$, then the ℓ^d-torsion points are defined over $\mathbb{F}_{p^{k'}}$. □

Remark 7.2. Note that Step 5 differs from the corresponding step in [11, Theorem 1]. Our version of the algorithm minimizes the amount of computation by terminating the algorithm in Step 5c as soon as the polynomials agree modulo two consecutive primes. For each prime p_i in an increasing sequence of primes, we compute a polynomial $F_{i,p}(x)$ that is congruent to the Igusa class polynomial $H_i(x)$ modulo the prime p_i [11, Theorem 2]. We then use the Chinese Remainder Theorem and the collection of polynomials $\{F_{i,p}(x)\}$ to compute a polynomial $F_i(x)$ modulo $b_i = \prod_{j=1}^{i} p_j$. If $F_i(x) = F_{i+1}(x)$, then with high probability the coefficients of $H_i(x)$ are less than b_{i+1}, and thus $F_i(x)$ is equal to $H_i(x)$ itself. This conclusion is justified by the fact that if an integer n has the property that it is the same modulo b_i and modulo b_{i+1}, then $n = a_i + r_i b_i = a_{i+1} + r_{i+1} b_{i+1}$, with $a_i < b_i$ and $a_i = a_{i+1}$. It follows that p_{i+1} divides r_i. Since the probability of this happening for a random number r_i is $1/p_{i+1}$, the probability that all coefficients would simultaneously satisfy this congruence is $(1/p_{i+1})^{D+1}$, so most likely we have that actually $r_{i+1} = 0$ for each coefficient.

Remark 7.3. The λ_i input into the algorithm can be taken to be products of primes bounded in [17], raised to a power that will be made explicit in

forthcoming work. In practice, the power can be taken to be a small multiple of 6.

Since we check after every prime p_i whether the algorithm is finished, we do not need to know in advance the number of primes p_i that we will need to use. Thus the only bounds that need to be computed in advance are the bounds λ_i on the denominators of the coefficients of the Igusa class polynomials. In particular, we do not need to have a bound on either the numerators or the absolute values of the coefficients.

8. Implementation notes

Our most significant observation is that in practice, the running time of the probabilistic CRT algorithm is dominated by generating p^3 curves for each small p. Steps 3a and 3b of Algorithm 7.1 generate a list of curves C for which $\mathrm{End}(\mathrm{Jac}(C))$ is an order in $\mathcal{O}_K$. Algorithms 4.3 and 5.1 determine which endomorphism rings are equal to $\mathcal{O}_K$. Data comparing the relative speeds of these two parts of the algorithm appear in Section 9. This section describes a number of ways to speed up Algorithm 7.1, which are reflected in the running times that appear in Section 9.

(1) If p and k are large, then arithmetic on $J(\mathbb{F}_{p^k})$ is prohibitively slow, which slows down Algorithms 4.3 and 5.1. Since for various ℓ dividing the index $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$, the extension degrees k depend only on the prime p and the CM field K and not on the curve C, these extension degrees may be computed in advance (via Algorithm 4.1) before generating any curves. We set some bound N and tell the program that if the extension degree k for some ℓ is such that $p^k > N$, we should skip that p and go on to the next prime. For example, if $K = \mathbb{Q}(i\sqrt{13 + 2\sqrt{13}})$ and $p = 53$ (see Example 9.2), we have $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]] = 3^2 \cdot 43$, and the 43-torsion of a Jacobian J with $\mathrm{End}(J) = \mathcal{O}_K$ will be defined over $\mathbb{F}_{p^{924}}$, a field of over 5000 bits that is far too large for our current implementation to handle efficiently.

(2) In a similar vein, since the speed of Algorithms 4.3 and 5.1 is determined by the size of the fields $\mathbb{F}_{p^k}$, for optimum performance one should perform these calculations in order of increasing k, so that as the fields get larger there are fewer curves to check.

(3) Algorithms 4.3 and 5.1 take a single curve as input. In Algorithm 7.1 those algorithms are executed with the same field K and many different curves, so any parameter that only depends on the field K and the prime p can be precomputed and stored for repeated reference. For example,

the representation $\alpha = (a_0 + a_1\pi + a_2\pi^2 + a_3\pi^3)/n$ and the extension degrees k in Step 3(c)i can be computed only once. In addition, all of the curves that pass Step 3b have one of a small number of given zeta functions. Since $\#J(\mathbb{F}_{p^k})$ is determined by the zeta function, this number can also be computed in advance.

(4) If $\mathbb{F}_{p^k}$ is small enough, it may be faster to check fields of definition using the brute force method of Section 4.1, rather than Algorithm 4.3. If ℓ is small (as must be the case for k to be small), then we often find that $\#J(\mathbb{F}_{p^k}) = \ell^s m$ with $s \gg 4d$, and thus the number of random points needed in Algorithms 4.3 and 5.1 will be very large. While computing the group structure is an exponential-time computation, we find that if the group has size at most 2^{200}, MAGMA can compute the group structure fairly quickly.

(5) If Step 5c has already output $H_j(x)$ for some j, the roots of this polynomial mod p can be used as the possible values of i_j in Step 3. This will greatly speed up the calculation of the $F_{i,p}$ for the remaining primes: if one H_j has been output then only p^2D curves need to be computed (instead of p^3), and if two H_j have been output then only pD^2 curves need to be computed.

(6) In practice, for small primes p ($p < 800$ in our MAGMA implementation), computing $\#C(\mathbb{F}_p)$ (Step 5b of Algorithm 2.1) is more efficient than choosing a random point on $J(\mathbb{F}_p)$ and determining whether it is killed by one of the potential group orders (Step 5a of Algorithm 2.1), so these two steps should be switched for maximum speed. However, as p grows, the order of the steps as presented will be the fastest.

9. Examples

This section describes the performance of Algorithm 7.1 on three quartic CM fields: $\mathbb{Q}(i\sqrt{2+\sqrt{2}})$, $\mathbb{Q}(i\sqrt{13+2\sqrt{13}})$, and $\mathbb{Q}(i\sqrt{29+2\sqrt{29}})$. These fields are all Galois and have class number 1, so the density of primes with the desired splitting behavior is maximal. The Igusa polynomials are linear; they have integral coefficients for the first two fields, and have denominators dividing 5^{12} for the last. In all three examples, as p grows, the running time of the algorithm becomes dominated by the computation of p^3 curves for each p, whereas it was previously suspected that the the endomorphism ring computation would be the slow step in the CRT algorithm. A fast implementation in C to produce the curves from their Igusa invariants and to test the numbers of points would thus significantly improve the running time of the CRT algorithm.

Details of the algorithms' execution are given below. The algorithms were run on a 2.39 GHz AMD Opteron with 4 GB of RAM. The table headings have the following meaning:

- p: Size of prime field over which curves were generated.
- ℓ^d: Prime powers appearing in the denominators n of elements α input into Algorithms 4.3 and 5.1, when written in the form (2).
- k: Degrees of extension fields over which ℓ^d-torsion points are expected to be defined. These are listed in the same order as the corresponding ℓ^d.
- Curves: Time taken to generate p^3 curves and determine which have CM by K (cf. Algorithm 2.1).
- #Curves: Number of curves computed whose Jacobians have CM by K.
- 4.3 & 5.1: Time taken to run Algorithms 4.3 and 5.1 to find the single curve whose Jacobian has endomorphism ring equal to $\mathcal{O}_K$.

Example 9.1. We ran Algorithm 7.1 with $K = \mathbb{Q}(i\sqrt{2+\sqrt{2}})$ and $\lambda_1, \lambda_2, \lambda_3 = 1$. The results appear in Table 1. The last column of the table shows the intermediate polynomials $F_i(x)$ computed via the Chinese Remainder Theorem in Step 5b. The algorithm output the $F_i(x)$ listed for $p = 151$ as the Igusa class polynomials of K.

The total time of this run was 3162 seconds, or about 53 minutes. We observe that the polynomials F_2 and F_3 agree for $p = 103$ and $p = 113$. We deduce that these polynomials are the correct Igusa polynomials, and following note 5 of Section 8, we use their roots for the values of i_2 and i_3 for $p = 151$. Thus instead of computing $151^3 \approx 2^{22}$ curves, we need to compute only 151 curves, out of which we can easily choose the right one. As a result, the computation for $p = 151$ takes practically no time at all. The same phenomenon also appears for the last prime in Examples 9.2 and 3.

Example 9.2. We ran Algorithm 7.1 with $K = \mathbb{Q}(i\sqrt{13+2\sqrt{13}})$ and $\lambda_1, \lambda_2, \lambda_3 = 1$. The results appear in Table 2. The algorithm output the following Igusa class polynomials:

$$x - 1836660096, \quad x - 28343520 \quad x - 9762768.$$

The total time of this run was 6969 seconds, or about 116 minutes. In this example we skip some primes because Algorithms 4.3 and 5.1 would need to compute in fields which are too large to be practical. In particular,

Table 1. Results for Algorithm 7.1 run with $K = \mathbb{Q}(i\sqrt{2+\sqrt{2}})$ and $\lambda_1, \lambda_2, \lambda_3 = 1$.

p	ℓ^d	k	Curves	#Curves	4.3 & 5.1	$F_i(x)$
7	2,4	2,4	0.5 sec	7	0.3 sec	$x+2$ $x+5$ $x+6$ (mod 7)
17	4,8	2,4	4 sec	39	0.2 sec	$x-54$ $x+19$ $x-8$ (mod 119)
23	2,4,7	2,4,3	9 sec	49	2.3 sec	$x+1017$ $x+852$ $x+111$ (mod 2737)
71	2,4	2,4	255 sec	7	0.7 sec	$x-75619$ $x+28222$ $x-46418$ (mod 194327)
97	4,8	2,4	680 sec	39	0.3 sec	$x-8237353$ $x+9355918$ $x+9086951$ (mod 18849719)
103	2,4,17	2,4,16	829 sec	119	17.6 sec	$x+104860961$ $x-28343520$ $x-9762768$ (mod 1941521057)
113	7,8,32	6,4,16	1334 sec	1281	28.8 sec	$x-1836660096$ $x-28343520$ $x-9762768$ (mod 219391879441)
151	2,4,7,17	2,4,6,16	0.2 sec	1	–	$x-1836660096$ $x-28343520$ $x-9762768$ (mod 33128173795591)

for $p = 29, 53, 107, 139$, the algorithms would run over extension fields of degree $264, 924, 308, 162$, all of which have well over 1000 bits. Skipping these primes has no effect on the ultimate outcome of the algorithm.

Example 9.3. We ran Algorithm 7.1 with $K = \mathbb{Q}(i\sqrt{29+2\sqrt{29}})$ and $\lambda_1, \lambda_2, \lambda_3 = 5^{12}$. The results appear in Table 3. The algorithm output the following Igusa class polynomials:

$$x - \frac{2614061544410821165056}{5^{12}}, \quad x + \frac{586040972673024}{5^6}, \quad x + \frac{203047103102976}{5^6}.$$

The total time of this run was 56585 seconds, or about 15 hours, 43 minutes. In this example we again skip some primes because the fields input to Algorithms 4.3 and 5.1 would be too large. We also note that for $p = 7$, $\mathcal{O}_K = \mathbb{Z}[\pi, \overline{\pi}]$, so any curve over $\mathbb{F}_7$ that has a correct zeta function already has CM by all of $\mathcal{O}_K$, and we do not need to run Algorithms 4.3 and 5.1.

Table 2. Results for Algorithm 7.1 with $K = \mathbb{Q}(i\sqrt{13+2\sqrt{13}})$ and $\lambda_1, \lambda_2, \lambda_3 = 1$.

p	ℓ^d	k	Curves	#Curves	4.3 & 5.1
29	3,23	2,264	–	–	–
53	3,43	2,924	–	–	–
61	3	2	167 sec	9	0.2 sec
79	27	18	376 sec	81	8.1 sec
107	9,43	6,308	–	–	–
113	3,53	1,52	1118 sec	159	137.2 sec
131	9,53	6,52	1872 sec	477	127.4 sec
139	9,243	6,162	–	–	–
157	9,81	6,54	3147 sec	243	16.5 sec
191	3,4,8	2,2,4	0.2 sec	1	–

Table 3. Results for Algorithm 7.1 with $K = \mathbb{Q}(i\sqrt{29+2\sqrt{29}})$ and $\lambda_1, \lambda_2, \lambda_3 = 5^{12}$.

p	ℓ^d	k	Curves	#Curves	4.3 & 5.1
7	–	–	0.3 sec	1	–
23	13	84	9 sec	15	70.7 sec
53	7	6	105 sec	7	0.5 sec
59	4,5,8	2,12,4	164 sec	322	6.4 sec
83	3,5	4,24	431 sec	77	9.8 sec
103	67	1122	–	–	–
107	7,13	6,42	963 sec	105	69.3 sec
139	7,25	2,60	2189 sec	259	62.1 sec
181	9,27	6,18	84 min	161	3.6 sec
197	5,109	24,5940	–	–	–
199	25	60	106 min	37	1355.3 sec
223	4,8,23	2,4,22	174 min	1058	35.1 sec
227	109	1485	–	–	–
233	5,7,13	8,3,28	193 min	735	141.6 sec
239	7,109	6,297	–	–	–
257	3,7,13	4,6,84	286 min	1155	382.8 sec
277	5,7,23	24,6,22	0.3 sec	1	–

Remark 9.4. The data in Examples 9.1, 9.2, and 9.3 suggest that odd primes dividing the index $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ always split in $\mathcal{O}_{K_0}$, the ring of integers of K_0. In fact the factorization of the index $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ was given in [11, Proposition 5] for primitive quartic CM fields K when K_0 has class number 1. We write $\pi = c_1 + c_2\sqrt{d} + (c_3 + c_4\sqrt{d})\eta$, where the c_i are rational numbers with only powers of 2 in the denominators and $\eta = i\sqrt{a + b\sqrt{d}}$ with $a, b, d \in \mathbb{Z}$, $d > 0$ and square-free. Then the index is, up to powers of 2, the product of c_2 with $(c_3^2 - c_4^2 d)$, where c_2 is the index of $\mathbb{Z}[\pi + \overline{\pi}]$ in

$\mathcal{O}_{K_0}$ up to a power of 2. If a prime divides $(c_3^2 - c_4^2 d)$ exactly, i.e. the square of the prime does not divide it, then the prime splits in K_0. Thus primes different from 2 dividing the index $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ exactly either split in K_0 or divide the index $[\mathcal{O}_{K_0} : \mathbb{Z}[\pi + \overline{\pi}]$. So except possibly for primes dividing c_2, no odd primes dividing the index $[\mathcal{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ exactly are inert or totally ramified in K. If K is Galois, then this is enough to ensure that the extension degree k determined by Proposition 6.2 is at most ℓ^2. This agrees with the data in our examples, all of which considered Galois fields.

In practice, if a prime ℓ is inert or totally ramified in K, it would almost certainly be skipped anyway, since Proposition 6.2 shows that the ℓ-torsion may be defined over an extension field of degree $k \sim \ell^3$, which is too large to be practical (cf. Note 1 of Section 8). However the theoretical running times of Algorithms 4.3 and 5.1, given by Propositions 4.6 and 5.5 respectively, improve if inert or ramified primes ℓ are not considered. The slow step of both algorithms is computing a random point on $J(\mathbb{F}_{p^k})$, which takes roughly $O(k^2 \log k (\log p)^2)$ $\mathbb{F}_p$ operations. Since the bound on ℓ is p^2, if k is bounded by ℓ^2 instead of ℓ^3, this step would run in $O(p^8 \log^3 p)$ instead of $O(p^{12} \log^3 p)$ time.

References

1. A. Agashe, K. Lauter, R. Venkatesan, "Constructing elliptic curves with a known number of points over a prime field," In *High Primes and Misdemeanours: lectures in honour of the 60th birthday of Hugh Cowie Williams*, Fields Institute Communications Series **42**, 2004, 1–17.
2. A.O.L. Atkin, F. Morain, "Elliptic curves and primality proving," *Math. Comp.* **61** (1993), 29–68.
3. D.J.Bernstein, "Multidigit modular multiplication with the Explicit Chinese Remainder Theorem," Chapter 4, Ph.D. thesis, University of California at Berkeley, 1995.
4. R. Bröker, "A p-adic algorithm to compute the Hilbert class polynomial," Preprint, 2006. `http://www.math.ucalgary.ca/~reinier/pub/padicj.pdf`
5. J. Chao, O. Nakamura, K. Sobataka, S. Tsujii, "Construction of secure elliptic cryptosystems using CM tests and liftings," in *ASIACRYPT '98* Springer LNCS **1514**, Beijing, 1998, 95–109.
6. H. Cohen, *A course in computational algebraic number theory*, Springer GTM **138**, 1993.
7. H. Cohn, K. Lauter, "Generating genus 2 curves with complex multiplication," Microsoft Research Internal Technical Report, 2001.
8. J.-M. Couveignes, "Linearizing torsion classes in the Picard group of algebraic curves over finite fields," Preprint, 2006. `http://www.picard.ups-tlse.fr/~couveig/publi/jaco.pdf`.
9. J.-M. Couveignes, T. Henocq, "Action of modular correspondences around

CM-points," in *ANTS-V*, Springer LNCS **2369**, 2002, 234–243.
10. G. Cardona, J. Quer, "Field of moduli and field of definition for curves of genus 2," in *Computational aspects of algebraic curves*, *Lecture Notes Ser. Comput.* **13**, World Sci. Publ., Hackensack, NJ, 2005, 71–83.
11. K. Eisenträger, K. Lauter, "A CRT algorithm for constructing genus 2 curves over finite fields," to appear in *AGCT-10*, 2007, `http://arxiv.org/abs/math.NT/0405305`.
12. A. Enge, "The complexity of class polynomial computation via floating point approximations," Preprint, 2006. `http://fr.arxiv.org/abs/cs.CC/0601104`.
13. J. von zur Gathen, J. Gerhard, *Modern Computer Algebra*, 2nd ed., Cambridge University Press, Cambridge, 2003.
14. P. Gaudry, R. Harley, "Counting Points on Hyperelliptic Curves over Finite Fields," in *ANTS-IV*, Springer LNCS **1838**, 2000, 297–312.
15. P. Gaudry, T. Houtmann, D. Kohel, C. Ritzenthaler, A. Weng, "The 2-adic CM method for genus 2 curves with application to cryptography," in *ASIACRYPT '06*, Springer LNCS **4284**, 2006, 114–129.
16. P. Gaudry, É. Schost, "Construction of secure random curves of genus 2 over prime fields," in *EUROCRYPT '04*, Springer LNCS **3027**, 2004, 239–256.
17. E. Goren, K. Lauter, "Class invariants for quartic CM fields," *Annales de l'Institut Fourier*, **57** (2007), 457–480.
18. E. Howe, "Principally polarized ordinary abelian varieties over finite fields", *Trans. Amer. Math. Soc.* **347** (1995) 2361–2401.
19. A. K. Lenstra, H. W. Lenstra, L. Lovász, "Factoring polynomials with rational coefficients," *Math. Ann.* **261** (1982), 515–534.
20. H. W. Lenstra, Jr., J. Pila, C. Pomerance, "A hyperelliptic smoothness test II," *Proc. London Math. Soc.*, (3) **84** (2002), 105–146.
21. J.-F. Mestre, "Construction de courbes de genre 2 à partir de leurs modules," in *Effective methods in algebraic geometry*, Birkhäuser Progr. Math. **94**, 1991, 313–334.
22. J. Neukirch, *Algebraic Number Theory*, trans. Norbert Schappacher, Springer-Verlag, Berlin, 1999.
23. S. Ross, *A First Course in Probability*, 5th ed., Prentice-Hall, Upper Saddle River, NJ, 1998.
24. E. Schaefer, "A new proof for the non-degeneracy of the Frey-Rück pairing and a connection to isogenies over the base field," in *Computational aspects of algebraic curves*, *Lecture Notes Ser. Comput.* **13**, World Sci. Publ., Hackensack, NJ, 2005, 1–12.
25. G. Shimura. *Abelian varieties with complex multiplication and modular functions*, *Princeton Mathematical Series* **46**, Princeton University Press, Princeton, NJ, 1998.
26. A.-M. Spallek, "Kurven vom Geschlecht 2 und ihre Anwendung in Public-Key-Kryptosystemen," Ph.D. thesis, Institut für Experimentelle Mathematik, Universität GH Essen, 1994.
27. B. K. Spearman, K. S. Williams, "Relative integral bases for quartic fields over quadratic subfields," *Acta Math. Hungar.* **70** (1996), 185–192.

28. P. van Wamelen, "Examples of genus two CM curves defined over the rationals," *Math. Comp.* **68** (1999), 307–320.
29. A. Weng, "Constructing hyperelliptic curves of genus 2 suitable for cryptography," *Math. Comp.* **72** (2003), 435–458.

Complex multiplication and canonical lifts

David R. Kohel*

Institut de Mathématiques de Luminy
case 930, F13288 Marseille cedex 9, France
E-mail : kohel@iml.univ-mrs.fr

The problem of constructing CM invariants of higher dimensional abelian varieties presents significant new challenges relative to CM constructions in dimension 1. Algorithms for p-adic canonical lifts give rise to very efficient means of constructing high-precision approximations to CM points on moduli spaces of abelian varieties. In particular, algorithms for 2-adic and 3-adic lifting of Frobenius give rise to CM constructions in dimension 2 (see [6] and [2]). We analyse the Galois-theoretic structure of CM points in higher dimension and combine geometric and arithmetic conditions to derive new p-adic canonical lifting algorithms using the ℓ-adic torsion structure of an ordinary abelian variety.

1. Introduction

The construction of CM invariants of abelian varieties holds interest from both a theoretical point of view, with connections to class field theory, and from a cryptographic point of view, with its application to the construction of abelian varieties over large finite fields with known prime group order. The advancement of canonical lifting algorithms in arithmetic geometry, following the original work of Satoh [13] on computing the zeta function of an elliptic curves, has provided a p-adic approach to constructive CM algorithms (following [3] and in higher dimension [6] and [2]).

In Section 2 we recall the classical theory of complex multiplication and class field theory with the view of understanding the geometry and Galois theory of the zero-dimensional schemes of CM invariants. We illustrate by examples the main pathologies which can arise in dimension 2. In Section 3 we recall the background on canonical lifts and the indicate main mechanism for constructing a canonical lift as the lift of an isogeny together with

*The author was employed at the University of Sydney at the time this research was carried out, with support from the Austrlian Research Council, grant DP0453134.

associated Galois relations. In Section 4 we treat explicit algorithms for constructing canonical lifts, in particular an approach to canonical lifting which lifts the ℓ-adic torsion structure of an abelian variety in characteristic p. We treat in detail an elementary and efficient 2-adic AGM algorithm for genus 2 curves and the ℓ-adic utilization of Richelot isogenies. We conclude with discussion of the main algorithm obstacles and potential directions for resolving them.

2. Complex multiplication

The Main Theorem of Complex Multiplication gives the relation between the ideal classes of a CM order $\mathcal{O}$ and abelian varieties with endomorphism ring $\mathcal{O}$. We recall this relationship for elliptic curves and its generalization to higher dimension.

2.1. *Complex multiplication in genus 1*

In genus 1, the j-variant of an elliptic curve with CM by a maximal order $\mathcal{O}_K$ in K, generates the Hilbert class field $H = K(j)/K$. More precisely, an embedding $K \to \mathbb{C}$ gives the relation between ideals of $\mathcal{O}_K$ and isomorphism classes of elliptic curves over $\mathbb{C}$:

$$\mathfrak{a} \longmapsto E_{\mathfrak{a}} = \mathbb{C}/\mathfrak{a}^{-1}.$$

The Artin isomorphism $\sigma : \mathrm{Gal}(H/K) \cong \mathrm{Cl}(\mathcal{O}_K)$, determines an action on $\{E_{\mathfrak{a}}\}$ compatible induced isogenies

$$E_{\mathfrak{a}} \to E_{\mathfrak{a}\mathfrak{p}} \cong_{\mathbb{C}} E_{\mathfrak{a}}^{\sigma(\mathfrak{p})}$$

The Galois action on $\{E_{\mathfrak{a}}\}$ may be determined on any model for $E_{\mathfrak{a}}$ over H.

A CM construction is an algorithm for the construction of invariants of an abelian variety with complex multiplication. The traditional method for elliptic curves is to evaluate the j-function at points τ in the upper half Poincaré plane, which correspond to lattices with complex multiplication. The objective of this algorithm is to determine the minimal polynomial $H_D(x)$ for $j(\tau)$ over $\mathbb{Q}$. Identifying the j-line $\mathbb{A}^1 = \mathrm{Spec}(\mathbb{Q}[x])$, this polynomial defines a zero dimensional subscheme of $\mathbb{A}^1 \subset \mathbb{P}^1 \cong X(1)$.

2.2. *Complex multiplication in higher dimension*

Suppose now that K is a CM field of degree $2g$ with totally real subfield F. We recall the analogous construction of an abelian variety over $\mathbb{C}$ with

complex multiplication by the maximal order $\mathcal{O}_K$ (c.f. Shimura [12, §14]). Let $\Phi = (\phi_1, \ldots, \phi_g)$ be a CM-type, consisting of a g-tuple of pairwise non-complex conjugate embeddings of K in $\mathbb{C}$. Then Φ defines a map $K \to \mathbb{C}^g$ by

$$z \longmapsto (z_1, \ldots, z_g) = (\phi_1(z), \ldots, \phi_g(z)).$$

The embedding Φ determines a complex abelian variety $A(\mathbb{C}) = \mathbb{C}^g/\Phi(\mathfrak{a}^{-1})$ with dual abelian variety $\hat{A}(\mathbb{C}) = \mathbb{C}^g/\Phi(\bar{\mathfrak{a}}\mathfrak{D}_K^{-1})$, where $\mathfrak{D}_K$ is the different $\{\alpha \in \mathcal{O}_K : \mathrm{Tr}_{\mathbb{Q}}^K(\alpha\mathcal{O}_K) \subseteq \mathbb{Z}\}$.

For any purely imaginary element ζ in K such that $\zeta\mathfrak{D}_K \subset \mathfrak{a}\bar{\mathfrak{a}}$, with $\mathrm{Im}(\phi_j(\zeta)) > 0$ for all j, we have a *polarization* of abelian varieties:

$$\Phi(\zeta) : A(\mathbb{C}) = \mathbb{C}^g/\Phi(\mathfrak{a}^{-1}) \longrightarrow \hat{A}(\mathbb{C}) = \mathbb{C}^g/\Phi(\bar{\mathfrak{a}}\mathfrak{D}_K^{-1}),$$

given by $z \mapsto \Phi(\zeta)z := (\phi_1(\zeta)z_1, \ldots, \phi_g(\zeta)z_g)$. The polarization is said to be *principal* if $\Phi(\zeta)$ is an isomorphism, which holds if and only if $\zeta\mathfrak{D}_K = \mathfrak{a}\bar{\mathfrak{a}}$. This motivates the following definition.

Definition. An ideal $\mathfrak{a}$ in $\mathcal{O}_K$ is *principally polarizable* if there exists a purely imaginary element ζ in $\mathcal{O}_K$ with $\mathrm{Im}(\phi_j(\zeta)) > 0$ for all j, and such that $\zeta\mathfrak{D}_K = \mathfrak{a}\bar{\mathfrak{a}}$.

The property of being principally polarizable is a property of the ideal class, hence we may refer to a principally polarizable ideal class in $\mathcal{O}_K$. In general the polarization class is defined to be an ideal $\mathfrak{c}$ of $\mathcal{O}_F$ such that $\mathfrak{c}\zeta\mathfrak{D}_K = \mathfrak{a}\bar{\mathfrak{a}}$, well-defined in $\mathrm{Cl}^+(\mathcal{O}_F)$ for any purely imaginary ζ as above.

The set of polarized abelian varieties with polarization class $\mathfrak{c}$ are acted on by pairs $(\mathfrak{a}, \alpha)$ such that $\mathfrak{a}\bar{\mathfrak{a}} = (\alpha)$ for totally positive α in $\mathcal{O}_F$. The existence of α is equivalent to $\mathfrak{a}$ being in the kernel of the homomorphism

$$\pi : \mathrm{Cl}(\mathcal{O}_K) \longrightarrow \mathrm{Cl}^+(\mathcal{O}_F),$$

where $\pi(\mathfrak{a}) = \mathcal{O}_F \cap \mathfrak{a}\bar{\mathfrak{a}}$. The set of pairs $(\mathfrak{a}, \alpha)$ forms a group $\mathfrak{C}(\mathcal{O}_K)$ with identity $(\mathcal{O}_K, 1)$, which is an extension of $\ker(N)$ by the group of totally positive units $(\mathcal{O}_F^*)^+$ modulo $N_F^K(\mathcal{O}_K^*)$ (either trivial or of exponent 2).

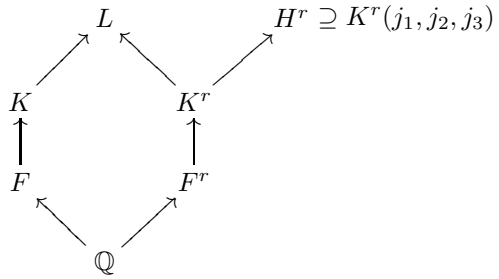
In general the maximal order may not be in the principally polarizable class but the following lemma asserts the existence of some polarized abelian variety in each polarization class $\mathfrak{c}$ in $\mathrm{Cl}^+(\mathcal{O}_F)$. It then follows that the isomorphism classes of polarized abelian varieties with CM by $\mathcal{O}_K$ are partioned into polarization classes by $\mathrm{Cl}^+(\mathcal{O}_F)$, each of which is acted on faithfully and transitively by $\mathfrak{C}(\mathcal{O}_K)$.

Lemma 2.1. *Let K be a CM field such that K/F is not unramified. Then for every class $\mathfrak{c}$ in $\mathrm{Cl}^+(\mathcal{O}_F)$ there exists $\mathfrak{a}$ in $\mathrm{Cl}(\mathcal{O}_K)$ with polarization*

class $\mathfrak{c}$. *Moreover there exists an ideal class* $\mathfrak{d}$ *in* $\mathrm{Cl}^+(\mathcal{O}_F)$ *such that* $\pi^{-1}(\mathfrak{d})$ *consists of the set of principally polarizable ideal classes in* $\mathrm{Cl}(\mathcal{O}_K)$.

Proof. Since $\mathfrak{D}_K$ is generated by elements of the form $z - \bar{z}$, both $\mathfrak{a}\bar{\mathfrak{a}}$ and $\zeta\mathfrak{D}_K$ (for any purely imaginary ζ) are generated by ideals of $\mathcal{O}_F$. Since K/F is not unramified, by class field theory $\mathrm{Cl}(\mathcal{O}_F)$ injects into $\mathrm{Cl}(\mathcal{O}_K)$, so we can find $\mathfrak{a}$ and ζ such that $\mathfrak{a}\bar{\mathfrak{a}}$ is in the class as $\mathfrak{c}\zeta\mathfrak{D}_K$. □

We now seek a description for the Galois action on the invariants of principally polarized CM abelian varieties. We specialize to CM abelian surfaces, for which the endomorphism algebra is a quartic CM field K. Generically such a field is non-Galois over $\mathbb{Q}$, and its normal closure is a degree 2 extension L/K with Galois group D_4 over $\mathbb{Q}$. There exist a triple of absolute Igusa invariants (j_1, j_2, j_3) associated to an ideal class in a maximal order $\mathcal{O}_K$ with principal polarization and CM-type Φ, contained in the Hilbert class field H^r of the reflex field K^r:



The field K^r may be constructed in terms of the CM-type Φ but is unique up to isomorphism.

The class group $\mathrm{Cl}(\mathcal{O}_{K^r})$ acts on the group $\mathfrak{C}(\mathcal{O}_K)$ by means of the homomorphism:

$$\begin{aligned} \mathrm{Gal}(H^r/K^r) \cong \mathrm{Cl}(\mathcal{O}_{K^r}) &\longrightarrow \mathfrak{C}(\mathcal{O}_K) \\ \mathfrak{c} &\longmapsto \left(\mathrm{N}_\Phi(\mathfrak{c}), \mathrm{N}_{\mathbb{Q}}^{K^r}(\mathfrak{c})\right) \end{aligned} \tag{1}$$

where $\mathrm{N}_\Phi(\mathfrak{c}) = \mathrm{N}_K^L(\mathfrak{c}\mathcal{O}_L)$. Composing with multiplication in $\mathfrak{C}(\mathcal{O}_K)$, we obtain the Galois action:

$$\mathrm{Gal}(H^r/K^r) \times \mathfrak{C}(\mathcal{O}_K) \to \mathfrak{C}(\mathcal{O}_K).$$

The homomorphism (1) can fail to be injective (hence $\{j_1, j_2, j_3\}$ does not generate H^r) or fail to be surjective (in which case the Galois action is not transitive, so there are multiple Galois orbits of invariants).

As an example, failure of injectivity occurs for the CM field $K \cong \mathbb{Q}[x]/(x^4 + 46x^2 + 257)$ of class number 1 and $\mathfrak{C}(\mathcal{O}_K) = \{(\mathcal{O}_K, 1)\}$. The reflex field $K^r \cong \mathbb{Q}[x]/(x^4 + 23x^2 + 68)$, on the other hand, has class number 3, so $\mathrm{Gal}(H^r/K^r)$ maps to the unique trivial class in $\mathfrak{C}(\mathcal{O}_K)$. Note, however, that $\mathfrak{C}(\mathcal{O}_{K^r})$ is also the trivial group since $\mathrm{Cl}^+(\mathcal{O}_{F^r})$ is a group of order 3, so that the Galois action is trivially transitive on both groups. The reflex field also provides an example where the maximal order does not admit a principal polarization since the different is not in the principal ideal class.

The first examples of multiple orbits occur with class number 2. In particular, the CM invariants associated to the maximal order of the quartic CM field $K \cong \mathbb{Q}[x]/(x^4+7x^2+5)$ and its reflex field $K^r \cong \mathbb{Q}[x]/(x^4+11x^2+29)$ have trivial action by $\mathrm{Gal}(H^r/K^r)$ and $\mathrm{Gal}(H/K)$, respectively. The maximal orders of K and K^r each determine two subschemes of degree 2 over $\mathbb{Q}$, which split over their totally real subfields — the Galois conjugate pairs determine distinct CM-types on the associated CM lattices.

The action of the absolute Galois group $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$, is more a subtle question, but relevant for determining the degree of the corresponding zero-dimensional schemes of CM points. The action of $\mathrm{Gal}(H^r/F^r)$ may be determined from the action of complex conjugation on ideal classes, and in general any automorphism of $\bar{\mathbb{Q}}/\mathbb{Q}$ which acts nontrivially on F^r will change the CM-type of a lattice. Thus the scheme over $\mathbb{Q}$ will represent Galois orbits from each of the possible CM-types.

Constructive CM

An analytic construction for dimension 2 uses theta functions on Siegel upper half space to determine points (j_1, j_2, j_3) in $\mathcal{M}_2(\mathbb{C})$, the moduli space of curves of genus 2 (which we identify with its image in the moduli space $\mathcal{A}_2(\mathbb{C})$ of principally polarized abelian surfaces). The result of a CM construction is an ideal in $\mathbb{Q}[j_1, j_2, j_3]$ defining the zero dimensional scheme over $\mathbb{Q}$ whose defining relations vanish on the Galois orbit of the point (j_1, j_2, j_3). In Section 4 we describe analogous algorithms for constructing these ideals, using p-adic canonical lifts.

Example. The curves $y^2 = x^5 + 1$ and $y^2 = x^6 + 1$ have absolute Igusa invariants (j_1, j_2, j_3) equal to

$$(0, 0, 0) \text{ and } (6400000/3, 440000/9, -32000/81).$$

Thus their respective defining ideals are

$$(j_1, j_2, j_3) \text{ and } (3j_1 - 6400000, 9j_2 - 440000, 81j_3 + 32000).$$

There are 19 such CM curve invariants known to exist over the rationals [14], each of which arises from an order in a cyclic quartic CM field.

In general the set of CM invariants forms a zero-dimensional subscheme of the moduli space $\mathcal{M}_2$ of genus two curves. To take an example of a non-normal quartic CM field of class number one, $K = \mathbb{Q}[x]/(x^4 + 13x^2 + 41)$, the set of absolute Igusa invariants (j_1, j_2, j_4) for a curve whose Jacobian has endormorphism ring $\mathcal{O}_K$ vanishes on the ideal of relations:

$$\begin{gathered}(4j_1^2 + 115322697j_1 - 10896201253125,\\ 64j_2^2 + 26342415j_2 + 74733890625,\\ 1024j_4^2 - 13091625j_4 + 4408171875,\\ 85j_1 - 8973j_2 - 97200j_4,\\ 25j_1 - 2920j_2 - 38016j_4 + 2460375).\end{gathered}$$

This ideal describes a subscheme of $\mathcal{M}_2$ of degree 2, which splits over the real quadratic subfield $\mathbb{Q}(\sqrt{41})$ of the reflex field of K. Here we prefer to work with

$$(j_1, j_2, j_4) = (\frac{J_2^5}{J_{10}}, \frac{J_2^3 J_4}{J_{10}}, \frac{J_2 J_8}{J_{10}}),$$

with $j_3 = J_2^2 J_6/J_{10}$ replaced by $j_4 = (j3 - j2^2/j_1)/4$, which provides local invariants for ordinary curves in characteristic 2.†

3. Canonical lifts

Let A be an ordinary, simple abelian variety over a finite field k of characteristic p, and let $R = W(k)$ be its Witt ring. Then R is an extension of $\mathbb{Z}_p$ such that $[R : \mathbb{Z}_p] = [k : \mathbb{F}_p]$ equipped with a surjective homomorphism $R \to k$. Then the Frobenius automorphism of k given by $\sigma(x) = x^p$ lifts uniquely to a Frobenius automorphism $\sigma : R \to R$.

A canonical lift is an abelian variety $\tilde{A}/R$ such that

$$\tilde{A}/R \times_R k = A/k \text{ and } \operatorname{End}(\tilde{A}) = \operatorname{End}(A).$$

By the theory of Serre and Tate (see [11]), we know that an ordinary abelian variety over a finite field admits a unique canonical lift to R.

† A curve over a field of characteristic 2 is ordinary if and only if $J_2 \neq 0$, and the equality $4J_8 = J_2 J_6 - J_4^2$ implies that $j_1 j_3$ is congruent to j_2^2 at 2.

3.1. *Canonical lifting conditions*

We describe the general idea for construction of a canonical lift in terms of isogenies induced by a decomposition of the modules of ℓ-torsion before passing to explicit algorithms in the next section. We assume that a modular correspondence for $(\ell, \ldots, \ell)$-isogenies has been precomputed as a subvariety $\mathcal{X}$ in the product $\mathcal{A} \times \mathcal{A}$ of moduli spaces of principally polarized abelian varieties (with prescribed torsion or theta structure). Such correspondences have been used for constructing the canonical lift as a lift of the Frobenius isogeny of ordinary abelian varieties in characteristic $p = \ell$. This makes use of the canonical decomposition

$$A[p] = A[p]^{loc} \oplus A[p]^{et},$$

induced by the kernels of Frobenius and Verschiebung, respectively.

Suppose now that A is ordinary over a finite field of characteristic $p \neq \ell$ and that $\ell\mathcal{O} = \mathfrak{a}\bar{\mathfrak{a}}$ where $\mathcal{O} = \mathrm{End}(A)$. Then we have an analogous decomposition

$$A[\ell] = A[\mathfrak{a}] \oplus A[\bar{\mathfrak{a}}],$$

determined by the ideal factorization. Moreover $A[\mathfrak{a}](\bar{k})$ is isomorphic to $(\mathbb{Z}/\ell\mathbb{Z})^g$ and there exists an $(\ell, \ldots, \ell)$-isogeny $\varphi : A \longrightarrow B$ with $\ker \varphi = A[\mathfrak{a}]$ and $\mathrm{End}(A) = \mathrm{End}(B)$. Suppose moreover that B is a Galois image of A (as happens when the image of $\mathfrak{a}$ under the Artin map is in the group generated by Frobenius at p). The canonical lift $\tilde{A}/R$ of A/k is determined by a lifting of isogenies:

$$\tilde{\varphi} : \tilde{A} \longrightarrow \tilde{B} = \tilde{A}^{\sigma^r},$$

preserving $\tilde{A}[\ell] = \ker(\varphi) \oplus \tilde{\varphi}(\tilde{A}[p])^{\sigma^{-r}}$ for some r.

3.2. *Canonical lifts as CM constructions*

An algorithm for the construction of the p-adic canonical lift of an elliptic curve was introduced by Satoh [13], to determine the number of points on a given $E/\mathbb{F}_q$ (in small characteristic p). The algorithm constructs the canonically lifted $\tilde{\jmath}$ of an given ordinary j-invariant j in $\mathbb{F}_q$, as the unique point $(\tilde{\jmath}, \tilde{\jmath}^\sigma)$ on

$$X_0(p) \to X(1) \times X(1).$$

An algorithm of Mestre, in 2000, introduced the use of theta functions and the AGM. This algorithm determines canonically lifted invariants $(\tilde{x}, \tilde{x}^\sigma)$

in $X_0(8) \times X_0(8)$ (and residue characteristic 2). The latter method extends naturally to higher dimension.

The idea to apply canonical lifting techniques to CM constructions was introduced in 2002 by Couveignes and Henocq [3], by determining a high precision approximation to moduli of the canonical lift as a means of computing the Hilbert class polynomial on $X(1)$. This idea was extended to abelian surfaces (Jacobians of genus 2 curves) by Gaudry et al. [6] in characteristic 2 and (extending Mestre's AGM to $(3,3)$-isogenies) by Carls et al. [2] in characteristic 3.

Example. The j-invariant $\tilde{j}$ of the canonical lift of $E/\mathbb{F}_p$, whose endomorphism ring is the maximal order of $K = \mathrm{End}(E) \otimes \mathbb{Q}$, is an element of $\mathbb{Z}_p$. Nevertheless, it is algebraic and integral over $\mathbb{Z}$ and generates the Hilbert class field of K. For instance

$$E/\mathbb{F}_{59} : y^2 = x^3 + 31x + 54$$

has j-invariant 20, but its canonical lift in $\mathbb{Z}_{59}$ is

$$\tilde{j} = 20 + 53 \cdot 59 + 0 \cdot 59^2 + 57 \cdot 59^3 + 9 \cdot 59^4 + 3 \cdot 59^5 + 5 \cdot 59^6 + \cdots$$

By lifting to sufficient precision we verify that $\tilde{j}$ is a root of the Hilbert class polynomial

$$x^3 + 3491750x^2 - 5151296875x + 12771880859375.$$

4. Constructive CM algorithms

In general, a p-adic algorithm for constructive CM must

- construct the lifted invariant (to some finite precision), and
- recognize an algebraic number from its approximation.

The first step replaces the p-adic numbers with complex numbers in analogous analytic constructions. Rather than a period lattice, the input is a suitable curve which we lift p-adically. The second step uses an LLL reconstruction, from one or multiple points on the CM subscheme. Finding suitable input curves, whose Jacobian has endomorphism ring which is a maximal order of low class number, is the primary difficulty in the first step. The height of the moduli points (hence the resulting output size) presents the major challenge to the LLL phase. Currently several constructive CM algorithms for genus 2 CM moduli exist:

- 2-adic lifting of $(2,2)$-isogenies (Gaudry, Houtmann, K., Ritzenthaler, Weng [6]), and

- 3-adic lifting of $(3,3)$-isogenies (Carls, K., Lubicz [2]).

We first describe an AGM recursion for 2-adic lifting, which provides a simplified yet efficient algorithm for carrying out Mestre's AGM lift in characteristic 2 (c.f. Lercier and Lubicz's treatment [10] and the construction in terms of Richelot isogenies in [6]). Then we introduce a new p-adic lifting of $(2,2)$-isogenies, by adapting the modular Richelot correspondences used in [6] to any odd characteristic p.

4.1. *Canonical 2-adic AGM algorithm*

We give an elementary version of the AGM recursion for ordinary curves of genus 2, by finding an explicit parametrisation of theta null points in terms of invariants of curves. We differ from the standard parametrisation of 2-theta null points in a neighborhood of a point $(1:1:1:1)$ which yields a less natural parametrisation.[‡] The simplicity and elegance of the equations justify giving particular treatment of the AGM algorithm relevant to the point $(1:0:0:0)$.

Rosenhain invariants in characteristic 2.

Over an extension splitting the Weierstrass points, a genus 2 curve C over a field k of characteristic 2 takes the form:

$$y^2 + x(x+1)y = x(x+1)u(x)$$

where $u(x)$ is a polynomial of degree 3, divisible by a linear factor $x + x_0$ for x_0 not in $\{0,1\}$. We set

$$a_1 = u(0),\ a_2 = u(1),\ a_3 = u(\infty),$$

where $u(\infty)$ is defined to be the leading coefficient of $u(x)$. The geometric isomorphism class of the curve is determined by the triple (a_1, a_2, a_3), independent of the value of x_0 $(\neq 0, 1)$, and provides a characteristic 2 analogue of the Rosenhain invariants $(\lambda_1, \lambda_2, \lambda_3)$ of a curve

$$y^2 = x(x-1)(x-\lambda_1)(x-\lambda_2)(x-\lambda_3)$$

[‡]In the neighborhood of $(1:1:1:1)$ suggested in Lercier and Lubicz [10] one obtains an affine parametrisation $(1:1+4t_1:1+4t_2:1+4(-t_1-t_2+2t_3)$ which lacks the symmetry and smoothness properties described here for the neighborhood of $(1:0:0:0)$. The use of a neighborhood of the latter point was suggested by Carls [1].

over any field of characteristic different from 2. Indeed, if $R = W(k)$ is the Witt ring of k, the curve

$$y^2 = x(x-1)(x-4\tilde{a}_1)(x-1-4\tilde{a}_2)(-4\tilde{a}_3x+1).$$

gives a lift of C to $K = R \otimes \mathbb{Q}$ for arbitrary lifts of a_i to $\tilde{a}_i$ in R. Thus (a_1, a_2, a_3) is a local system of coordinates at 2 for the Rosenhain invariants $(4\tilde{a}_1, 1+4\tilde{a}_2, 1/(4\tilde{a}_3))$.

Theta null points.

We refer to a theta null point with respect to a $(\mathbb{Z}/2\mathbb{Z})^g$-theta structure as a 2-theta null point. We consider a projective embedding provided by the system of 2-theta null constants:

$$(x_{00} : x_{01} : x_{10} : x_{11}) = \big(\vartheta\left[\begin{smallmatrix}00\\00\end{smallmatrix}\right](0,\tau) : \vartheta\left[\begin{smallmatrix}01\\00\end{smallmatrix}\right](0,\tau) : \vartheta\left[\begin{smallmatrix}10\\00\end{smallmatrix}\right](0,\tau) : \vartheta\left[\begin{smallmatrix}11\\00\end{smallmatrix}\right](0,\tau)\big).$$

Given a genus 2 curve C/k over a finite field k of characteristic 2 with Witt ring $R = W(k)$, the canonical lift of $\mathrm{Jac}(C)$ to R admits a canonical $(\mathbb{Z}/2\mathbb{Z})^2$-theta null structure over R in the neighborhood of $(1:0:0:0)$, parametrised by (x_1, x_2, x_3), where

$$x_1 = \sqrt{a_2a_3},\ x_2 = \sqrt{a_1a_3},\ x_3 = \sqrt{a_1a_2},$$

by means of the map

$$(x_1, x_2, x_3) \longmapsto (1 : 2x_1 : 2x_2 : 2x_3).$$

Here $2x_i$ is well-defined as an element of $2R/4R \cong R/2R = k$ from x_i in k. Conversely we recover the curve from affine parameters (x_1, x_2, x_3), by setting

$$a_1 = x_2x_3/x_1,\ a_2 = x_1x_3/x_2,\ a_3 = x_1x_2/x_3.$$

This gives an initialisation of 2-theta null points, from which we derive a modular correspondence for 2-theta null points.

Duplication formulae.

Let $x_\varepsilon = \vartheta\left[\begin{smallmatrix}\varepsilon\\00\end{smallmatrix}\right](0,\tau)$ and $y_\varepsilon = \vartheta\left[\begin{smallmatrix}\varepsilon\\00\end{smallmatrix}\right](0,2\tau)$ be 2-theta null constants. Then the classical duplication formulae give the relations between 2-theta null points $\mathbf{x} = (x_{00} : x_{01} : x_{10} : x_{11})$ and $\mathbf{y} = (y_{00} : y_{01} : y_{10} : y_{11})$. Precisely the Riemann duplication formulas [4] are

$$y_\varepsilon^2 = \sum_{\delta \in \mathbb{F}_2^2} x_\varepsilon x_{\varepsilon+\delta}.$$

This yields the following defining relations for the modular correspondence defining 2-theta null points with $(\mathbb{Z}/2\mathbb{Z})^g$-isogenies

$$\begin{aligned}\Phi(\mathbf{x},\mathbf{y}) = \Big(&\frac{(x_{00}x_{02}+x_{20}x_{22})}{2}y_{00}^2 - \frac{(x_{00}^2+x_{02}^2+x_{20}^2+x_{22}^2)}{4}y_{01}^2,\\ &\frac{(x_{00}x_{20}+x_{02}x_{22})}{2}y_{00}^2 - \frac{(x_{00}^2+x_{02}^2+x_{20}^2+x_{22}^2)}{4}y_{10}^2,\\ &\frac{(x_{00}x_{22}+x_{02}x_{20})}{2}y_{00}^2 - \frac{(x_{00}^2+x_{02}^2+x_{20}^2+x_{22}^2)}{4}y_{11}^2\Big) = (0,0,0)\cdot\end{aligned}$$

In terms of our affine parametrisations $\mathbf{x} = (1 : 2x_1 : 2x_2 : 2x_3)$ and $\mathbf{y} = (1 : 2y_1 : 2y_2 : 2y_3)$, this gives the system of local equations:

$$(y_1+2y_2y_3-x_1^2u(y),\ y_2+2y_1y_3-x_2^2u(y),\ y_3+2y_1y_2-x_3^2u(y)) = (0,0,0), \quad (2)$$

where $u(y) = 1 + 4(y_1^2 + y_2^2 + y_3^2)$.

$$D_{\mathbf{x}}\Phi(\mathbf{x},\mathbf{y}) = -2u(y)\begin{pmatrix} x_1 & 0 & 0\\ 0 & x_2 & 0\\ 0 & 0 & x_3\end{pmatrix} \equiv 0 \bmod 2$$

and

$$D_{\mathbf{y}}\Phi(\mathbf{x},\mathbf{y}) = \begin{pmatrix} 1-8x_1^2y_1 & 2y_3-8x_2^2y_1 & 2y_2-8x_3^2y_1\\ 2y_3-8x_1^2y_2 & 1-8x_2^2y_2 & 2y_1-8x_3^2y_2\\ 2y_2-8x_1^2y_3 & 2y_1-8x_2^2y_3 & 1-8x_3^2y_3\end{pmatrix} \equiv 1 \bmod 2.$$

Moreover one sees from equations (2) that $(y_1, y_2, y_3) \equiv (x_1^2, x_2^2, x_3^2) \bmod 2$. The simultaneous solution of a root $\Phi(\mathbf{x},\mathbf{y}) = (0,0,0)$ by Newton-Raphson iteration, satisfying $\mathbf{y} = \mathbf{x}^\sigma$, yields an Artin-Schreier equation as described in Lercier and Lubicz [10].

Example. Let $C/\mathbb{F}_2$ be the curve

$$y^2 + (x^3 + x^2 + 1)y = (x^2 + 1)(x^3 + x^2 + 1).$$

By naïve point counting we find the characteristic polynomial of Frobenius $x^4 + x^3 + x^2 + 2x + 4$, which generates a quartic CM field of class number 1. Over the extension $\mathbb{F}_8 = \mathbb{F}_2[w]/(w^3 + w + 1)$, we obtain a model

$$y^2 + x(x + 1)y = x(x + 1)(w^5x^3 + w^6x^2 + w^2x + w^3),$$

whence $(a_0, a_1, a_2) = (w^3, w^6, w^5)$. By means of the above canonical lifting algorithm, we determine the lifted invariants and compute the absolute Igusa invariants $(\tilde{\jmath}_1, \tilde{\jmath}_2, \tilde{\jmath}_4)$ to sufficient precision to recover the ideal of

relations:

$$\begin{aligned}
&4j_1^2 + 8218017j_1 + 146211169851,\\
&32j_2^2 + 1394199j_2 + 12065509143,\\
&2048j_4^2 - 2807745_j4 + 615519801,\\
&644j_1 - 45615j_2 - 484928j_4 + 267501,\\
&777j_1 - 55059j_2 - 584512j_4 - 576156.
\end{aligned}$$

The group $\mathfrak{C}(\mathcal{O}_K)$ has order 1, and the resulting scheme splits into two rational points over the totally real subfield $\mathbb{Q}(\sqrt{17})$ of the reflex field, representing the two CM-types on $\mathcal{O}_K$.

4.2. *Canonical ℓ-adic Richelot lifting algorithm*

We show how the principles of this analytic parametrization can be applied to yield a canonical lifting algorithm where a correspondence is only implicitly defined in the product of rational spaces. The method applies to the above AGM correspondence, but uses the Richelot correspondences of Rosenhain invariants, as in Gaudry et al [6], which in general can be defined over a smaller degree extension of $\mathbb{F}_p$ (and $\mathbb{Z}_p$).

Let $C_{\mathbf{t}}/k$ be the genus 2 curve $y^2 = x(x-1)(x-t_0)(x-t_1)(x-t_2)$ over a finite field k of odd characteristic. A Richelot isogeny of the Jacobian of $C_{\mathbf{t}}$ is determined by a splitting

$$x(x-1)(x-t_0)(x-t_1)(x-t_2) = G_0(x)G_1(x)G_2(x)$$

where $G_0(x) = x(x-t_0)$, $G_2(x) = (x-1)(x-t_1)$, $G_2(x) = x - t_2$.

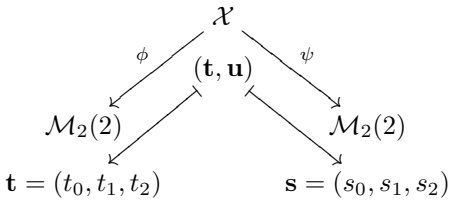
The codomain is the Jacobian of the curve $C'_{\mathbf{t}} : y^2 = \delta H_0(x)H_1(x)H_2(x)$ where δ is an explicit constant, and over some splitting field of the $H_i(x)$ we have:

$$\begin{aligned}
H_0(x) &= x^2 - 2t_2x + t_1t_2 - t_1 + t_2 = (x-u_0)(x-v_0),\\
H_1(x) &= -x^2 - 2t_2x + t_0t_2 = (x-u_1)(x-v_1),\\
H_2(x) &= (t_0 - t_1 - 1)x^2 + 2t_1x - t_0t_1 = (t_0 - t_1 - 1)(x-u_2)(x-v_2).
\end{aligned}$$

For any such triple $\mathbf{u} = (u_0, u_1, u_2)$, the conjugates v_i are determined from $\mathbf{t} = (t_0, t_1, t_2)$. By a choice of ordering for $\{u_0, u_1, u_2, v_0, v_1, v_2\}$ we obtain an isomorphism

$$C'_{\mathbf{t}} \cong C_{\mathbf{s}} : y^2 = x(x-1)(x-s_0)(x+s_1)(x+s_2).$$

This gives a space $\mathcal{X}$ with two finite morphisms to $\mathcal{M}_2(2)$:



such that $\mathbf{s} = (\psi_i(\mathbf{t},\mathbf{u}))$. Then $\mathcal{X}$ is determined in $\mathbb{A}^3 \times \mathbb{A}^3 \times \mathbb{A}^3$ by the polynomial equations

$$\Phi(\mathbf{t},\mathbf{u}) = (H_0(u_0), H_1(u_1), H_2(u_2)) = (0,0,0)$$
$$\Psi(\mathbf{t},\mathbf{u},\mathbf{s}) = (\Psi_0(\mathbf{t},\mathbf{u},\mathbf{s}), \Psi_1(t,u,s), \Psi_2(\mathbf{t},\mathbf{u},\mathbf{s})) = (0,0,0)$$

where Ψ_i is the numerator of the rational function $s_i - \psi_i(\mathbf{t},\mathbf{u})$.

We want to solve for $\mathbf{t}$ in $\mathbb{A}^3(R)$ to high precision, with auxillary point $\mathbf{u}$, such that $(\mathbf{t},\mathbf{u},\mathbf{s}) = (\mathbf{t},\mathbf{u},\mathbf{t}^{\sigma^r})$ satisfy these equations, given only the image of $\mathbf{t}$ in $\mathbb{A}^3(k)$. Assuming we have already determined $\mathbf{t}$ in $\mathbb{A}^3(R/p^{2m}R)$ such that its image in $\mathbb{A}^3(R/p^mR)$ is the canonical lift, we set

$$\mathbf{t}' = \mathbf{t} + p^m\Delta_{\mathbf{t}}, \text{ and } \mathbf{s}' = \mathbf{s} + p^m\Delta_{\mathbf{s}} = \mathbf{t}^{\sigma^r} + p^m\Delta_{\mathbf{t}}^{\sigma^r}$$

where $\mathbf{t}'$ is the canonical lift to $\mathbb{A}^3(R/p^{2m}R)$. We find $\mathbf{u}$ in $\mathbb{A}^3(R/p^{2m}R)$ by Hensel lifting such that $\Phi(\mathbf{t},\mathbf{u}) = (0,0,0)$, and suppose that

$$\mathbf{u}' = \mathbf{u} + p^m\Delta_{\mathbf{u}}$$

satisfies $\Phi(\mathbf{t}',\mathbf{u}') = (0,0,0)$. This gives the vector-matrix equation

$$\begin{aligned}(0,0,0) &= \Phi(\mathbf{t},\mathbf{u}) + p^m\Delta_{\mathbf{t}}D_{\mathbf{t}}\Phi(\mathbf{t},\mathbf{u}) + p^m\Delta_{\mathbf{u}}D_{\mathbf{u}}\Phi(\mathbf{t},\mathbf{u})\\ &= p^m\Big(\Delta_{\mathbf{t}}D_{\mathbf{t}}\Phi(\mathbf{t},\mathbf{u}) + \Delta_{\mathbf{u}}D_{\mathbf{u}}\Phi(\mathbf{t},\mathbf{u})\Big).\end{aligned}$$

Hence we have $\Delta_{\mathbf{u}} = -\Delta_{\mathbf{t}}D_{\mathbf{t}}\Phi(\mathbf{t},\mathbf{u})D_{\mathbf{u}}\Phi(\mathbf{t},\mathbf{u})^{-1}$. Similarly, we solve for $\Delta_{\mathbf{t}}$ such that

$$\begin{aligned}(0,0,0) = \Psi(\mathbf{t},\mathbf{u},\mathbf{s}) &+ p^m\Delta_{\mathbf{t}}D_{\mathbf{t}}\Psi(\mathbf{t},\mathbf{u},\mathbf{t}^{\sigma^r})\\ &+ p^m\Delta_{\mathbf{u}}D_{\mathbf{u}}\Psi(\mathbf{t},\mathbf{u},\mathbf{t}^{\sigma^r}) + p^m\Delta_{\mathbf{t}}^{\sigma}D_{\mathbf{s}}\Psi(\mathbf{t},\mathbf{u},\mathbf{t}^{\sigma^r}).\end{aligned}$$

Dividing by p^m and eliminating $\Delta_{\mathbf{u}}$, we obtain a vector-matrix equation

$$\Delta_{\mathbf{t}}^{\sigma^r}A + \Delta_{\mathbf{t}}B + \mathbf{c} = (0,0,0) \tag{3}$$

Unlike in the case of Frobenius lifts, the vector-matrix equations so obtained in general do not satisfy $B \equiv 0 \bmod p$, thus is not in the form of an Artin–Schreier equation. This means that there generally exist multiple solutions

modulo p to equation (3), and one must test whether each extends to the unique solution.

Example. Let $\mathbb{F}_{27} = \mathbb{F}_3[w]/(w^3 - w + 1)$, and let C be the curve

$$y^2 = x(x-1)(x-t_0)(x-t_1)(x-t_2),$$

where $\mathbf{t} = (t_0, t_1, t_2) = (w^{14}, w^8, 2)$. The point $\mathbf{s} = (s_0, s_1, s_2) = (w^{16}, w^{24}, 2)$ is the image of $\mathbf{t}$ by Frobenius and defines a second curve

$$y^2 = x(x-1)(x-s_0)(x-s_1)(x-s_2),$$

connected to the first by a Richelot correspondence (after renormalization). Applying the above lifting algorithm, we obtain a high precision lift of $\mathbf{t} = (t_0, t_1, t_2)$, and compute the absolute Igusa invariants (j_1, j_2, j_4) in R/p^mR, and use LLL lattice reduction to recover the algebraic relations among them. This yields the following polynomials for which (j_1, j_2, j_4) are zeros:

$$\begin{aligned}
&10460353203j_1^6 - 2580575774371539210j_1^5 + \\
&\quad 24762467241323829203127831j_1^4 - \\
&\quad 11315274154291362251887420761693 1j_1^3 - \\
&\quad 11614283201572167934644349880291166628 8j_1^2 - \\
&\quad 70782776480135088514937849133086022245140736j_1 - \\
&\quad 62317304708077035966402728779551311876832467230 72, \\
&282429536481j_2^6 - 101720638067873841 0j_2^5 + \\
&\quad 24881230456016762392454 7j_2^4 - \\
&\quad 93569901113311479610902034073j_2^3 + \\
&\quad 21637107789746630425279273638830 74j_2^2 - \\
&\quad 11272146035292913758697580625298514138 8j_2 + \\
&\quad 22265377293416386582386758988724792363081576, \\
&84333007705968230 4j_4^6 - 69928198180577770146048j_4^5 - \\
&\quad 14026747871338192671359918 4j_4^4 + \\
&\quad 33322274480663624199233151469 97j_4^3 + \\
&\quad 19431755806296265925420352017482148j_4^2 - \\
&\quad 32480753189175363543835184657189382877j_4 + \\
&\quad 34295760875987608803808408216247577819433
\end{aligned}$$

5. Conclusion

Several algorithmic obstacles present themselves when applying a p-adic CM construction. Since these algorithms take as input a curve over a small finite field, finding suitable input curves such that the endomorphism ring

of the Jacobian is a maximal order of small class number is crucial to their application. For this reason, the determination of the exact endomorphism ring $\mathcal{O} = \mathrm{End}(J)$, with

$$\mathbb{Z}[\pi, \bar{\pi}] \subseteq \mathcal{O} \subseteq \mathcal{O}_K,$$

is necessary in order to determine suitability of a chosen input curve. Recent work of Freeman and Lauter [5] addresses this problem by analysing the Frobenius action on ℓ-torsion points, when only small primes ℓ divide the index $[\mathcal{O}_K : \mathbb{Z}[\pi, \bar{\pi}]]$. A general method for constructing the graphs of (ℓ, ℓ)-isogenies is still needed to differentiate the orders between $\mathbb{Z}[\pi, \bar{\pi}]$ and $\mathcal{O}_K$ when the index $[\mathcal{O}_K : \mathbb{Z}[\pi, \bar{\pi}]]$ is divisible by a large prime (applying the same algorithmic approach as for elliptic curves [8]).

Once a suitable input curve has been found, the LLL reconstruction of algebraic relations (over $\mathbb{Q}$) for the invariants remains the limiting step of p-adic CM constructions. Combining the knowledge of the Galois action on CM points with explicit class field constructions has the potential to minimise this phase of the algorithm.

One of the motivations for CM constructions is the cryptographic application to producing abelian varieties whose number of points is prime or nearly prime over a large prime field $\mathbb{F}_p$. Currently, the performance of algorithms for determining the zeta function of genus 2 curves over prime fields place limitations on the use of random genus 2 curves over $\mathbb{F}_p$ in cryptography. Instead curve generation by CM construction is typically used, which we demonstrate in the folowing example.

Example. Let C be the curve $y^2 + x(x+1)y = x(x+1)(x^3 + x^2 + w^2x + w^3)$ over the finite field $\mathbb{F}_8 = \mathbb{F}_2[w]/(w^3 + w + 1)$. By naïve point counting, we find the characteristic polynomial of Frobenius is $x^4 + 4x^3 + 15x^2 + 32x + 64$. The curve is ordinary and has complex multiplication by the maximal order of $K = \mathbb{Q}[x]/(x^4 + 26x^2 + 449)$. The maximal order has class number 3, and there exist 6 isomorphism classes of principally polarized abelian varieties.

We construct the ideal of relations in Igusa invariants (j_1, j_2, j_4) from the canonical lift of the Jacobian of C. For example, the invariant j_1 satisfies a minimal polynomial:

$$\begin{aligned}
&H_1(x) = 2^{18}5^{36}7^{24}\,x^6 \\
&- 1118773039927368977400974047014016967290290543651580810546875000\,x^5 \\
&+ 5015125276905916795044208327674714215126845014038345476446629882636718750000\,x^4 \\
&- 1011240924278739178667628463373057504761454313557202566746822143270426385780826292 3\,x^3 \\
&+ 1182870002505886675645407447394061543981359784477927719285355412407973869920918282135218 75\,x^2 \\
&- 2^1 3^{50} 5^{10} 11^1 13^1 53^1 701^1 16319^1 69938793494948953569198870004032131926868578084899317\,x \\
&+ 3^{60} 5^{15} 23^5 409^5 179364113^5
\end{aligned}$$

Choosing the 120-bit prime

$$p = 954090659715830612807582649452910809,$$

and solving a norm equation in the endomorphism ring $\mathcal{O}_K$, we determine that the Jacobian of some curve over $\mathbb{F}_p$ with CM by $\mathcal{O}_K$ will have prime order

$$91028898695698888575311855828448102931141112827604802758431052540888444 9.$$

Solving for a solution to the system of equations over $\mathbb{F}_p$, we find a corresponding curve

$$\begin{aligned} C : y^2 = x^6 &+ 827864728926129278937584622188769650\, x^4 \\ &+ 102877610579816483342116736180407060\, x^3 \\ &+ 335099510136640078379392471445640199\, x^2 \\ &+ 351831044709132324687022261714141411\, x \\ &+ 274535330436225557527308493450553085. \end{aligned}$$

A test of a random point on the Jacobian verifies the group order.

Cryptographic CM database. A comprehensive database for CM invariants in genera 1 and 2 is being developed to provide a relational interface to CM fields K, their Hilbert class fields, and moduli of CM abelian varieties [9]. This database includes the output of CM constructions using the p-adic algorithms of Gaudry et al. [6], Carls et al. [2], the ℓ-adic variants described in this work, and complex analytic algorithms of Houtmann [7].

Acknowledgements. The author thanks the group CACAO of INRIA/LORIA for the invitation where this article was completed, the organizers of SAGA for the invitation to present this work, and discussions with R. Carls, P. Gaudry, and D. Lubicz.

References

1. R. Carls. Theta null points of 2-adic canonical lifts, Preprint available at `http://arxiv.org/abs/math.NT/0509092`, 2005.
2. R. Carls, D. Kohel, D. Lubicz, Higher dimensional 3-adic CM construction, to appear in *Journal of Algebra*.
3. J.-M. Couveignes and T. Henocq, Action of modular correspondences around CM points, in *Algorithmic number theory (Sydney, 2002)*, *Lecture Notes in Comp. Sci.*, **2369**, 234–243, 2002.
4. John D. Fay, *Theta functions on Riemann surfaces*, *Lect. Notes in Comp. Sci.*, **389**, Springer–Verlag, 1973.
5. D. Freeman and K. Lauter, Computing endomorphism rings of Jacobians of genus 2 curves over finite fields, Preprint available at `http://arxiv.org/abs/math.NT/0701305`, 2007.
6. P. Gaudry, T. Houtmann, D. Kohel, C. Ritzenthaler, and A. Weng, The 2-adic CM method for genus 2 curves with application to cryptography, *Asiacrypt 2006 (Shanghai) Lect. Notes in Comp. Sci.*, **4284**, 114–129, Springer–Verlag, 2006.

7. T. Houtmann, These, École Polytechnique, in preparation, 2007.
8. D. Kohel, *Endomorphism rings of elliptic curves over finite fields*, PhD thesis, University of California, Berkeley, 1996.
9. D. Kohel et al., ECHIDNA Databases for Algebra and Geometry Experimentation, `http://echidna.maths.usyd.edu.au/~kohel/dbs/`, 2007.
10. R. Lercier and D. Lubicz. A quasi-quadratic time algorithm for hyperelliptic curve point counting, *Ramanujan J.*, **12**, no. 3, 399–423, 2006.
11. J. Lubin, J. P. Serre and J. Tate, *Elliptic Curves and formal groups*, Notes available at `http://ma.utexas.edu/users/voloch/lst.html`, 1964.
12. G. Shimura. *Abelian Varieties with Complex Multiplication and Modular Functions*, Princeton University Press, 1998.
13. T. Satoh. The canonical lift of an ordinary elliptic curve over a finite field and its point counting, preprint, 1999.
14. P. van Wamelen. Examples of genus two CM curves defined over the rationals, *Math. Comp.*, **68**, no. 225, 227–320, 1999.

Two letters to Jaap Top

Jean-Pierre Serre

Collège de France
3, rue d'Ulm
75231 Paris cedex 05,
E-mail : serre@noos.fr

The article " On some questions of Serre on abelian threefolds " *by G. Lachaud and C. Ritzenthaler (cf. page 88) is based on a series of questions by J.-P. Serre on curves of genus* 3*. Since they have not been published, the Editors considered it worthwhile to include them here together with the related article.*

Paris, February 28, 2003

Dear M. Top,

Since you are interested in curves of genus 3 over finite fields, I would like to ask you some questions about this very interesting case. Maybe you have already thought about them ? Anyway, here they are:

Let us consider first an arbitrary field k, and an abelian variety A over k, of dim.3, with a principal polarization a. Assume that (A, a) is indecomposable, so that it is isomorphic (over an extension of k) with the Jacobian (J, θ) of a smooth algebraic curve X. By Torelli's theorem, we know that X has a natural k-structure, and that:

i) If X is hyperelliptic, one can find an isomorphism $(J, \theta) \to (A, a)$ which is defined over k.

ii) If X is not hyperelliptic, there exists a quadratic character

$$\epsilon : \mathrm{Gal}(k_s/k) \to \{1, -1\}$$

such that (J, θ) is isomorphic to the twist $(A, a)_\epsilon$ of (A, a) by ϵ.
(When k is finite, this implies that the trace of Frobenius for X is equal to $\pm$ (trace of Frobenius for A), a very annoying ambiguity !)

This raises two questions :

i?) How to decide (knowing only (A, a)) that we are in the hyperelliptic case ?

ii?) How to find the quadratic character ϵ of case ii) ?

Question i?) has been more or less "solved" by analysts, but their solution is not of much use. Question ii?) has not been seriously attacked, as far as I know. Still, classical literature (e.g. work of Klein and Igusa) contains the elements of an answer. I am not sure I understand the situation correctly, but here is what I think is known (or should be known):

Let me call $\omega = \omega(A)$ the dualizing sheaf of A, i.e. the 1-dimensional k-vector space $\omega(A) = \det t(A)^*$, where $t(A)$ is the tangent space of A, and $t(A)^*$ is its dual, i. e. $H^0(A, \Omega^1)$. Define similarly $\omega(X)$. For every integer m, write $\omega^{\otimes m}$ for the m-th tensor power of ω. Then, *there is a canonical element* $\Delta(A, a)$ of $\omega(A)^{\otimes 18}$ with the properties:

1) One has $\Delta(A, a) = 0$ if and only if X is hyperelliptic.

2) The quadratic character ϵ of ii) above is the one defined by the action of $\mathrm{Gal}(k_s/k)$ on the square roots of $\Delta(A, a)$.*

(Note that the square roots of $\Delta(A, a)$ make sense in $\omega(A)^{\otimes 9}$.)
The map $(A, a) \mapsto \Delta(A, a)$ may be viewed as a *Siegel modular form of weight* 18 *and level* 1. Up to a constant factor (which it would be important to compute), it should be the modular form χ_{18} of Igusa (Amer.J.Math. 89 (1967), pp.850-851). A proof of property 1) is given in Igusa's paper. As for 2), it should come from a result of Klein (Gesam.Abh.III, pp.460-462) which can be stated as follows:

For any smooth quartic X in $\mathbf{P}_2$, given by an equation $F(x, y, z) = 0$, we may consider the *discriminant* $\mathrm{Disc}(F)$ (I should say how it is defined: one takes the resultant of the partial derivatives $\partial F/\partial x$, $\partial F/\partial y$, $\partial F/\partial z$, and one divides it by the "best possible" integer, which is here 2^{14}, if I am not mistaken†. One then finds a polynomial of degree 27 in the coefficients of F which vanishes if and only if $F = 0$ is not smooth (in any characteristic, even char.2!)) Of course F depends on the coordinates used. But it is not

* I am assuming here that the characteristic is $\neq 2$.

† The correcting factor for a polynomial of degree d in n variables is $d^{((d-1)^n-(-1)^n)/d}$. Here $d = 4, n = 3$.

hard to transform it into a *canonical element* of $\omega(X)^9$, which I shall denote by $\mathrm{Disc}(X)$. Now Klein's result is (or should be) that $\mathrm{Disc}(X)$ is a square root of $\Delta(J,\theta)$. In particular, (A,a) is k-isomorphic to (J,θ) implies that $\Delta(A,a)$ is a square over k (and conversely). This implies assertion 2) above.

(I should make clear that I have not proved these statements. To do so, one should revise carefully the arguments of Igusa and Klein, and compare their definitions of $\Delta(A,a)$; both of them say that $\Delta(A,a)$ is the product of the 36 "even theta constants", but Igusa interprets these algebraically, and Klein analytically. There are periods involved. Moreover, Igusa does not consider the discriminant $\mathrm{Disc}(X)$.)

Assuming all these can be fixed up, we now have a way to decide concretely whether a given (A,a) is a Jacobian over k, or not: it is enough (!) to compute $\Delta(A,a)$, and to see whether it is a square or not.

How can we do that? The case which interests me most is the one where (A,a) is obtained from an elliptic curve E, with C.M. by a quadratic ring R, by "tensoring" E with an hermitian module M (rank 3, discr.1, positive definite, indecomposable). We get $(A,a) = E \otimes_R M$. One wants to compute its Δ-invariant in terms of E and M. This is a serious problem, for which I don't know a solution. But it can be attacked by computer means. Indeed, we may assume E ordinary (that is the interesting case anyway) and lift it to characteristic 0, hence to $\mathbf{C}$. When on $\mathbf{C}$, one can use the analytic definition of Δ by theta series and get at least an approximation of Δ. Presumably, this Δ should be an algebraic integer (or have an innocent denominator); hence an approximate value should determine it. One would then be able to reduce to char.p. It may even be that there is an explicit recipe for $\Delta(E \otimes M)$ using only the equation of E and some algebraic invariants of M; I don't know.

Sorry to be so vague! l hope you (or one of your students) can transform this sketch into something precise.

Best wishes

J-P. Serre

Paris, March 6, 2003

Dear M. Top,

I feel l should be more precise about the definition of the invariant which I calI Disc(X), and which belongs to $\omega(X)^{\otimes 9}$.

Consider a smooth quartic X in $\mathbf{P}_2$. Choose coordinates x_l, x_2, x_3 and choose an equation $F = 0$ for X, where F is homogeneous of degree 4. Call F_1, F_2, F_3 the partial derivatives of F. Put:

$$\mathrm{disc}(F) = 2^{-14}\mathrm{res}(F_1, F_2, F_3),$$

where res is the resultant. If we put

$$\alpha_1 = x_1(x_2.\mathrm{d}x_3 - x_3\mathrm{d}x_2)/F_1,$$

$$\alpha_2 = x_2(x_3.\mathrm{d}x_1 - x_1\mathrm{d}x_3)/F_2,$$

$$\alpha_3 = x_3(x_1.\mathrm{d}x_2 - x_2\mathrm{d}x_1)/F_3,$$

we get a basis of the differential forms of first kind on X, and one has $\alpha_i/\alpha_j = x_i/x_j$. The 1-dimensional vector space $\omega(X) = \det H^0(X, \Omega^1)$ contains the element $\alpha = \alpha_1 \wedge \alpha_2 \wedge \alpha_3$.

Now, put:

$$\mathrm{Disc}(X) = \mathrm{Disc}(F).\alpha^{\otimes 9} \quad \text{in} \quad \omega(X)^{\otimes 9}.$$

This element *does not depend on the choices* we have made (namely the coordinates in $\mathbf{P}_2$, and the equation F). It deserves to be called "the" discriminant of X.

Best wishes

J-P. Serre

On some questions of Serre on abelian threefolds

Gilles Lachaud

Institut de Mathématiques de Luminy
Université Aix-Marseille - CNRS
Luminy Case 907, 13288 Marseille Cedex 9 - FRANCE
E-mail: lachaud@iml.univ-mrs.fr

Christophe Ritzenthaler

Institut de Mathématiques de Luminy
Université Aix-Marseille - CNRS
Luminy Case 907, 13288 Marseille Cedex 9 - FRANCE
E-Mail : ritzent@iml.univ-mrs.fr

En genre 3*, le théorème de Torelli s'applique de façon moins satisfaisante : on doit extraire une mystérieuse racine carrée (J.-P.S., Collected Papers, n^o 129)*

J.-P. Serre asserted a precise form of Torelli Theorem for genus 3 curves, namely, an indecomposable principally polarized abelian threefold is a Jacobian if and only if some specific invariant is a square. We study here a three dimensional family of such threefolds, introduced by Howe, Leprevost and Poonen. By a new formulation, we link their results to Serre' s assersion. Then, we recover a formula of Klein related to the question for complex threefolds. In this case the invariant is a modular form of weight 18, and the result is proved using theta functions identities.

Keywords: Curve, Jacobian, abelian threefold, discriminant, Ciani quartic, modular form

1. Introduction

1.1. *Geometric Torelli's theorem*

Let K be an algebraically closed field. If X is a (smooth algebraic projective) curve of genus g over K, the Jacobian $\mathrm{Jac}\, X$ of X is an abelian variety of dimension g, and $\mathrm{Jac}\, X$ has a canonical principal polarization λ. We obtain in this way a morphism

$$\mathrm{Jac} : \mathsf{M}_g \longrightarrow \mathsf{A}_g$$

from the space M_g of (K-isomorphism classes of) curves of genus g to the space A_g of (K-isomorphism classes of) g-dimensional principally polarized abelian varieties (p.p.a.v.).

According to Torelli's Theorem, proved one century ago, the map $X \mapsto (\mathrm{Jac}\, X, \Theta)$ is injective. An algebraic proof was provided by Weil [18] half a century ago, and it is a long time studied question to characterize the image of this map.

If $g = 3$, these spaces are both of dimension $3g - 3 = g(g+1)/2 = 6$. According to Hoyt [7] and Oort and Ueno [15], the image of M_g is exactly the space of indecomposable principally polarized threefolds. Recall that (A, λ) is decomposable if there is an abelian subvariety B of A neither equal to 0 nor to A, such that the restriction of λ to B is a principal polarization, and indecomposable otherwise. This was a problem left unsolved by Weil in [18].

Given a principally polarized abelian threefold (A, λ) over K, two natural questions arise :

(1) How can we decide if the polarization is indecomposable ?
(2) How can we decide if A is the Jacobian of a a hyperelliptic curve ?

Actually, both questions were answered by Igusa in 1967 [9] when $K = \mathbb{C}$, making use of a particular modular form χ_{18} on the Siegel upper half-space (see Th. 3.1 below).

1.2. *Arithmetic Torelli's theorem*

Assume now that K is an arbitrary field. Then, as Serre noticed in [12], the above correspondence is no longer true. Let (A, λ) be a p.p.a.v. of dimension g over K, and assume that (A, λ) is isomorphic over $\overline{K}$ to the Jacobian of a curve $\mathcal{X}$ of genus g.

Theorem 1.1 (Serre). *The following alternative holds :*

(1) If $\mathcal{X}$ is hyperelliptic, there exists a model X/K of $\mathcal{X}$ and a K-isomorphism between the p.p.a.v. $(\mathrm{Jac}\, X, \Theta)$ and (A, λ).
(2) If $\mathcal{X}$ is non hyperelliptic, there exists a model X/K of $\mathcal{X}$ and a quadratic character $\varepsilon : \mathrm{Gal}(K_s/K) \to \{\pm 1\}$ such that $(\mathrm{Jac}\, X, \Theta)$ is isomorphic to the twist $(A, \lambda)_\varepsilon$ of (A, λ) by ε.

In particular, if ε is not trivial, this implies that $\mathrm{Jac}\, X$ is not isomorphic to A over K, but only over a quadratic extension, and (A, λ) is not isomorphic over K to the Jacobian of a curve.

1.3. *Serre's questions*

Let us come back to the case $g = 3$. Let there be given an indecomposable principally polarized abelian threefold (A, λ) defined over K. In a letter to J. Top (cf. page 84 of this book) in 2003, J.-P. Serre asked two questions:

(1) How to decide, knowing only (A, λ), that X is hyperelliptic ?
(2) If X is not hyperelliptic, how to find the quadratic character ε ?

He asserted, in the case $K \subset \mathbb{C}$, the following answer to these questions :

Assertion 1.1. *Let (A, λ) be an undecomposable principally polarized abelian threefold over K isomorphic over $\overline{K}$ to the Jacobian of a curve $\mathcal{X}$ of genus 3. Then there is an invariant $\chi_{18}(A, \lambda)$ such that*

(1) $\chi_{18}(A, \lambda) = 0$ is and only if $\mathcal{X}$ is hyperelliptic;
(2) the character ε is the one defined by the action of $\mathrm{Gal}(\bar{K}/K)$ on the square root of $\chi_{18}(A, \lambda)$.

We use here the notation χ_{18} to emphasize that this assertion was inspired by the results obtained by Igusa, and also much earlier by Klein [11] (see the remark after Cor. 4.2). In this article Klein relates (up to an undetermined constant) the modular form χ_{18} and the square of the discriminant of the quartic $\mathcal{X}$ (when $\chi_{18}(\tau) \neq 0$). This invariant seemed to Serre a good choice to find this "mysterious square root".

We plan to prove this assertion for a family of abelian threefolds which are isogenous to the product of three elliptic curves (see Cor. 4.1). This will rely on the work of Howe, Leprevost and Poonen [8] for which we propose a natural rephrasing. For any field K of characteristic different from 2, they consider abelian threefolds (A, λ) defined as a quotient of three elliptic curves (with the trivial polarization) by a certain subgroup of 2-torsion points. For this three-dimensional family, they make explicit the equation

of the related curve and express the character ε by a invariant T involving the coefficients of the elliptic curves. In the first part, we show that T can be naturally interpreted as a determinant. In a second phase, we take $K \subset \mathbb{C}$ and by uniformization, we express T in terms of certain Thetanullwerte of the elliptic curves. Then using the duplication and transformation formula we express the modular form $\chi_{18}(A, \lambda)$ in terms of the same Thetanullwerte and compare the two expressions. We also obtain a proof of Klein's result in this particular case and give the constant involved (see Cor. 4.2).

We describe now briefly the different sections. In Sec. 2, we define Ciani quartics, go back to the aforementioned results of [8], and show the relation with Serre's assertion (§ 2.5). In Sec. 3, we recall some general facts about abelian varieties over $\mathbb{C}$ (of arbitrary dimension) and introduce the modular function χ_k (§ 3.5). We prove Serre's assertion in Sec. 4. Finally, an appendix gathers some technical proofs, in particular the modularity of the form χ_k.

Acknowledgements. We would like to thank J.-P. Serre for interesting discussions and S. Meagher for relevant remarks.

2. Ciani Quartics

In this section we reformulate a result of [8] on a three-dimensional family of non-hyperelliptic genus 3 curves. In particular, this gives a more natural point of view on Prop. 15 of [8].

2.1. *Definition of Ciani quartics*

Edgardo Ciani gave in 1899 [3] a classification of nonsingular complex plane quartics curves based on the number of involutions in their automorphism group. We describe below the family of quartics admitting (at least) two commuting involutions (different from identity).

Let K be a field with char$K \neq 2$, and $\mathbf{Sym}_3(K)$ the vector space of symmetric matrices of size 3 with coefficients in K. Let

$$Q_m(x, y, z) = {}^t v.m.v, \quad v = (x^2, y^2, z^2), \quad m = \begin{bmatrix} a_1 & b_3 & b_2 \\ b_3 & a_2 & b_1 \\ b_2 & b_1 & a_3 \end{bmatrix} \in \mathbf{Sym}_3(K).$$

Then

$$Q_m(x, y, z) = a_1x^4 + a_2y^4 + a_3z^4 + 2(b_1y^2z^2 + b_2x^2z^2 + b_3x^2y^2)$$

is a ternary quartic, and the map $m \mapsto Q_m$ is an isomorphism of $\mathbf{Sym}_3(K)$ to the vector space of ternary quartic forms invariant under the three involutions

$$\sigma_1(x,y,z) = (-x,y,z), \quad \sigma_2(x,y,z) = (x,-y,z), \quad \sigma_3(x,y,z) = (x,y,-z).$$

The form Q_m is the zero locus of a plane quartic curve X_m, whose automorphism group contains the Vierergruppe $V_4 = (\mathbb{Z}/2\mathbb{Z})^2$.
If X_m is a nonsingular curve, we say that X_m is a *Ciani quartic* and that Q_m is a *Ciani form*. Now, E. Ciani (*loc. cit.*) proved that a plane quartic admitting two commuting involutions is geometrically isomorphic to a Ciani quartic (a more recent reference is [1]).

Proposition 2.1. *If X is a plane quartic curve defined over K, admitting at least two commuting involutions, also defined over K, then there is $m \in \mathbf{Sym}_3(K)$ such that X is isomorphic to X_m over K.*

Proof. Let M_1 and M_2 in $\mathrm{PGL}_3(K)$ inducing two commuting involutions of X. Then $M_i^2 = \alpha_i \mathbf{I}$ with $\alpha_i \in K$ and $\det(M_i)^2 = \alpha_i^3$, hence α_i is a square and we can assume, by dividing M_i by $\sqrt{\alpha_i}$, that $M_i^2 = \mathbf{I}$. The two matrices M_i commute so we can diagonalize them in the same basis : after a change of coordinates, we can suppose that M_i are (projectively) equal to

$$I_1 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This implies that a quartic equation $Q(x,y,z) = 0$ of X in these new coordinates must be invariant by the involutions σ_1 and σ_2 above, hence, Q is a Ciani form. □

2.2. *Discriminant of a ternary form*

Our definition of a Ciani form includes that its zero locus must be a nonsingular curve. This condition is fulfilled if and only if the discriminant of the form is not 0. In order to obtain a criterium for this condition, we develop an algorithm for the discriminant of a general ternary form.

The *multivariate resultant* $\mathrm{Res}(f_1, \ldots, f_n)$ of n forms $f_1, \ldots f_n$ in n variables with coefficients in a field K is an irreducible polynomial in the coefficients of $f_1, \ldots f_n$ which vanishes whenever $f_1, \ldots f_n$ have a common non-zero root. One requires that the resultant is irreducible over $\mathbb{Z}$, *i.e.* it has integral coefficients with greatest divisor equal to 1, and moreover

$$\mathrm{Res}(x_1^{d_1}, \ldots, x_n^{d_n}) = 1$$

for any $(d_1, \ldots, d_n) \in \mathbb{N}^n$. The resultant exists and is unique. There is a remarkable determinantal formula for the resultant of 3 ternary forms of the same degree d, due to Sylvester; see [6] for a modern exposition and a proof. We give this formula in the case $d = 3$. Then $\mathrm{Res}(f_1, f_2, f_3)$ is a form of degree 27 in 30 unknowns. We shall express $\mathrm{Res}(f_1, f_2, f_3)$ as the determinant of a square matrix of size 15.

Let I_d be the set of sequences $\nu = (\nu_1, \nu_2, \nu_3)$ with $\nu_1 + \nu_2 + \nu_3 = d$. The monomials $x^\nu = x_1^{\nu_1} x_2^{\nu_2} x_3^{\nu_3}$, where $\nu \in I_d$, form the canonical basis of the space V_d of ternary forms of degree d, which is of dimension $(d+1)(d+2)/2$.

For any monomial $x^\nu \in V_2$ with $\nu_1 + \nu_2 + \nu_3 = 2$, we choose arbitrary representations

$$f_i = x_1^{\nu_1+1} f_{i,1} + x_2^{\nu_2+1} f_{i,2} + x_3^{\nu_3+1} f_{i,3} \quad (1 \leq i \leq 3),$$

where $f_{i,j}$ are forms of degree $2 - \nu_j$, for $1 \leq i, j \leq 3$. Such a representation is always possible, although not unique. Now we define

$$S(x^\nu) = \det \begin{bmatrix} f_{1,1} & f_{1,2} & f_{1,3} \\ f_{2,1} & f_{2,2} & f_{2,3} \\ f_{3,1} & f_{3,2} & f_{3,3} \end{bmatrix}.$$

Note that this determinant is indeed a ternary form of degree

$$(2 - \nu_1) + (2 - \nu_2) + (2 - \nu_3) = 6 - 2 = 4.$$

Since the monomials x^ν with $\nu \in I_2$ make up a basis of V_2, we have thus defined a linear map $S : V_2 \longrightarrow V_4$. We consider the linear map

$$T : V_1 \times V_1 \times V_1 \times V_2 \longrightarrow V_4$$

given by

$$T(l_1, l_2, l_3, g) = l_1 f_1 + l_2 f_2 + l_3 f_3 + S(g).$$

Now one proves that the determinant of T is independent of the choices made in the definition of S, and *Sylvester's formula* holds :

$$\mathrm{Res}(f_1, f_2, f_3) = \det T.$$

Generally speaking, the matrix of T involves 864 monomials. Now, let Q be a ternary form of degree d, and X be the plane projective curve which is the zero locus of Q. Call q_1, q_2, q_3 the partial derivatives of Q. The *discriminant* of Q is $\mathrm{Disc}(Q) = \mathrm{Res}(q_1, q_2, q_3)$. It is a form of degree $3(d-1)^2$ in the coefficients of Q, and X is non singular if and only if $\mathrm{Disc}(Q) \neq 0$. The discriminant is an invariant of ternary forms : if $g \in \mathrm{GL}_3(K)$, then

$$\mathrm{Disc}(Q \circ g) = (\det g)^w \ \mathrm{Disc}(Q), \qquad \text{where } w = d(d-1)^2. \tag{1}$$

If Q is a quartic form $\mathrm{Disc}(Q)$ is a form of degree 27 in the coefficients, with $w = 36$ in (1). Applying Sylvester's formula, we get :

Proposition 2.2. *Let $m \in \mathbf{Sym}_3(K)$ and $c_i = a_j a_k - b_i^2$ the cofactor of a_i for $1 \leq i \leq 3$. If (q_1, q_2, q_3) are the partial derivatives of the ternary quartic Q_m, then $\mathrm{Disc}(Q_m = 2^{54}\mathsf{D}(m)$, where*

$$\mathsf{D}(m) = a_1\, a_2\, a_3\, (c_1\, c_2\, c_3)^2\, \det(m)^4.$$

Note that this result was obtained by Edge [4], in a more intricate way. We denote by S the set of $m \in \mathbf{Sym}_3(K)$ such that

$$a_1 a_2 a_3 \neq 0, \quad c_1 c_2 c_3 \neq 0.$$

Now, Prop. 2.2 implies that the curve X_m is nonsingular if and only if m belongs to the set $\mathsf{S}^\times = \mathsf{S} \cap GL_3(K)$.

Lemma 2.1. *The map $m \mapsto Q_m$ from $\mathsf{S}^\times$ to the set Q of Ciani forms is a bijection.*

The automorphisms of X_m induce a simple description of its Jacobian. In order to make it explicit, we need to introduce a certain product of elliptic curves.

2.3. *Product of elliptic curves*

We introduce the following notations : let

$$E_i : y^2 = x(x^2 - 4b_i x - 4c_i), \qquad (b_i \in K, c_i \in K^\times) \quad (i = 1, 2, 3),$$

three elliptic curves with $(0, 0)$ as a rational 2-torsion point. The discriminant of E_i is $\Delta_i = 2^{12} c_i^2 \delta_i$, where $\delta_i = b_i^2 + c_i \in K^\times$. We assume that there exists a square root $\rho \in K^\times$ of $\delta(A) = \delta_1\delta_2\delta_3$, that is, $\Delta_1\Delta_2\Delta_3$ is a square in K. We denote by A the set of products $E_1 \times E_2 \times E_3$ of such curves and we define

$$\widetilde{\mathsf{A}} = \left\{ \widetilde{A} = (A, \rho) \in \mathsf{A} \times K^\times \mid \rho^2 = \delta(A) \right\}.$$

If $\widetilde{A} \in \widetilde{\mathsf{A}}$, we put $a_i = \rho/\delta_i$ and

$$\mathbf{Mat}(\widetilde{A}) = \begin{bmatrix} a_1 & b_3 & b_2 \\ b_3 & a_2 & b_1 \\ b_2 & b_1 & a_3 \end{bmatrix} \in \mathsf{S}.$$

Conversely, a matrix $m \in \mathsf{S}$ defines an abelian threefold $A(m) \in \mathsf{A}$, which is the product of the curves

$$E_i : y^2 = x(x^2 - 4b_i x - 4c_i) \qquad (i = 1, 2, 3).$$

Then

$\delta_i = b_i^2 + c_i = a_j a_k \in K^\times, \Delta_1\Delta_2\Delta_3 = (2^{18}\, a_1a_2a_3c_1c_2c_3)^2, \delta(A) = (a_1a_2a_3)^2.$

We define $\rho(m) = a_1a_2a_3$, and $\mathbf{Ab}(m) = (A(m), \rho(m)) \in \widetilde{\mathsf{A}}$.

Lemma 2.2. *The maps*

$$\mathbf{Mat}\colon \widetilde{\mathsf{A}} \longrightarrow \mathsf{S}, \qquad \mathbf{Ab}\colon \mathsf{S} \longrightarrow \widetilde{\mathsf{A}},$$

are mutually inverse bijections.

The two lemmas 2.1 and 2.2 provide a natural map from the set Q of Ciani quartic forms to $\widetilde{\mathsf{A}}$. This map has actually a geometric meaning, and in order to explain it, we introduce the following notation. If $m \in \mathrm{GL}_3(K)$, we denote by $\mathrm{Cof}(m)$ the *cofactor matrix* of m, satisfying

$$m.{}^t\mathrm{Cof}\, m = \det m.\mathbf{I}, \quad \det \mathrm{Cof}\, m = (\det m)^2, \quad \mathrm{Cof}\,\mathrm{Cof}\, m = (\det m).m.$$

Let Q_m be a Ciani form associated to $m \in \mathsf{S}^\times$ and X_m be the corresponding Ciani quartic

$$X_m : Q_m(x,y,z) = F_m(x^2, y^2, z^2) = 0.$$

By quotient, we get three genus one curves

$$\begin{aligned} C_1 &:= X_m/ < 1, \sigma_1 > \ : F(yz, x^2, y^2) = 0, \\ C_2 &:= X_m/ < 1, \sigma_2 > \ : F(zx, y^2, z^2) = 0, \\ C_3 &:= X_m/ < 1, \sigma_3 > \ : F(xy, z^2, x^2) = 0, \end{aligned}$$

where σ_i $(i = 1, 2, 3)$ are the involutions of X_m. Another change of variables maps the genus 1 quartics C_i to the elliptic curves

$$F_i : y^2 = x(x^2 - 4d_i x - 4a_i \det(m)), \qquad (i = 1, 2, 3).$$

In this way, we get a map

$$\varphi : X_m \longrightarrow B_m = F_1 \times F_2 \times F_3.$$

Let us now look more closely at B_m. The identity

$$\mathrm{Cof}\,\mathrm{Cof}\, m = (\det m).m$$

implies that the cofactor of c_i is $a_i \det m$. Hence,

$$\mathbf{Ab}(\mathrm{Cof}\, m) = (B_m, c_1\, c_2\, c_3).$$

Since the Jacobian is the Albanese variety of X_m, we get a factorization

$$\begin{array}{ccc} X_m & & \\ {\scriptstyle\iota}\downarrow & \searrow^{\varphi} & \\ \operatorname{Jac} X_m & \xrightarrow{\Phi} & B_m \end{array}$$

where ι is a canonical embedding. Since the images of the regular differential forms on F_i make a basis of those on X_m, we obtain:

Proposition 2.3. *The map*

$$\Phi : \operatorname{Jac} X_m \longrightarrow A(\operatorname{Cof} m)$$

is a $(2,2,2)$*-isogeny defined over* K*.*

The correspondences

$$\begin{array}{ccccc} m & \longrightarrow & \operatorname{Cof} m & & \\ \downarrow & & \downarrow & & \\ Q_m & \longrightarrow & A(\operatorname{Cof} m) & \xrightarrow{isg} & \operatorname{Jac} X_m \end{array}$$

lead to a commutative diagram, where Q is the space of Ciani quartics over K:

$$\begin{array}{ccc} \mathsf{S}^{\times} & \xrightarrow{\text{Cof}} & \mathsf{S}^{\times} \\ \downarrow = & & \downarrow \mathbf{Ab} \\ \mathsf{Q} & \xrightarrow{\text{“Jac”}} & \widetilde{\mathsf{A}} \end{array}$$

In the next section we describe the kernel of the isogeny Φ.

2.4. *The theory of Howe, Leprevost and Poonen revisited*

The previous isogeny can be made more precise as we recall from [8]. There are some differences between their notation and ours, see the remark at the end of Sec.2.1 for a comparison. Let us introduce couples (A, W), where:

(1) $A \in \mathsf{A}$ as defined in § 2.3.

The Weil pairings on the factors combine to give a non degenerate alternating pairing e_2 on the finite group scheme $A[2]$ over K.

(2) W is a totally isotropic indecomposable subspace of $A[2]$ defined over K.

Choose a basis $(P_i, Q_i) \in E_i[2]$, that is, a level 2 structure on E_i. This defines a level 2 structure on A. In [8] (Lem.13) it is proved that after a labeling of the 2-torsion points we can write

$$W = \left\{ \begin{matrix} (O,O,O), & (O,Q_2,Q_3), & (Q_1,O,Q_3), & (Q_1,Q_2,O), \\ (P_1,P_2,P_3), & (P_1,R_2,R_3), & (R_1,P_2,R_3), & (R_1,R_2,P_3) \end{matrix} \right\} \tag{2}$$

with

$$Q_i = (0,0), \quad P_i = (0, 2b_i + \rho_i\,), \quad R_i = (0, 2b_i - \rho_i\,), \quad \rho_1\rho_2\rho_3 = \rho_W,$$

and the four possible choices of ρ_1, ρ_2, ρ_3 leading to the same value of $\rho_1\rho_2\rho_3$ give the same subgroup W. Conversely, if $(A, \rho) \in \widetilde{\mathsf{A}}$ is given, if we choose ρ_1, ρ_2, ρ_3 in such a way that $\rho_1\rho_2\rho_3 = \rho$, and if we define P_i and R_i as above, then we can define a subgroup W_ρ by (2).

Lemma 2.3. *The map $(A, \rho) \mapsto (A, W_\rho)$ from $\widetilde{\mathsf{A}}$ to the set of couples (A, W) as defined above is a bijection.*

We take on $A = A(m)$ the principal polarization λ which is the product of the canonical polarizations on each factor. Then we have a commutative diagram

$$\begin{array}{ccc} A & \xrightarrow{2\lambda} & A^\vee \\ {\scriptstyle\pi}\big\downarrow{\scriptstyle d^\circ 8} & & {\scriptstyle d^\circ 8}\big\uparrow{\scriptstyle\widehat{\pi}} \\ A' & \xrightarrow{\lambda'} & (A')^\vee \end{array}$$

with a unique principal polarization λ' on $A' = A'(m) = A(m)/W_{\rho(m)}$. From [8] (Prop.15) we get :

Theorem 2.1. *The composition of the isogeny Φ and of the projection π leads to an isomorphism of p.p.a.v.:*

$$\operatorname{Jac} X_m \longrightarrow A'(\operatorname{Cof} m).$$

As a corollary, for any m, the p.p.a.v. (A', λ') is indecomposable.

The reverse direction is more interesting and will give an algebraic answer to Serre's assertion.

2.5. *Relation with Serre's assertion*

We need the following elementary lemma from linear algebra.

Lemma 2.4. *The map $m \mapsto \operatorname{Cof} m$ induces an exact sequence*

$$1 \longrightarrow \{\pm 1\} \longrightarrow \mathrm{GL}_3(K) \xrightarrow{\mathrm{Cof}} G^{\times 2}(K) \longrightarrow 1$$

with

$$G^{\times 2}(K) = \left\{ m \in \mathrm{GL}_3(K) \mid \det m \in K^{\times 2} \right\}.$$

Let $(A, \rho) = \widetilde{A} \in \widetilde{\mathsf{A}}$ and $m = \mathbf{Mat}(\widetilde{A})$. Denote

$$\mathsf{T}(\widetilde{A}) := \det(m) = 2b_1b_2b_3 - \rho(\frac{b_1^2}{\delta_1} + \frac{b_2^2}{\delta_2} + \frac{b_3^2}{\delta_3} - 1).$$

Theorem 2.2. *The following results hold.*

(1) if $\mathsf{T}(\widetilde{A}) = 0$*, that is,* $m \in \mathsf{S} \setminus \mathsf{S}^{\times}$*, there is a hyperelliptic curve* X *of genus* 3 *such that* $A'(m)$ *is isomorphic to the Jacobian of* X*.*

(2) if $\mathsf{T}(\widetilde{A}) \neq 0$*, that is,* $m \in \mathsf{S}^{\times}$*, then there exists a non hyperelliptic curve of genus* 3 *defined over* K *whose Jacobian is isomorphic to* $A'(m)$ *if and only if* $\mathsf{T}(\widetilde{A})$ *is a square in* K*.*

Proof. The first part is [8] (Prop.14) where the hyperelliptic curve is constructed explicitly. For the second part, if $\det(m)$ is a square, then using Lem. 2.4, we see that there exists a matrix $m' \in \mathsf{S}^{\times}$ such that $m = \mathrm{Cof}(m')$ and we apply Th. 2.1. If $d = \det(m)$ is not a square, let $m_d = dm$ and $\widetilde{A}_d = \mathbf{Ab}(m_d)$. $A(m_d)$ is defined by

$$E_i : y^2 = x(x^2 - 4b_i dx - 4c_i d^2) \qquad (i = 1, 2, 3),$$

Thus $A(m_d)$ is a quadratic twist of $A(m)$. Now, $\det(m_d)$ is a square, so there exists m' such that $\mathrm{Jac}(X_{m'})$ is isomorphic to $A'(m_d)$. Since $A'(m_d)$ is a quadratic twist of $A'(m)$ and is the Jacobian of a non hyperelliptic curve, Th. 1.1 shows that $A'(m)$ cannot be a Jacobian. □

Corollary 2.1. *With the same notation as above:*

(1) if $\mathsf{T}(\widetilde{A}) \in K^{\times 2}$*, there is an isogeny defined over* K

$$\mathrm{Jac}\, X_{m'} \longrightarrow A(m), \quad \mathrm{Cof}\, m' = m,$$

(2) If $\mathsf{T}(\widetilde{A}) \notin K^{\times 2}$*, there is an isogeny defined over* K

$$\mathrm{Jac}\, X_{m'} \longrightarrow A(m_d) \quad \mathrm{Cof}\, m' = dm.$$

We hope to give in a near future a geometric interpretation of the connection of Serre's problem with the determinant of certain quadratic forms in the general case.

Remark 2.1. In [8], Howe, Leprevost and Poonen write the elliptic curves

$$y^2 = x(x^2 + A_i x + B_i) \qquad \text{avec } A_i, B_i \in K \qquad (i = 1, 2, 3).$$

So

$$A_i = -4b_i, \quad B_i = -4c_i,$$

$$\Delta_i = A_i^2 - 4B_i = 16(b_i^2 + c_i) = 16\delta_i,$$

$$d_i = -(A_i + 2x(P_i)) = 4b_i - 4(b_i + \rho_i) = -4\rho_i, \quad d_i^2 = \Delta_i.$$

And the factor

$$T_0(\widetilde{A}) = d_1d_2d_3(\frac{A_1^2}{\Delta_1} + \frac{A_2^2}{\Delta_2} + \frac{A_3^2}{\Delta_3} - 1) - 2A_1A_2A_3,$$

which is

$$T_0(\widetilde{A}) = 64[-\rho(\frac{b_1^2}{\delta_1} + \frac{b_2^2}{\delta_2} + \frac{b_3^2}{\delta_3} - 1) + 2b_1b_2b_3].$$

We then have $T_0(\widetilde{A}) = 64\,\mathsf{T}(\widetilde{A})$.

3. Complex abelian varieties

We recall in this section some well known propositions on abelian varieties over $\mathbb{C}$ and fix the notation.

3.1. *The symplectic group*

If V is a module of rank $2g$ over a commutative ring R and if E is a non-degenerate alternating bilinear form on V, a basis $(a_i)_{1\leq i\leq 2g}$ of V is said *symplectic* if the matrix $(E(a_i, a_j)) = J$, where

$$J = \begin{bmatrix} 0 & \mathbf{1}_g \\ -\mathbf{1}_g & 0 \end{bmatrix}.$$

The group of matrices

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathrm{SL}_{2g}(R)$$

such that $M.J.{}^tM = J$ is the *symplectic group* $\mathrm{Sp}_{2g}(R)$. It acts simply transitively on the set of symplectic bases of V.

Lemma 3.1 ([2] (Lem.8.2.1)). *If $M \in \mathrm{Sp}_{2g}(R)$ the following conditions are equivalent.*

(1) $M \in \mathrm{Sp}_{2g}(R)$.
(2) ${}^tA.C$ and ${}^tB.D$ are symmetric, and ${}^tA.D - {}^tC.B = \mathbf{1}_g$.
(3) $A{}^tB$ and $C{}^tD$ are symmetric and $A.{}^tD - B.{}^tC = \mathbf{1}_g$.

The group $\mathrm{Sp}_{2g}(R)$ is the group of R-rational points of a Chevalley group scheme Sp_{2g}, which contains certain remarkable subgroups defined as follows. The reductive subgroup $\mathbf{M}$ of Sp_{2g} is the subgroup which respects the canonical decomposition $\mathbb{Z}^{2g} = \mathbb{Z}^g \oplus \mathbb{Z}^g$. Elements of $\mathbf{M}(\mathbb{Z})$ are

$$M = \begin{bmatrix} A & 0 \\ 0 & {}^tA^{-1} \end{bmatrix}, \quad A \in \mathrm{GL}_g(\mathbb{Z}).$$

The unipotent subgroup $\mathbf{U}$ is the stability group leaving pointwise fixed the canonical totally isotropic subspace V_0, which is the first direct summand in the standard decomposition. Elements of $\mathbf{U}(\mathbb{Z})$ are

$$U = \begin{bmatrix} \mathbf{1}_g & B \\ 0 & \mathbf{1}_g \end{bmatrix}, \quad {}^tB = B, \quad B \in \mathbf{M}_g(\mathbb{Z}).$$

The unipotent subgroup $\mathbf{V}$ opposite to $\mathbf{U}$ is the stability group of the second direct summand in the standard decomposition. One has
$\mathbf{V} = {}^t\mathbf{U} = J.\mathbf{U}.J^{-1}$. Elements of $\mathbf{V}(\mathbb{Z})$ are

$$V = \begin{bmatrix} \mathbf{1}_g & 0 \\ C & \mathbf{1}_g \end{bmatrix}, \quad {}^tC = C, \quad C \in \mathbf{M}_g(\mathbb{Z}).$$

The subgroup $\mathbf{P} = \mathbf{M} \ltimes \mathbf{U}$ is the parabolic subgroup of Sp_{2g} normalizing V_0, and $\mathbf{P}$ is actually a maximal parabolic subgroup. Elements of $\mathbf{P}(\mathbb{Z})$ are

$$P = \begin{bmatrix} A & B \\ 0 & {}^tA^{-1} \end{bmatrix}, \quad A.{}^tB = B.{}^tA, \quad A \in \mathrm{GL}_g(\mathbb{Z}), \quad B \in \mathbf{M}_g(\mathbb{Z}).$$

3.2. *Abelian varieties*

Let $\Omega = [w_1 \dots w_{2g}] \in \mathbf{M}_{g,2g}(\mathbb{C})$, where $w_1, \dots, w_{2g}$ are columns vectors giving a basis of $\mathbb{C}^g$ on $\mathbb{R}$. It generates a lattice

$$\Lambda = \Omega\mathbb{Z}^{2g} \subset \mathbb{C}^g.$$

Let $\mathcal{R}_g$ be the set of matrices $\Omega \in \mathbf{M}_{g,2g}(\mathbb{C})$ satisfying the *Riemann conditions*

$$\Omega.J.{}^t\Omega = 0, \quad 2i(\overline{\Omega}.J^{-1}.{}^t\Omega)^{-1} > 0$$

(> 0 means positive definite). We call such a matrix Ω a *period matrix*. If $\Omega \in \mathcal{R}_g$, the torus $A_\Omega = \mathbb{C}^g/\Lambda$ is an abelian variety of dimension g with a principal polarization λ represented by the hermitian form
$H = 2i(\overline{\Omega}.J^{-1}.{}^t\Omega)^{-1}$ (see [2] (Lem.4.2.3)).
The group $\mathrm{GL}_g(\mathbb{C})$ acts on the left on $\mathcal{R}_g$. If we write

$$\Omega = [(w_1 \ \dots \ w_g)\,(w_{g+1} \ \dots \ w_{2g})] = [\Omega_1 \ \Omega_2], \quad \text{where } \Omega_i \in \mathbf{M}_g(\mathbb{C}),$$

we get $W.[\Omega_1\ \Omega_2] = [W.\Omega_1\ W.\Omega_2]$ for any $W \in \mathrm{GL}_g(\mathbb{C})$. This action induces an isomorphism of p.p.a.v. In particular if we choose $W = \Omega_2^{-1}$, we see that A_Ω is isomorphic to the p.p.a.v.

$$A_\tau = A_{\boldsymbol{\Omega}(\tau)}, \quad \boldsymbol{\Omega}(\tau) = [\tau\ \mathbf{1}_g], \quad \tau = \boldsymbol{\tau}(\Omega) = \Omega_2^{-1}\Omega_1,$$

and $\Omega \in \mathcal{R}_g$ if and only if $\boldsymbol{\tau}(\Omega)$ belongs to the Siegel upper half plane

$$\mathbb{H}_g = \left\{\tau \in \mathbf{M}_g(\mathbb{C}) \mid {}^t\tau = \tau,\ \Im\tau > 0\right\}.$$

We call a matrix $\tau \in \mathbb{H}_g$ a *Riemann matrix*. The *Siegel modular group* $\Gamma_g = \mathrm{Sp}_{2g}(\mathbb{Z})$ acts on the right on $\mathcal{R}_g$: if $\Omega \in \mathcal{R}_g$ and if $M \in \Gamma_g$,

$$\Omega.M = [\Omega_1\ \Omega_2]\begin{bmatrix} A & B \\ C & D \end{bmatrix} = [\Omega_1 A + \Omega_2 C\ \Omega_1 B + \Omega_2 D].$$

This action corresponds to a change of symplectic basis. The group Γ_g also acts on the left on the Siegel upper half plane : if $\tau \in \mathbb{H}_g$, we denote

$$M.\tau = (A\tau + B)(C\tau + D)^{-1}.$$

Both actions are linked by

$$M.\boldsymbol{\tau}(\Omega) = \boldsymbol{\tau}(\Omega.{}^tM). \tag{3}$$

3.3. *Isotropy and quotients*

For any maximal isotropic subgroup $V \subset \mathbb{F}_2^{2g}$, we have the *transporter*

$$\mathrm{Trans}(V) = \left\{M \in \mathrm{Sp}_{2g}(\mathbb{F}_2) \mid MV_0 = V\right\},$$

V_0 being the canonical maximal isotropic subgroup generated by the vectors $e_1, \ldots, e_g$ of the canonical basis. Since $\mathrm{Sp}_{2g}(\mathbb{F}_2)$ permutes transitively the maximal isotropic subgroups of $\mathbb{F}_2^{2g}$, the transporter is a left coset: $\mathrm{Trans}(V) = M_0\mathbf{P}(\mathbb{F}_2)$, for any $M_0 \in \mathrm{Trans}(V)$. Hence, the set of maximal isotropic subgroups is the quotient set $\mathrm{Sp}_{2g}(\mathbb{F}_2)/\mathbf{P}(\mathbb{F}_2)$, a set with 135 elements if $g = 3$.

Let now $\Omega \in \mathcal{R}_g$, $\Lambda = \Omega\mathbb{Z}^{2g}$ and $(A, \lambda) = (A_\Omega, H)$ be the corresponding p.p.a.v. of dimension g. The linear map $\alpha : \mathbb{Z}^{2g} \to \frac{1}{2}\Lambda$ such that

$$\alpha(x) = \frac{1}{2}\Omega.x$$

defines a level 2 symplectic structure on $A[2]$, that is, an isomorphism

$$\bar{\alpha} : \mathbb{F}_2^{2g} \xrightarrow{\sim} A[2]$$

and if $V \subset \mathbb{F}_2^{2g}$ is a maximal isotropic subgroup, the same property holds for $W = \bar{\alpha}(V) \subset A[2]$. If $\pi : \mathbb{C}^3 \to \mathbb{C}^3/\Lambda$ is the canonical projection, the lattice $\Lambda_W = \pi^{-1}(W)$ is associated to A/W as the following diagram shows:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \Lambda_W/\Lambda & \longrightarrow & \mathbb{C}^3/\Lambda & \longrightarrow & \mathbb{C}^3/\Lambda_W & \longrightarrow & 0 \\ & & \| & & \| & & \| & & \\ 0 & \longrightarrow & W & \longrightarrow & A & \longrightarrow & A/W & \longrightarrow & 0. \end{array}$$

We define

$$\mathrm{Trans}(W) = \{M \in \Gamma_g \mid M(\mathrm{mod}\, 2) \in \mathrm{Trans}(V)\}.$$

We introduce now the congruence subgroup

$$\Gamma_{0,g}(2) = \left\{ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \Gamma_g \mid C \equiv 0 (\mathrm{mod}\, 2) \right\}.$$

This is the transposed subgroup of the group $\Gamma_g^0(2)$, see § A. From Prop. A.1 we deduce that $\Gamma_{0,g}(2) = \mathbf{P}(\mathbb{Z}).\Gamma_g(2)$, hence

$$\mathrm{Trans}(W) = M\Gamma_{0,g}(2)$$

for any $M \in \mathrm{Trans}(W)$.

Proposition 3.1. *With the previous notation, if $\tau = \boldsymbol{\tau}(\Omega)$ then $\frac{1}{2}{}^tM.\tau$ is a Riemann matrix of the p.p.a.v. A/W for all $M \in \mathrm{Trans}(W)$.*

Proof. If $M \in \mathrm{Trans}(W)$, $W \bmod \Lambda$ is generated by the vectors

$$\frac{1}{2}\,\Omega.Me_1, \quad \ldots \quad \frac{1}{2}\,\Omega.Me_g,$$

and the matrix

$$\Omega' = \Omega.M.H, \qquad H = \begin{bmatrix} \frac{1}{2}\mathbf{1}_g & 0 \\ 0 & \mathbf{1}_g \end{bmatrix},$$

generates Λ_W. Using (3), we get

$$\boldsymbol{\tau}(\Omega') = {}^t(M.H).\tau = \frac{1}{2}{}^tM\tau.$$

By [14] (Prop. 16.8), the polarization 2λ of A reduces to a principal polarization λ' on $A' = A/W$. This last corresponds canonically to Ω' since

$$2i(\overline{\Omega'}.J^{-1}.{}^t\Omega')^{-1} = 2i(\overline{\Omega}MHJ^{-1}{}^tH{}^tM{}^t\Omega)^{-1} = 2 \cdot 2i(\overline{\Omega}J^{-1}{}^t\Omega)^{-1}. \qquad \square$$

3.4. *Theta functions*

We recall the definition of theta functions with (entire) characteristics $[\boldsymbol{\varepsilon}] = \left[\begin{smallmatrix}\varepsilon_1\\ \varepsilon_2\end{smallmatrix}\right]$ where $\varepsilon_1, \varepsilon_2 \in \mathbb{Z}^g$, following [2]. The *(classical) theta function* is

$$\vartheta \left[\begin{smallmatrix}\varepsilon_1\\ \varepsilon_2\end{smallmatrix}\right](z,\tau) = \sum_{n\in\mathbb{Z}^g} e^{i\pi(n+\varepsilon/2)\tau(n+\varepsilon/2)+2(n+\varepsilon/2)(z+\varepsilon_2/2)} \quad (\tau\in\mathbb{H}_g, z\in\mathbb{C}^g).$$

The *Thetanullwerte* are the values at $z = 0$ of these functions, and we denote

$$\vartheta \left[\begin{smallmatrix}\varepsilon_1\\ \varepsilon_2\end{smallmatrix}\right](\tau) = \vartheta \left[\begin{smallmatrix}\varepsilon_1\\ \varepsilon_2\end{smallmatrix}\right](0,\tau).$$

We now state two formulas.

Proposition 3.2 (duplication formula). *(see [16] (Cor.IIA2.1) and [10] (IV.th.2)).*
Let $\left[\begin{smallmatrix}\varepsilon_1\\ \varepsilon_2\end{smallmatrix}\right]$ *and* $\left[\begin{smallmatrix}\varepsilon_1\\ \delta\end{smallmatrix}\right]$ *be two characteristics and* $\tau\in\mathbb{H}_g$. *Then*

$$\vartheta \left[\begin{smallmatrix}\varepsilon_1\\ \varepsilon_2\end{smallmatrix}\right](\tau/2)\vartheta \left[\begin{smallmatrix}\varepsilon_1\\ \delta\end{smallmatrix}\right](\tau/2) = \sum_{\mu\in(\mathbb{Z}/2\mathbb{Z})^g} (-1)^{\mu\delta}\cdot \vartheta \left[\begin{smallmatrix}\varepsilon_1-\mu\\ \varepsilon_2-\delta\end{smallmatrix}\right](\tau)\cdot \vartheta \left[\begin{smallmatrix}\mu\\ \varepsilon_2-\delta\end{smallmatrix}\right](\tau). \quad (4)$$

The second is called *transformation formula*.
Let $M = \begin{pmatrix} A & B\\ C & D\end{pmatrix} \in \mathrm{Sp}_{2g}(\mathbb{Z})$. We let M acts on the characteristics in the following way

$$[M.\boldsymbol{\varepsilon}] = M.\left[\begin{smallmatrix}\varepsilon_1\\ \varepsilon_2\end{smallmatrix}\right] = \left[\begin{smallmatrix} D\varepsilon_1 - C\varepsilon_2 + (C\,{}^tD)_0\\ -B\varepsilon_1 + A\varepsilon_2 + (A\,{}^tB)_0\end{smallmatrix}\right]$$

where P_0 denotes the diagonal of the matrix P.

Proposition 3.3 ([10] (V.§.2)).

$$\vartheta[M.\boldsymbol{\varepsilon}](M.\tau) = \kappa(M)\cdot w^{\phi_{[\varepsilon_1,\varepsilon_2]}(M)}\cdot j(M,\tau)^{1/2}\cdot\vartheta[\boldsymbol{\varepsilon}](\tau)$$

where $\kappa(M)^2$ *is a root of* 1 *depending only on* M, $w = e^{i\pi/4}$,

$$j(M,\tau) = \det(C\tau + D)$$

and

$$\phi_{[\varepsilon_1,\varepsilon_2]}(M) = \varepsilon_1\,{}^tDB\varepsilon_1 - 2\varepsilon_1\,{}^tBC\varepsilon_2 + \varepsilon_2\,{}^tCA\varepsilon_2 - 2(D\varepsilon_1 - C\varepsilon_2)\cdot(A\,{}^tB)_0.$$

We will need a slightly modified version of the previous result.

Corollary 3.1. *For any characteristic* $\left[\begin{smallmatrix}\varepsilon_1'\\ \varepsilon_2'\end{smallmatrix}\right]$ *and for any* $M\in\mathrm{Sp}_{2g}(\mathbb{Z})$ *we have*

$$\vartheta \left[\begin{smallmatrix}\varepsilon_1'\\ \varepsilon_2'\end{smallmatrix}\right](M.\tau) = c(M,\tau)\cdot\omega^{-\phi_{[\varepsilon_1,\varepsilon_2]}(M^{-1})}\cdot\vartheta \left[\begin{smallmatrix} {}^tA(\varepsilon_1' - (C\,{}^tD)_0) + {}^tC(\varepsilon_2' - (A\,{}^tB)_0)\\ {}^tB(\varepsilon_1' - (C\,{}^tD)_0) + {}^tD(\varepsilon_2' - (A\,{}^tB)_0)\end{smallmatrix}\right](\tau) \quad (5)$$

where

$$c(M,\tau) = \kappa(M^{-1})^{-1} \cdot j(M,\tau).$$

Proof. To inverse the action on the characteristics, we let $\tau' = M^{-1}.\tau$ in the transformation formula. Note that

$$M^{-1} = \begin{pmatrix} {}^tD & -{}^tB \\ -{}^tC & {}^tA \end{pmatrix}$$

and that $[M.\boldsymbol{\varepsilon}] = [{}^tM^{-1}\binom{\varepsilon_1}{\varepsilon_2}] + \begin{bmatrix} (C{}^tD)_0 \\ (A{}^tB)_0 \end{bmatrix}$. Thus we get the action on the characteristics. For the factor $j(M,\tau)$ note that

$$\begin{aligned} j(M,\tau) &= \det(C\tau + D) = \det(CM^{-1}.\tau' + D) \\ &= \det(C({}^tD\tau' - {}^tB)(-{}^tC\tau' + {}^tA)^{-1} + D) \\ &= \det(C({}^tD\tau' - {}^tB) + D(-{}^tC\tau' + {}^tA)) \det(-{}^tC\tau' + {}^tA)^{-1} \\ &= \det(-{}^tC\tau' + {}^tA)^{-1} = j(M^{-1},\tau')^{-1} \end{aligned}$$

using Lem. 3.1. □

Corollary 3.2. *Let* $\Omega = [\Omega_1\ \Omega_2]$ *be a period matrix and* $\tau = \boldsymbol{\tau}(\Omega) = \Omega_2^{-1}\Omega_1 \in \mathbb{H}_g$. *Let* $\Omega' = \Omega {}^tM = [\Omega'_1\ \Omega'_2]$. *Then*

$$j(M, \boldsymbol{\tau}(\Omega)) = \det(\Omega_2)^{-1} \cdot \det(\Omega'_2).$$

Proof. We compute

$$\begin{aligned} \det(C\tau + D) &= \det(\tau {}^tC + {}^tD) \\ &= \det(\Omega_2)^{-1} \det(\Omega_1 {}^tC + \Omega_2 {}^tD) \\ &= \det(\Omega_2)^{-1} \cdot \det(\Omega'_2), \end{aligned}$$

the last expression coming from (3). □

3.5. *The modular function* χ_k

Recall that a characteristic $\left[\begin{smallmatrix}\varepsilon_1\\ \varepsilon_2\end{smallmatrix}\right]$ is *even* (resp. *odd*) if $\varepsilon_1.\varepsilon_2 \equiv 0 \pmod 2$ (resp. $\varepsilon_1.\varepsilon_2 \equiv 1 \pmod 2$). Let S_g (resp. U_g) be the set of even (resp. odd) characteristics $\left[\begin{smallmatrix}\varepsilon_1\\ \varepsilon_2\end{smallmatrix}\right]$ with coefficients in $\{0,1\}$. It is well known that

$$\#S_g = 2^{g-1}(2^g+1), \quad \#U_g = 2^{g-1}(2^g-1).$$

Let $\Omega = [\Omega_1\ \Omega_2] \in \mathcal{R}_g$ and $\tau = \Omega_2^{-1}\Omega_1 \in \mathbb{H}_g$ be a Riemann matrix. For $g \geq 2$, we denote $k = \#S_g/2$ and we are interested in the following expressions :

$$\chi_k(\tau) = \prod_{\boldsymbol{\varepsilon} \in S_g} \vartheta[\boldsymbol{\varepsilon}](\tau).$$

Recall that a function f is a *modular form* of weight k for the congruence subgroup $\Gamma \in \Gamma_g$ if for all $\tau \in \mathbb{H}_g$ and $M \in \Gamma$ one has

$$f(M.\tau) = j(M,\tau)^k f(\tau).$$

Using Cor.3.2, we get

Corollary 3.3. *Let f be a modular form of weight k for Γ on $\mathbb{H}_g$. For $\Omega = [\Omega_1\ \Omega_2] \in \mathcal{R}_g$, we define $\tau = \Omega_2^{-1}\Omega_1 \in \mathbb{H}_g$ a Riemann matrix and*

$$f(\Omega) := \det(\Omega_2)^{-k} \cdot f(\tau).$$

Then for all $M \in \Gamma$

$$f(\Omega.M) = f(\Omega).$$

In his beautiful paper [9], Igusa proves the following result [*loc. cit.*, Lem. 10 & 11]. Denote by Σ_{140} the thirty-fifth elementary symmetric function of the eighth power of the even Thetanullwerte.

Theorem 3.1. *For $g \geq 3$, the product $\chi_k(\tau)$ is a modular form of weight k for the group Γ_g. Moreover, If $g = 3$ and $\tau \in \mathbb{H}_3$, then:*

(1) A_τ is decomposable if $\chi_{18}(\tau) = \Sigma_{140}(\tau) = 0$.
(2) A_τ is a hyperelliptic Jacobian if $\chi_{18}(\tau) = 0$ and $\Sigma_{140}(\tau) \neq 0$.
(3) A_τ is a non hyperelliptic Jacobian if $\chi_{18}(\tau) \neq 0$.

This theorem gives answers to the two questions raised in Sec.1.1 over $\mathbb{C}$.

In the sequel, we will need the following result to prove the independence of our results from the choices we will make. The proof is the case $g = 3$ of Th. A.2.

Proposition 3.4. *The product $\tau \mapsto \chi_{18}(\frac{1}{2}\tau)$ is a modular form on $\mathbb{H}_3$ of weight 18 for $\Gamma_3^0(2)$.*

4. Comparison of analytic and algebraic discriminants

In this part, we make the link between the algebraic result Th.2.2 and Serre's assertion on the modular function χ_{18}. To do so, we first compute a quantity related easily to $T(\widetilde{A})$ in terms of the Thetanullwerte on the elliptic curves. Then, after a good choice of a symplectic matrix N (related to the subgroup W we use for the quotient), we compute $\chi_{18}(({}^tN.\tau)/2)$ in terms of the same Thetanullwerte. Thus, we express χ_{18} on the quotient A/W. Finally we compare the expressions to prove Serre's assertion.

4.1. *Expression of the algebraic discriminant*

We come back to the hypotheses of § 2.3, and specialize to the case $K \subset \mathbb{C}$. Let $A = E_1 \times E_2 \times E_3 \in \mathsf{A}$, where

$$E_i : y^2 = x(x^2 - 4b_i x - 4c_i), \qquad (b_i \in K, c_i \in K^\times) \quad (i = 1, 2, 3).$$

We choose a root ρ of $\delta(A)$, and put $m = \mathbf{Mat}(\widetilde{A})$ with $\widetilde{A} = (A, \rho)$. Let

$$\mathsf{X}(m) := (a_1\, a_2\, a_3)^4\, (c_1\, c_2\, c_3)^2\, \det m.$$

Since $\mathsf{T}(\widetilde{A}) = \det m$, $\mathsf{T}(\widetilde{A})$ is a square in K if and only if $X(m)$ is a square in K. The function X appears naturally in our problem since it is related to the function $\mathsf{D}(m)$ (Prop. 2.2) by

$$\mathsf{X}(\mathrm{Cof}\, m) = \mathsf{D}(m)^2, \tag{6}$$

and this reflects Serre's assertion according to Th.2.1. In order to determine the expression of $X(m)$ is terms of the Thetanullwerte, we use the following uniformization. The curves E_i can be written as

$$E(\omega_{1i}, \omega_{2i}) : y^2 = x(x + \frac{\pi^2}{\omega_2^2}\vartheta\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right](\tau_i)^4)(X + \frac{\pi^2}{\omega_2^2}\vartheta\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right](\tau_i)^4),$$

with $[\omega_{1i}\ \omega_{2i}] \in \mathcal{R}_1$ and $\tau_i = \frac{\omega_{1i}}{\omega_{2i}}$. We identify $\mathcal{R}_1^3$ with the set of matrices

$$\Omega = [\Omega_1\ \Omega_2] = \left[\begin{pmatrix} \omega_{11} & 0 & 0 \\ 0 & \omega_{12} & 0 \\ 0 & 0 & \omega_{13} \end{pmatrix} \begin{pmatrix} \omega_{21} & 0 & 0 \\ 0 & \omega_{22} & 0 \\ 0 & 0 & \omega_{23} \end{pmatrix} \right]$$

such that

$$\tau = \tau(\Omega) = \Omega_2^{-1}\Omega_1 = \begin{bmatrix} \tau_1 & 0 & 0 \\ 0 & \tau_2 & 0 \\ 0 & 0 & \tau_3 \end{bmatrix} \in \mathbb{H}_3.$$

We define

$$A(\Omega) := E(\omega_{11}, \omega_{21}) \times E(\omega_{12}, \omega_{22}) \times E(\omega_{13}, \omega_{23}),$$

$$\rho(\Omega) := \frac{\pi^6}{64(\det \Omega_2)^2}\, \theta^4\begin{bmatrix}111\\000\end{bmatrix}(\tau).$$

This defines an element $\widetilde{A}(\Omega) := (A(\Omega), \rho(\Omega)) \in \widetilde{\mathsf{A}}$, and a matrix $m(\Omega) := \mathbf{Mat}(\widetilde{A}(\Omega))$. For $1 \le i \le 3$, denote

$$\vartheta_{0i} = \vartheta\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right](\tau_i), \quad \vartheta_{1i} = \vartheta\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right](\tau_i), \quad \vartheta_{2i} = \vartheta\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right](\tau_i).$$

The coefficients of $A(\Omega)$ and $m(\Omega)$ are

$$a_i = -\frac{\pi^2}{4}\,\frac{\omega_{2i}^2}{(\omega_{2j}\omega_{2k})^2}\,\frac{\vartheta_{1j}^4\vartheta_{1k}^4}{\vartheta_{1i}^4}\,,$$
$$b_i = -\frac{\pi^2}{4\omega_{2i}^2}(\vartheta_{0i}^4+\vartheta_{2i}^4),$$
$$c_i = -\frac{\pi^4}{4\omega_{2i}^4}\,\vartheta_{0i}^4\,\vartheta_{2i}^4,$$

where (i,j,k) is a cyclic permutation. The determinant of $m(\Omega)$ is expressed as follows. Let

$$\mathsf{a}=\vartheta_{01}^2\vartheta_{02}^2\vartheta_{23}^2,\quad \mathsf{b}=\vartheta_{01}^2\vartheta_{22}^2\vartheta_{03}^2,\quad \mathsf{c}=\vartheta_{21}^2\vartheta_{02}^2\vartheta_{03}^2,\quad \mathsf{d}=\vartheta_{21}^2\vartheta_{22}^2\vartheta_{23}^2,$$

$$\mathsf{R}_1=(\mathsf{a}+\mathsf{b}+\mathsf{c}+\mathsf{d})(\mathsf{a}+\mathsf{b}-\mathsf{c}-\mathsf{d})(\mathsf{a}-\mathsf{b}-\mathsf{c}+\mathsf{d})(\mathsf{a}-\mathsf{b}+\mathsf{c}-\mathsf{d}),$$

Then

$$\det m(\Omega)=\frac{\pi^6}{2^4\cdot\prod_{i=1}^3(\omega_{2i}^2\cdot(\vartheta_{0i}^4-\vartheta_{2i}^4))}\cdot\mathsf{R}_1.$$

Thus we get

$$\begin{aligned}\mathsf{X}(m(\Omega)) &= \Big(\frac{\pi^{12}}{2^6\cdot\prod_{i=1}^3\omega_{2i}^4}\cdot\prod_{i=1}^3\vartheta_{0i}^4\vartheta_{2i}^4\Big)^2\cdot\Big(\frac{\pi^6}{2^6\cdot\prod_{i=1}^3\omega_{2i}^2}\cdot\prod_{i=1}^3(\vartheta_{0i}^4-\vartheta_{2i}^4)\Big)^4\\ &\quad\cdot\Big(\frac{\pi^6}{2^4\cdot\prod_{i=1}^3(\omega_{2i}^2\cdot(\vartheta_{0i}^4-\vartheta_{2i}^4))}\cdot\mathsf{R}_1\Big)\\ &= \frac{\pi^{54}}{2^{40}}\cdot\det(\Omega_2)^{-18}\cdot\Big(\prod_{i=1}^3\vartheta_{0i}^8\vartheta_{2i}^8(\vartheta_{0i}^4-\vartheta_{2i}^4)^3\Big)\cdot\mathsf{R}_1.\end{aligned}$$

4.2. *The subgroup W*

With the notation of Sec. 2.4, we can always assume the following correspondences

$$P_i\leftrightarrow\frac{\omega_{1i}}{2},\quad Q_i\leftrightarrow\frac{\omega_{2i}}{2},\quad R_i=P_i+Q_i\leftrightarrow\frac{\omega_{1i}+\omega_{2i}}{2}$$

for the points of E_i. The characteristics associated to the points of W (see § 2.4) are

$$\left[\begin{smallmatrix}000\\000\end{smallmatrix}\right],\left[\begin{smallmatrix}000\\011\end{smallmatrix}\right],\left[\begin{smallmatrix}000\\101\end{smallmatrix}\right],\left[\begin{smallmatrix}000\\110\end{smallmatrix}\right],\left[\begin{smallmatrix}111\\000\end{smallmatrix}\right],\left[\begin{smallmatrix}111\\011\end{smallmatrix}\right],\left[\begin{smallmatrix}111\\101\end{smallmatrix}\right],\left[\begin{smallmatrix}111\\110\end{smallmatrix}\right].$$

It defines a maximal isotropic subgroup V of $\mathbb{F}_2^6$. A basis of V over $\mathbb{F}_2$ is given by the three vectors

$$\alpha_1=\left[\begin{smallmatrix}000\\011\end{smallmatrix}\right],\ \alpha_2=\left[\begin{smallmatrix}000\\110\end{smallmatrix}\right],\ \alpha_3=\left[\begin{smallmatrix}111\\000\end{smallmatrix}\right].$$

The matrix

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

belongs to Γ_3 and satisfies $N.e_i \equiv \alpha_i \pmod 2$ if $1 \leq i \leq 3$, thus $N \in \mathrm{Trans}(W)$.
The set

$$\Gamma_g(1,2) = \left\{ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \Gamma_g \mid (A.{}^tB)_0 \equiv (C.{}^tD)_0 \equiv 0 (\mathrm{mod}\, 2) \right\},$$

is a subgroup of Γ_g, and κ^2 is a character of $\Gamma_g(1,2)$ [10] (p. 181).

Lemma 4.1. *The matrices N and ${}^tN$ are in $\Gamma_3(1,2)$, and $\kappa(N)^2 = \kappa({}^tN)^2 = \pm 1$.*

Proof. We have $N = LQ$, where

$$L = \begin{bmatrix} A & 0 \\ 0 & {}^tA^{-1} \end{bmatrix}, \quad \text{with } A = \begin{bmatrix} 0 & -1 & 1 \\ 0 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix},$$

and

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

One checks easily that $L, {}^tL, Q, {}^tQ$ belong to $\Gamma_3(1,2)$, hence, N and ${}^tN$ are in $\Gamma_3(1,2)$ as well. If

$$M = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} \in \mathbf{P}(\mathbb{Z}),$$

then $\kappa(M)^2 = \det D$, see [9] (Lem. 7, p. 181). Now

$$Q^2 = \begin{bmatrix} S & 0 \\ 0 & S \end{bmatrix}, \quad \text{with } S = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

From this we deduce that

$$\kappa(L)^2 = \det A = 1, \quad \kappa(Q)^4 = \kappa(Q^2)^2 = \det S = 1,$$

hence, $\kappa(N)^2 = \kappa({}^tN)^2 = \pm 1$. □

Proposition 4.1. *Let $\Omega' = \Omega N H$. Then*

$$\tau(\Omega') = \frac{1}{2}{}^tN.\tau$$

is a Riemann matrix for $A'(m)$. Moreover, the value $\chi_{18}(\Omega')$ is independent on the choice of $N \in \mathrm{Trans}(W)$.

Proof. The first assertion comes from Prop. 3.1, the second from Prop. 3.4. □

4.3. *Expression of $\chi_{18}(\Omega')$ as a discriminant*

Our main result in this section is the following

Theorem 4.1. *Let $\Omega \in \mathcal{R}_1^3$ and $A(\Omega)$ be the corresponding abelian threefold, let $m = m(\Omega) \in \mathsf{S}$ be the associated matrix, and $\Omega' \in \mathcal{R}_3$ be a Riemann matrix of $A(\Omega)/W$. Then*

$$\left(\frac{\pi}{2}\right)^{54} \cdot \chi_{18}(\Omega') = \mathsf{X}(m).$$

Proof. The strategy is the following. Let N be the matrix defined in § 4.2, and define $\tau' = {}^tN.\tau = 2\tau(\Omega')$.

(1) Pair the Thetanullwerte in $\tau'/2$ such that one can apply the duplication formula (4). We then obtain expressions in terms of Thetanullwerte in τ'. Such a pairing is not unique and one makes here a choice which allows an easy comparison of the final formulas.
(2) For each of the Thetanullwerte in τ', apply the transformation formula (5) to obtain an expression in τ.
(3) Finally, since $\tau = \mathrm{diag}(\tau_1, \tau_2, \tau_3)$, we get

$$\vartheta \begin{bmatrix} a_1 b_1 c_1 \\ a_2 b_2 c_2 \end{bmatrix} (\tau) = \prod_{i=1}^{3} \vartheta \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} (\tau_i).$$

Let

$$c(N) = \kappa({}^tN^{-1})^{-2} \det(\Omega_2)^{-1} \det(\Omega_2') = \pm \det(\Omega_2)^{-1} \det(\Omega_2')$$

by Lem. 4.1.
Applying steps (1) to (3) with the software MAGMA (see `http://iml.univ-mrs.fr/~ritzenth/programme/check2.m`), we get the following 18 identities, where we write

$$\vartheta\left[\begin{smallmatrix}000\\000\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}000\\001\end{smallmatrix}\right] = \vartheta\left[\begin{smallmatrix}000\\000\end{smallmatrix}\right](\tau'/2)\vartheta\left[\begin{smallmatrix}000\\001\end{smallmatrix}\right](\tau'/2), \quad c = c(N).$$

We make the pairing in such a way that the expressions of $\vartheta\left[\begin{smallmatrix}000\\ \varepsilon_2\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}000\\ \delta\end{smallmatrix}\right]$ do not contain ϑ_{1i} terms. The first four are, with the preceding notation,

$$\begin{aligned}
\vartheta\left[\begin{smallmatrix}000\\000\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}000\\001\end{smallmatrix}\right] &= c\,(\mathsf{a}+\mathsf{b}+\mathsf{c}+\mathsf{d})\\
\vartheta\left[\begin{smallmatrix}000\\010\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}000\\011\end{smallmatrix}\right] &= c\,(\mathsf{a}+\mathsf{b}-\mathsf{c}-\mathsf{d})\\
\vartheta\left[\begin{smallmatrix}000\\100\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}000\\101\end{smallmatrix}\right] &= -c\,(\mathsf{a}-\mathsf{b}-\mathsf{c}+\mathsf{d})\\
\vartheta\left[\begin{smallmatrix}000\\110\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}000\\111\end{smallmatrix}\right] &= -c\,(\mathsf{a}-\mathsf{b}+\mathsf{c}-\mathsf{d})
\end{aligned}$$

and the remaining 14 are

$$\begin{aligned}
\vartheta\left[\begin{smallmatrix}010\\000\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}010\\001\end{smallmatrix}\right] &= 2c\,(\vartheta_{01}\vartheta_{21}\vartheta_{02}\vartheta_{22}\vartheta_{03}^2+\vartheta_{01}\vartheta_{21}\vartheta_{02}\vartheta_{22}\vartheta_{23}^2)\\
\vartheta\left[\begin{smallmatrix}100\\000\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}100\\001\end{smallmatrix}\right] &= 2c\,(\vartheta_{01}^2\vartheta_{02}\vartheta_{22}\vartheta_{03}\vartheta_{23}+\vartheta_{21}^2\vartheta_{02}\vartheta_{22}\vartheta_{03}\vartheta_{23})\\
\vartheta\left[\begin{smallmatrix}110\\000\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}110\\001\end{smallmatrix}\right] &= 2c\,(\vartheta_{01}^2\vartheta_{21}\vartheta_{02}^2\vartheta_{03}\vartheta_{23}+\vartheta_{01}\vartheta_{21}\vartheta_{22}^2\vartheta_{03}\vartheta_{23})\\
\vartheta\left[\begin{smallmatrix}010\\100\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}010\\101\end{smallmatrix}\right] &= 2c\,(\vartheta_{01}^2\vartheta_{21}\vartheta_{02}\vartheta_{22}\vartheta_{03}^2-\vartheta_{01}^2\vartheta_{21}\vartheta_{02}\vartheta_{22}\vartheta_{23}^2)\\
\vartheta\left[\begin{smallmatrix}100\\010\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}100\\011\end{smallmatrix}\right] &= 2c\,(\vartheta_{01}^2\vartheta_{02}\vartheta_{22}\vartheta_{03}\vartheta_{23}-\vartheta_{21}^2\vartheta_{02}\vartheta_{22}\vartheta_{03}\vartheta_{23})\\
\vartheta\left[\begin{smallmatrix}110\\110\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}110\\111\end{smallmatrix}\right] &= -2c\,(\vartheta_{01}\vartheta_{21}\vartheta_{02}^2\vartheta_{03}\vartheta_{23}-\vartheta_{01}\vartheta_{21}\vartheta_{22}^2\vartheta_{03}\vartheta_{23})\\
\vartheta\left[\begin{smallmatrix}001\\000\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}001\\010\end{smallmatrix}\right] &= 2c\,(\vartheta_{01}\vartheta_{11}\vartheta_{02}\vartheta_{12}\vartheta_{03}\vartheta_{13})\\
\vartheta\left[\begin{smallmatrix}001\\100\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}001\\110\end{smallmatrix}\right] &= 2c\,(\vartheta_{01}\vartheta_{11}\vartheta_{02}\vartheta_{12}\vartheta_{03}\vartheta_{13})\\
\vartheta\left[\begin{smallmatrix}011\\110\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}011\\100\end{smallmatrix}\right] &= 2c\,(\vartheta_{11}\vartheta_{21}\vartheta_{12}\vartheta_{22}\vartheta_{03}\vartheta_{13})\\
\vartheta\left[\begin{smallmatrix}101\\000\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}101\\010\end{smallmatrix}\right] &= 2c\,(\vartheta_{01}\vartheta_{11}\vartheta_{12}\vartheta_{22}\vartheta_{13}\vartheta_{23})\\
\vartheta\left[\begin{smallmatrix}111\\000\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}111\\110\end{smallmatrix}\right] &= 2c\,(\vartheta_{11}\vartheta_{21}\vartheta_{02}\vartheta_{12}\vartheta_{13}\vartheta_{23})\\
\vartheta\left[\begin{smallmatrix}011\\011\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}011\\111\end{smallmatrix}\right] &= -2c\,(\vartheta_{11}\vartheta_{21}\vartheta_{12}\vartheta_{22}\vartheta_{03}\vartheta_{13})\\
\vartheta\left[\begin{smallmatrix}111\\011\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}111\\101\end{smallmatrix}\right] &= -2c\,(\vartheta_{11}\vartheta_{21}\vartheta_{02}\vartheta_{12}\vartheta_{13}\vartheta_{23})\\
\vartheta\left[\begin{smallmatrix}101\\101\end{smallmatrix}\right]\vartheta\left[\begin{smallmatrix}101\\111\end{smallmatrix}\right] &= -2c\,(\vartheta_{01}\vartheta_{11}\vartheta_{12}\vartheta_{22}\vartheta_{13}\vartheta_{23})
\end{aligned}$$

Denote by R_1' the product of the first four lines.
Obviously $\mathsf{R}_1' = c(N)^4\mathsf{R}_1$. Calling R_2' the product of the last fourteen lines, we get

$$\mathsf{R}_2' = 2^{14}\cdot c(N)^{14}\cdot\left(\prod_{i=1}^{3}\vartheta_{0i}^8\vartheta_{2i}^8(\vartheta_{0i}^4-\vartheta_{2i}^4)^3\right).$$

So

$$\chi_{18}(\tau'/2) = \mathsf{R}'_1\mathsf{R}'_2 = 2^{14}\cdot c(N)^{18}\cdot\left(\frac{2^{40}}{\pi^{54}}\cdot\det(\Omega_2)^{18}\right)\cdot\mathsf{X}(m)$$
$$= \left(\frac{2}{\pi}\right)^{54}\cdot\det(\Omega'_2)^{18}\cdot\mathsf{X}(m),$$

which is the expected result. □

Since $\mathsf{X}(m)$ is equal to $\mathsf{T}(\widetilde{A})$ up to a square, Th.2.2 and Th.4.1 show Serre's assertion.

Corollary 4.1. *Let $K\subset\mathbb{C}$ and $m\in\mathsf{S}^\times$ with coefficients in K. Let $A'(m)$ be the associated abelian threefold and Ω' be one of its period matrix. Then*

$$\left(\frac{\pi}{2}\right)^{54}\cdot\chi_{18}(\Omega')\in K^{\times 2}$$

if and only if $A'(m)$ is the Jacobian of a non hyperelliptic genus 3 curve.

In other words, Serre's assertion is true for our three dimensional family A of abelian threefolds.

Corollary 4.2. *If $m\in\mathsf{S}^\times$ and Ω_m is a period matrix associated to the non hyperelliptic genus 3 curve X_m with Ciani form Q_m then*

$$\chi_{18}(\Omega_m) = \left(\frac{1}{2\pi}\right)^{54}\cdot\mathrm{Disc}(Q_m)^2.$$

Proof. Using Th.2.1 and (6) we get

$$\left(\frac{\pi}{2}\right)^{54}\cdot\chi_{18}(\Omega_m) = \mathsf{X}(\mathrm{Cof}\,m)$$
$$= \mathsf{D}(m)^2 = \left(2^{-54}\cdot\mathrm{Disc}(Q_m)\right)^2.$$ □

Remark 4.1. When $m\in\mathsf{S}\setminus\mathsf{S}^\times$, the abelian variety $A'(m)$ comes from a hyperelliptic curve and the above formula degenerates. However in [13] and [5] we find a beautiful formula for the hyperelliptic case in every genus. Let

$$C: Y_2 = a_{2g+2}X^{2g+2}+\ldots+a_0 = a_{2g+2}(X-\alpha_1)\cdots(X-\alpha_{2g+2})$$

and

$$\Delta_{\mathrm{alg}}(C) = a_{2g+2}^{4g+2}\prod_{j<k}(\alpha_j-\alpha_k)^2.$$

They define also a modular form on $\mathbb{H}_g$

$$\delta(\tau) = \prod_{\boldsymbol{\varepsilon} \in T} \vartheta[\boldsymbol{\varepsilon}](\tau)^8$$

where T is a certain subset of even theta characteristic. One has

$$\Delta_{\mathrm{alg}}(C)^{2n} = (2\pi)^{4rg} \det(\Omega_1)^{-4r} \delta(\tau)^2$$

where

$$r = \binom{2g+2}{g+1}, \quad n = \binom{2g}{g+1},$$

and $\tau = \tau(\Omega)$ for a certain period matrix $\Omega = [\Omega_1, \Omega_2]$ of $\mathrm{Jac}(C)$.

Remark 4.2. Denote by V_3^4 the 15-dimensional affine open set of ternary quartics. Felix Klein proved in 1889 that there is a map

$$\Omega : V_3^4 \longrightarrow \mathcal{R}_g$$

such that if $\Omega(Q) = [\Omega_1\ \Omega_2]$ and $X : Q = 0$, then $\mathrm{Jac}\, X = A_{\Omega(Q)}$ and

$$\chi_{18}(\Omega) = c\, \mathrm{Disc}(Q)^2,$$

with some unspecified constant $c \in \mathbb{C}$. We prove here that $c = (1/2\pi)^{54}$. Using this precise version of Klein's formula, it is almost obvious to extend our theorem to the general case. However, we did not include it, for we think that a good presentation should include a modern proof of Klein's result. We plan to do this in a forthcoming article.

Appendix A. Modularity of χ_k

Let

$$\Gamma_g(2) = \{M \in \Gamma_g \mid M \equiv \mathbf{1}_{2g} (\mathrm{mod}\, 2)\}$$

an recall that the sequence

$$1 \to \Gamma_g(2) \to \Gamma_g \to \mathrm{Sp}_{2g}(\mathbb{F}_2) \to 1$$

is exact. We introduce the congruence subgroup

$$\Gamma_g^0(2) = \left\{ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathrm{Sp}_g(\mathbb{Z}) \mid B \equiv 0 (\mathrm{mod}\, 2) \right\}.$$

We need a set of generators for this subgroup. For any integer $n \geq 1$, define

$$M(n) = \mathbf{M}(\mathbb{Z}) \cap \Gamma_g(n), \quad U(n) = \mathbf{U}(\mathbb{Z}) \cap \Gamma_g(n), \quad V(n) = \mathbf{V}(\mathbb{Z}) \cap \Gamma_g(n).$$

with

$$\Gamma_g(n) = \{M \in \Gamma_g \mid M \equiv \mathbf{1}_{2g} (\mathrm{mod}\, n)\}.$$

Theorem A.1. *The subgroups $M(1)$, $U(2)$ and $V(1)$ generate $\Gamma_g^0(2)$, and $\Gamma_g^0(2) = \Gamma_g(2).\mathbf{M}(\mathbb{Z}).\mathbf{V}(\mathbb{Z})$.*

Proof. First, the subgroups $M(2)$, $U(2)$ and $V(2)$ generate $\Gamma_g(2)$, see [10] (p. 179). Let

$$\Gamma_g^1(2) = \left\{ \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in \mathrm{Sp}_g(\mathbb{Z}) \mid A \equiv D \equiv 1 (\mathrm{mod}\, 2) \text{ and } B \equiv 0 (\mathrm{mod}\, 2) \right\}.$$

There is the following diagram, where the vertical arrow is the transpose of the reduction modulo 2:

$$\begin{array}{ccccccc} \Gamma_g(2) & \subset & \Gamma_g^1(2) & \subset & \Gamma_g^0(2) & \subset & \Gamma_g(1) \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ 0 & \subset & \mathbf{U}(\mathbb{F}_2) & \subset & \mathbf{P}(\mathbb{F}_2) & \subset & \mathrm{Sp}_g(\mathbb{F}_2) \end{array}$$

Then, if $M \in \Gamma_g^1(2)$ is written as usual

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathbf{1}_g & 0 \\ C & \mathbf{1}_g \end{bmatrix} = \begin{bmatrix} A + BC & B \\ C + DC & D \end{bmatrix} \in \Gamma_g(2),$$

and if $M \in \Gamma_g^0(2)$, then

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & {}^tA \end{bmatrix} = \begin{bmatrix} \mathbf{1}_g & B{}^tA \\ CA^{-1} & D{}^tA \end{bmatrix} \in \Gamma_g^1(2). \qquad \square$$

Theorem A.2. *Assume $g \geq 3$. The function $\chi_k(\frac{1}{2}\tau)$ is a modular form on $\mathbb{H}^g$ of weight k for $\Gamma_g^0(2)$.*

Proof. J.-I. Igusa proved [9] (p. 850) that $\chi_k(\tau)$ is a modular form of weight k for $\Gamma_g(1)$ if $g \geq 3$. Let

$$H = \begin{bmatrix} \frac{1}{2}\mathbf{1}_g & 0 \\ 0 & \mathbf{1}_g \end{bmatrix}, \quad H.\tau = \tfrac{1}{2}\tau.$$

Let $f(\tau) = \chi_k(\frac{1}{2}\tau) = \chi_k(H.\tau)$. It is sufficient to check that

$$f(M.\tau) = j(M, \tau)^k f(\tau)$$

if M belongs to one of the generating subgroups described in Theorem A.1. First, if $M \in M(1)$, then $H.M = M.H$, hence,

$$f(M.\tau) = \chi_k(H.M.\tau) = \chi_k(M.H.\tau) = j(M, H.\tau)^k \chi_k(H.\tau) = j(M, \tau)^k f(\tau),$$

since $j(M,\tau) = \pm 1$ for every $M \in \mathbf{M}(\mathbb{Z})$ does not depend on $\tau \in \mathbb{H}_g$. Now, if $U \in U(2)$, then

$$U = U'^2 = \begin{bmatrix} \mathbf{1}_3 & 2B \\ 0 & \mathbf{1}_3 \end{bmatrix}, \quad \text{where } U' = \begin{bmatrix} \mathbf{1}_3 & B \\ 0 & \mathbf{1}_3 \end{bmatrix} \in \mathbf{U}(\mathbb{Z}),$$

and $H.U = H.U'^2 = U'.H$. This implies

$$f(U.\tau) = \chi_k(H.U'^2.\tau) = \chi_k(U'.H.\tau) = j(U', H.\tau)^k \chi_k(H.\tau) = j(U,\tau)^k f(\tau),$$

since $j(U^n,\tau) = 1$ for every $U \in \mathbf{U}(\mathbb{Z})$. If $V \in V(1)$, then $H.V = V^2.H$. Hence

$$\begin{aligned} f(V.\tau) = \chi_k(H.V.\tau) = \chi_k(V^2.H.\tau) &= j(V^2, H.\tau)^k \, \chi_k(H.\tau) \\ &= j(V,\tau)^k \chi_k(H.\tau) = j(V,\tau)^k f(\tau), \end{aligned}$$

since $j(V^2,\tau) = j(V, 2\tau)$ for every $V \in V(1)$. □

References

1. Bars, Francesc, Automorphism groups of genus 3 curves, in *Corbes de gènere 3*, Notes del Seminari de Teoria de Nombres de Barcelona **14**, 2006, 27-62.
2. C. Birkenhake, , H. Lange, *Complex abelian varieties* Second edition. Grundlehren der Mathematischen Wissenschaften, **302** Springer-Verlag, Berlin, 2004.
3. E. Ciani, I Varii Tipi Possibili di Quartiche Piane più Volte Omologico-Armoniche, Rend. Circ. Mat. Palermo **13** (1899), 347-373.
4. W.L. Edge, The discriminant of a certain ternary quartic, Proc. Roy. Soc. Edinburgh, Sect. A. **62** (1948), 268-272.
5. J. Guàrdia, Jacobian nullwerte and algebraic equations. J. Algebra **253** (2002), 112-132.
6. I.M. Gel'fand, M.M. Kapranov, A.V. Zelevinsky *Discriminants, resultants, and multidimensional determinants*, Birkhäuser, Boston, (1994).
7. W.L. Hoyt, On products and algebraic families of Jacobian varieties. Ann. of Math. **77**, (1963), 415-423.
8. E. Howe, F. Leprévost, B. Poonen, Large torsion subgroups of split Jacobians of curves of genus two or three. Forum Math. **12**, (2000), 315-364.
9. J.-I. Igusa, Modular forms and projective invariants, Amer. J. Math, **89**, (1967), 817-855.
10. J.-I. Igusa, *Theta functions*, Grundlehren der mathematischen Wissenschaften, **194**, Springer Verlag, (1972).
11. F. Klein, Zur Theorie der Abelschen Funktionen. Math. Annalen, **36** (1889-90) = Gesammelte mathematische Abhandlungen, **XCVII**, 388-474.
12. K. Lauter, Geometric methods for improving the upper bounds on the number of rational points on algebraic curves over finite fields. With an appendix by Jean-Pierre Serre. J. Algebraic Geom. **10**, (2001), 19-36.

13. P. Lockhart, On the discriminant of a hyperelliptic curve. Trans. Amer. Math. Soc. **342**, (1994), 729-752.
14. J.S. Milne, Abelian varieties, in *Arithmetic geometry* (Storrs, Conn., 1984), 103-150, Springer, New York, (1986).
15. F. Oort, K. Ueno, Principally polarized abelian varieties of dimension two or three are Jacobian varieties. J. Fac. Sci. Univ. Tokyo Sect. IA Math., **20**, (1973), 377-381.
16. H.E. Rauch, H.M. Farkas, *Theta functions with applications to Riemann surfaces* The Williams & Wilkins Co., Baltimore, Md., (1974).
17. J.-P. Serre, Letter to Jaap Top, February 28, private communication, (2003).
18. A. Weil, Zum Beweis des Torellischen Satzes. Nachr. Akad. Wiss. Göttingen. Math.-Phys. Kl. IIa. (1957), 33-53 ; = Œuvres Sc., vol. II, [1957a], 307-327, Springer, New York, (1979).

Pseudorandom Points on Elliptic Curves over Finite Fields

Igor E. Shparlinski

Department of Computing, Macquarie University
Sydney, NSW 2109, Australia
E-mail : igor@ics.mq.edu.au

We give a brief survey of several recently suggested constructions of generating sequences of pseudorandom points on elliptic curves. Such constructions are of interest for both classical and elliptic curve cryptography and are also of intrinsic mathematical interest. We present an account of various results obtained for such sequences and outline several open questions (of different level of difficulty) and directions for further research.

1. Introduction

1.1. *Motivation*

There is a vast literature devoted to generating pseudorandom numbers using arithmetic of finite fields and residue rings, see [46–49,54,60] and references therein. Here we consider a reasonably new source of pseudorandom numbers which is based on using the group structure of elliptic curves over finite fields. We remark that the idea of using elliptic curves as a source of randomness is not new and dates back to [32]. However here we consider different constructions which also have a different emphasis. Besides intrinsic mathematical interest, there are two main reasons motivating such constructions:

- Many standard pseudorandom number generators based on finite fields and residue rings have proved to be insecure or at least requiring great care in their use, see [4–6,10,11,15,22–24,31,35,37] (however, the recent result of [27] shows that pseudorandom number generators on elliptic curves are not immune to this kind of attack and should also be used with great care).
- Many cryptographic protocols explicitly require to generate random points on a given elliptic curve, see [1,8].

It should be noted, that usually there is nothing too exciting in the constructions themselves, which in most of the cases are merely straightforward analogues of well-known constructions over finite fields (however, see Section 6). Typically, the most interesting part lies in the analysis and proving various results about their distribution and other properties.

Here we give a survey of several recently proposed construction together with a representative sample of results which have been obtained. We also propose several open questions and directions for further research. Some of these questions can probably be solved by a rather straight forward extension of already known results, but some may require principally new approaches.

1.2. *Notation*

For a prime power q and a positive integer m, we use $\mathbb{F}_q$ to denote a finite field of q elements and $\mathbb{F}_q$ and $\mathbb{Z}/m\mathbb{Z}$ to denote the residue ring modulo m.

The implied constants in the symbols 'O', '$\ll$' and '$\gg$' may occasionally, where obvious, depend on the real numbers ε and δ, but are absolute otherwise. We recall that the notations $U = O(V)$, $U \ll V$, and $V \gg U$ are all equivalent to the assertion that the inequality $|U| \le cV$ holds for some constant $c > 0$. We also use the symbol $o(1)$ to denote a function which tends to 0 and depends only on ε and δ.

For a real $x > 0$ we use $\log x$ to denote the binary logarithm of x.

1.3. *Basic Facts on Elliptic Curves*

Let $\mathbf{E}$ be an elliptic curve defined over $\mathbb{F}_q$, the finite field of q elements given by an *affine Weierstrass equation*, which for $\gcd(q, 6) = 1$ takes form

$$Y^2 = X^3 + aX + b, \tag{1}$$

for some $a, b \in \mathbb{F}_q$ with $4a^3 + 27b^2 \neq 0$.

We recall that the set $\mathbf{E}(\mathbb{F}_q)$ of $\mathbb{F}_q$-rational points forms an abelian group whose satisfies the Hasse–Weil inequality

$$|\#\mathbf{E}(\mathbb{F}_q) - q - 1| \le 2q^{1/2}, \tag{2}$$

with the *point at infinity* $\mathcal{O}$ as the neutral element of this group (which does not have affine coordinates), see [1,7,57] for this and other general properties of elliptic curves.

We use $\oplus$ to denote the group operation. For example, $Q \oplus \mathcal{O} = Q$ for any point $Q \in \mathbf{E}(\mathbb{F}_q)$.

It is well-known that the group of $\mathbb{F}_q$-rational points $\mathbf{E}(\mathbb{F}_q)$ is of the form

$$\mathbf{E}(\mathbb{F}_q) \cong \mathbb{Z}/L\mathbb{Z} \times \mathbb{Z}/M\mathbb{Z},$$

where the integers L and M are uniquely determined with $M \mid L$. We also recall (although we do not use this fact) that the *Weil pairing* implies that $M \mid q-1$.

For a point $Q \in \mathbf{E}(\mathbb{F}_q)$ we use $x(Q)$ and $y(Q)$ to denote its components, that is, $Q = (x(Q), y(Q))$.

1.4. *Some Important Characteristics of Pseudorandom Number Generators*

There are several important criteria which a good sequence of pseudorandom numbers should satisfy. However here we concentrate only on its *period*, *discrepancy* and *linear complexity*.

Certainly any decent sequence of pseudorandom numbers should have a large period. In fact, it has turned out that most of the known bounds on the discrepancy and linear complexity are nontrivial only if the corresponding sequence is of sufficiently large period.

For a sequence of N points in the s-dimensional unit cube $[0,1)^s$

$$\Gamma = \{(\gamma_{1,n}, \ldots \gamma_{s,n}) \in [0,1)^s, \quad 1 \le n \le N\} \tag{3}$$

its discrepancy $\Delta(\Gamma)$ is defined as

$$\Delta(\Gamma) = \sup_{B \subseteq [0,1)^s} \left| \frac{T_\Gamma(B)}{N} - |B| \right|,$$

where $T_\Gamma(B)$ is the number of points of Γ inside the box

$$B = [\alpha_1, \beta_1) \times \ldots \times [\alpha_s, \beta_s) \subseteq [0,1)^s$$

and the supremum is taken over all such boxes, see [17,36]. It is easy to see that the discrepancy is a quantitative measure of uniformity of distribution of sequences, and thus good pseudorandom sequences should (after an appropriate scaling) have a small discrepancy.

Given a sequence $s_0, \ldots, s_{N-1}$ of N elements of a ring $\mathcal{R}$ we call its linear complexity the smallest value of ℓ for which

$$s_{n+\ell} = c_{\ell-1} s_{n+\ell-1} + \ldots + c_0 s_n, \qquad n = 0, \ldots, N-\ell-1,$$

with some coefficients $c_0, \ldots, c_{\ell-1} \in \mathcal{R}$. We use the convention that the linear complexity is 0 if $s_0 = s_1 = \ldots = s_{N-1} = 0$ and N if $s_0 = s_1 = \ldots =$

$s_{N-2} = 0$ but $s_{N-1} \neq 0$. If s_n is an infinite sequence which is periodic with period T, then the linear complexity of the first N elements stabilizes at $N = 2T$.

Linear complexity is an important cryptographic characteristic of sequences and provides information on the predictability and thus suitability for cryptography. Linear complexity of various sequences of cryptographic interest has been studied in a vast number of works, see [16,54] and references therein.

1.5. *Main Tools*

Typically the bounds on the discrepancy of a sequence are derived from bounds of exponential sums with elements of this sequence. The relation is made explicit in the celebrated *Koksma–Szüsz inequality*, see Theorem 1.21 of [17], which we present it in the following form.

Lemma 1.1. *There exists an absolute constant $C > 0$ such that, for any integer $L > 1$ and any sequence Γ of N points* (3) *the discrepancy $\Delta(\Gamma)$ satisfies the following bound:*

$$\Delta(\Gamma) < C^s \left(\frac{1}{L} + \frac{1}{N} \sum_{\substack{0 \le a_1, \ldots a_s \le L \\ a_1 + \ldots + a_s > 0}} \prod_{j=1}^{s} \frac{1}{a_j + 1} \left| \sum_{n=1}^{N} \exp\left(2\pi i \sum_{j=1}^{s} a_j \gamma_{j,n} \right) \right| \right).$$

For estimate the corresponding exponential sums for various sequences of pseudorandom numbers, the following bound from [34] has been used.

Lemma 1.2. *The bound*

$$\sum_{Q \in \mathcal{H}} \exp\left(2\pi i f(Q)/p\right) = O(p^{1/2})$$

holds for any subgroup $\mathcal{H} \in \mathbf{E}(\mathbb{F}_p)$ and any rational function $f(X, Y) \in \mathbb{F}_q(X, Y)$ of degree d which is not constant on $\mathbf{E}$.

In particular, choosing $f(X, Y) = X$ or $f(X, Y) = y$ in Lemma 1.2 we obtain nontrivial bounds with exponential sums with $x(Q)$ and $y(Q)$, respectively.

2. Constructions

2.1. *General Conventions*

As we have mentioned, the construction present here are direct analogues of the corresponding constructions over finite fields and residue rings.

We always assume that the elliptic curve $\mathbf{E}$ is defined over a prime field $\mathbb{F}_p$ which is represented by the elements of the set $\{0, 1, \ldots, p-1\}$. For example, if $Q \in \mathbf{E}(\mathbb{F}_p)$ then we treat $x(Q)/p$ as a rational number in a unit interval $[0, 1]$.

The more general case of arbitrary finite fields $\mathbb{F}_q$ can be considered without any substantial difficulties (however some of the results depends on how the field is given).

2.2. *Linear Congruential Generator on Elliptic Curves, EC-LCG*

For a given point $G \in \mathbf{E}(\mathbb{F}_p)$, the **EC-LCG** is defined as the sequence:

$$U_n = G \oplus U_{n-1} = nG \oplus U_0, \qquad n = 1, 2, \ldots, \tag{4}$$

where $U_0 \in \mathbf{E}(\mathbb{F}_q)$ is the "initial value".

The **EC-LCG** generator has been suggested in [28] and than studied in a number of papers [3,19,25,26,29].

2.3. *Power Generator on Elliptic Curves, EC-PG*

For a given point $G \in \mathbf{E}(\mathbb{F}_p)$ and an integer $e \geq 2$, the **EC-LCG** is defined as the sequence:

$$W_n = eW_{n-1} = e^n G, \qquad n = 1, 2, \ldots, \tag{5}$$

where $W_0 \in \mathbf{E}(\mathbb{F}_q)$ is the "initial value".

The **EC-PG** generator has been introduced and studied in [40], see also [20].

2.4. *Naor-Reingold Generator on Elliptic Curves, EC-NRG*

For a given point $G \in \mathbf{E}(\mathbb{F}_p)$ of order t and a k-dimensional an integer vector $\mathbf{a} = (a_1, \ldots a_k) \in (\mathbb{Z}/t\mathbb{Z})^k$, the **EC-NRG** is defined as the sequence:

$$F_{\mathbf{a}}(n) = a_1^{\nu_1} \ldots a_k^{\nu_k} G, \qquad n = 1, 2, \ldots, \tag{6}$$

where $n = \nu_1 \ldots \nu_k$ is the bit representation of n, $0 \leq n \leq 2^k - 1$.

The **EC-NRG** generator has been introduced and studied in [53,56] and is a full analogue of the original construction of [45] described over a finite field.

Since this construction is more complicated than the other two, we give a numerical example. Let $G \in \mathbf{E}(\mathbb{F}_p)$ be of order $t = 19$, $k = 5$ and $\mathbf{a} =$

$(3, 2, 5, 3, 4)$. Then,

$$\begin{aligned}
F_{\mathbf{a}}(0) &= 3^0 2^0 5^0 3^0 4^0 G = G,\\
F_{\mathbf{a}}(1) &= 3^0 2^0 5^0 3^0 4^1 G = 4G,\\
F_{\mathbf{a}}(2) &= 3^0 2^0 5^0 3^1 4^0 G = 3G,\\
F_{\mathbf{a}}(3) &= 3^0 2^0 5^0 3^1 4^1 G = 12G,\\
&\dots \quad \dots\\
F_{\mathbf{a}}(11) &= 3^0 2^1 5^0 3^1 4^1 G = 24G = 5G,\\
&\dots \quad \dots\\
F_{\mathbf{a}}(31) &= 3^1 2^1 5^1 3^1 4^1 G = 360G = 18G.
\end{aligned}$$

Note that in the above computation we have used that $19G = \mathcal{O}$, the point at infinity.

2.5. *Subset-Sum Generator on Elliptic Curves, EC-SSG*

Let $u(n)$ be a linear recurrence sequence of elements of $\mathbb{F}_2$ of order k, that is,

$$u(n+k)+c_{k-1}u(n+k-1)+\ldots+c_1u(n+1)+c_0u(n)=0, \qquad n=1,2,\ldots,$$

for some $c_0, \ldots, c_{k-1} \in \mathbb{F}_2$, $c_0 \neq 0$, see [43, Chapter 8].

Following [18], given k points $G_1, \ldots, G_k \in \mathbf{E}(\mathbb{F}_p)$, we define the **EC-SSG** as the sequence:

$$S_n = \sum_{j=1}^{k} u(n+j-1)G_j, \qquad n = 1, 2, \ldots, \tag{7}$$

see also [23]. This construction is an elliptic curve analogue of the subset sum generator over finite fields and residue rings, which has been introduced in [52] and studied in [14,23,50,51].

2.6. *Some Other Constructions*

We note that after [32], there have been several other suggestions and approaches to extracting pseudorandomness from elliptic curves, see [12,13, 21,30]. However, these methods and results have a slightly different focus and we do not discuss them in this paper.

3. Periodic Structure

3.1. *Period of EC-LCG*

It is clear that the period T of the sequence (4) is equal to the order $T = t$ of G. This naturally leads to a question how often this order is large.

Denote by $\mathcal{T}_p$ the set of all triples (a, b, G), where $a, b \in \mathbb{F}_p$ are such that $4a^3 + 27b^2 \neq 0$ and G is a point on the corresponding curve $\mathbf{E}_{a,b}(\mathbb{F}_p)$, given by (1). From (2) we see that

$$\#\mathcal{T}_p = \left(p^2 + O(p)\right)\left(p + O(p^{1/2}\right) \sim p^3.$$

The following result, showing that typically the period of (4) is large has been established in [55].

Theorem 3.1. *For any prime $p \geq 5$, $\delta > 0$ and $\varepsilon > 0$ the number of triples $(a, b, G) \in \mathcal{T}_p$ such that the period T of the sequence* (4) *satisfies the inequality*

$$T < p^{1-\delta}$$

is at most $O\left(\#\mathcal{T}_p p^{-2\delta/3+\varepsilon}\right)$.

3.2. *Period of EC-PG*

The period T of the sequence (4) is equal to the multiplicative order of e modulo the order t of G (provided $\gcd(e, t) = 1$ and thus is a little harder to control.

Given an arbitrary $X \geq 2$, we denote by $\mathcal{S}(X)$ the set of all quintuples of the form (p, a, b, e, G), where $p \in [X/2, X]$ is prime, $a, b \in \mathbb{F}_p$ are such that $4a^3 + 27b^2 \neq 0$, G is a point of order t on the corresponding curve $\mathbf{E}_{a,b}(\mathbb{F}_p)$, given by (1) and $e \in [1, \#\mathbf{E}_{a,b}(\mathbb{F}_p) - 1]$ is an integer with $\gcd(e, t) = 1$. We see that

$$\#\mathcal{S}(X) = \sum_{X/2 \leq p \leq X} \sum_{(a,b,G) \in \mathcal{T}_p} \frac{\varphi(t_p(a, b, G))}{t_p(a, b, G)} \#\mathbf{E}_{a,b}(\mathbb{F}_p).$$

It has been shown in [55] that

$$\frac{X^5}{\log X} \gg \#\mathcal{S}(X) \gg \frac{X^5}{\log X \log\log X}.$$

The following result from [55] shows that "on average" over all parameters from $\mathcal{S}(X)$ the period of (5) is still large.

Theorem 3.2. *For any sufficiently large X, any $\varepsilon > 0$ and $\Delta = \varepsilon \log X$, the number of quintuples $(p, a, b, e, G) \in \mathcal{S}(X)$ such that the period T of the sequence* (4) *satisfies the inequality*

$$T < p \exp(-\Delta)$$

is at most $O\left(\#\mathcal{S}(X) \exp\left(-0.45 \left(\Delta \log \Delta\right)^{1/3}\right)\right)$.

3.3. *Period of EC-NRG*

The period T of the sequence (6) has never been studied in detail. Certainly, any lower bound on the linear complexity of this sequence (for example, see Theorem 5.3 below) implies the same lower bound on the period. However, there is little doubt that one can get better estimates. In fact there is no reason to expect that for almost all vectors $\mathbf{a} = (\mathbb{Z}/t\mathbb{Z})^k$ there is any periodicity at all. It would be interesting to clarify this issue at least in the most important case when $k \sim \log p$ (and also G is of order $t = p^{1+o(1)}$).

Question 3.1. Prove that for any $\delta > 0$, for almost all vectors $\mathbf{a} = (\mathbb{Z}/t\mathbb{Z})^k$ of dimension $k \sim \log p$ and almost all points $G \in \mathbf{E}(\mathbb{F}_p)$, and any positive integer $T < 2^{k(1-\delta)}$

$$F_{\mathbf{a}}(n + T) \neq F_{\mathbf{a}}(n)$$

for at least one n with $0 \leq n < 2^k - T$.

Clearly, if $a_1 = \ldots = a_{k-r} = 1$ then obviously

$$F_{\mathbf{a}}(n) = F_{\mathbf{a}}(n + 2^r), \qquad 0 \leq n \leq 2^k - 2^r - 1,$$

thus in this case the period of the sequence $F_{\mathbf{a}}(n)$ is at most 2^r. Which shows that only "statistical" results which apply to almost all (but not all) vectors $\mathbf{a} = (\mathbb{Z}/t\mathbb{Z})^k$ are possible.

3.4. *Period of EC-SSG*

Since every linear recurrence sequence $u(n)$ of order k over $\mathbb{F}_2$ is periodic with period $\tau \leq 2^k - 1$ then the **EC-SSG** given by (7) is also periodic with some period $T \mid \tau$. However it is conceivable that $T < \tau$ and it is certainly an interesting question to describe the cases when this may happen.

Question 3.2. Give a sufficiently broad conditions which guarantee that the period of the **EC-SSG** given by (7) is the same as the period of the underlying linear recurrence sequence $u(n)$.

4. Discrepancy

4.1. *Discrepancy of EC-LCG*

Let U_n be the sequence given by (4) for $G \in \mathbf{E}(F_p)$ where the elliptic curve $\mathbf{E}$ is defined over $\mathbb{F}_p$ and p is prime.

For an integer $s \ge 1$ we consider the $2s$-dimensional points

$$\left(\frac{x\left(U_n\right)}{p}, \frac{y\left(U_n\right)}{p}, \ldots, \frac{x\left(U_{n+s-1}\right)}{p}, \frac{y\left(U_{n+s-1}\right)}{p}\right). \tag{8}$$

The following result is a special partial case of a more general estimate from [29].

Theorem 4.1. *Assume that* $\mathbf{E}$ *is an elliptic curve over* $\mathbb{F}_p$ *where* p *is prime and* t *is the order of* $G \in \mathbf{E}(\mathbb{F}_p)$*. Then for the* $2s$*-dimensional discrepancy* D_s *of the points* (8) *for* $n = 1, \ldots, t$ *the following bound holds:*

$$D_s = O(t^{-1} p^{1/2} (\log p)^s).$$

Clearly the bound in nontrivial only if $t \ge p^{1/2}(\log p)^s$, but as we have seen in Theorem 3.1 this holds for most of the random choices of the parameters.

It is very plausible that the same method can produce a similar bound (probably only with an extra factor of $\log p$) on the $2s$-dimensional discrepancy of the points (8) taken over a part of the period, however this has not been worked out.

Question 4.1. Prove that the $2s$-dimensional discrepancy of the set of points (8) for $n = 1, \ldots, N$, is $O(t^{-1} p^{1/2} (\log p)^{s+1})$ for any integer $N \le t$.

As we have mentioned, one can find in [29] a more general bound which applies to points constructed by using other functions defined on points on elliptic curves, rather than just $x(Q)$ and $y(Q)$.

4.2. *Discrepancy of EC-PG*

The discrepancy of the points associated with the sequence (5) has been estimated in [40].

Theorem 4.2. *Assume that* $\mathbf{E}$ *is an elliptic curve over* $\mathbb{F}_p$ *where* p *is prime and* t *is the order of* $G \in \mathbf{E}(\mathbb{F}_p)$*. Let* T *be the period of the sequence* (5)*. Then for any fixed integer* $\nu \ge 1$*, the discrepancy* D *of the points*

$$\frac{x\left(W_n\right)}{p}, \qquad n = 1, \ldots, T,$$

the following bound holds:

$$D = O\left(T^{-(3\nu+2)/2\nu(\nu+2)} t^{(\nu+1)/\nu(\nu+2)} p^{1/4(\nu+2)} \log p\right).$$

The optimal choice of ν depends on the relation between T, t and p. For example, if $T = p^{1+o(1)}$, which is typically the case, see Theorem 3.1, then $\nu = 1$ is the optimal choice which leads to the bound $D = O\left(p^{-1/12+o(1)}\right)$.

One the other hand, if $T \geq t^{2/3} q^{1/6+\varepsilon}$ for some fixed $\varepsilon > 0$, then taking sufficiently large ν makes the bound of Theorem 4.2 nontrivial.

A nontrivial upper bound on the discrepancy of the points (5) for $n = 1, \ldots, N$, where $N \leq T$ has been given in [20].

We remark that an alternative way of estimating exponential sums, and thus the discrepancy for the sequence (5) is given in [2].

We however do not see any plausible approaches to estimating the multidimensional discrepancy (on full or a part of the period) in the style of Theorem 4.1.

Question 4.2. Obtain a nontrivial bound on the s-dimensional discrepancy of the points

$$\left(\frac{x\left(W_n\right)}{p}, \ldots, \frac{x\left(W_{n+s-1}\right)}{p}\right)$$

for $n = 1, \ldots, N$, where $N \leq t$.

As we have mentioned even the case of $N = t$ is of interest and seems to require new ideas to be resolved.

4.3. *Discrepancy of EC-NRG*

We now turn our attention to the sequence $F_{\mathbf{a}}(n)$ given by (6). The following result has been obtained in [53].

Theorem 4.3. *Assume that* $\mathbf{E}$ *is an elliptic curve over* $\mathbb{F}_p$ *where* p *is prime Let* $G \in \mathbf{E}(\mathbb{F}_p)$ *be of prime order* t*. Then for any* $\delta > 0$*, for a random vector* $\mathbf{a}$ *chosen uniformly from* $(\mathbb{Z}/t\mathbb{Z})^k$ *and the sequence* $F_{\mathbf{a}}(n)$ *given by* (6)*, with probability at least* $1 - \delta$ *discrepancy* $D_{\mathbf{a}}$ *of the points*

$$\frac{x\left(F_{\mathbf{a}}(n)\right)}{p}, \qquad n = 1, \ldots, 2^k,$$

the following bound holds:

$$D_{\mathbf{a}} = O\left(\delta^{-1} B(k, t, p) \log^2 p\right),$$

where

$$B(n,l,p) = 2^{-k/2} + 3^{k/2}2^{-k}t^{-1/2}p^{1/4} + k^{1/2}t^{-1/2} + t^{-1}p^{1/2}.$$

It is easy to check that the bound of Theorem 4.3 is nontrivial beginning with

$$t \ge \max\left\{p^{1/2+\varepsilon}, k^{1+\varepsilon}\right\}$$

with any fixed $\varepsilon > 0$. It is natural to select k of order $\log p$ (thus the definition domain of $F_{\mathbf{a}}$ and the value domain are of approximately the same size) the second term can be dropped. In fact, in the most interesting case when k is about the bit length of p, thus $k = \log p + O(1)$ we obtain $B(k,t,p) \ll B(t,p)$ where

$$B(t,p) = \begin{cases} t^{-1/2}p^{1/2-\gamma/2}, & \text{if } l \ge p^{\gamma}; \\ t^{-1}p^{1/2}, & \text{if } t < p^{\gamma}; \end{cases}$$

and $\gamma = 2.5 - \log 3 = 0.9150\ldots$.

We now mention several open questions.

Question 4.3. Obtain a nontrivial upper bound on the discrepancy of the points

$$\frac{x\left(F_{\mathbf{a}}(n)\right)}{p}, \qquad n = 1, \ldots, N,$$

where $N < 2^k$.

In principle, the method of proof of Theorem 4.3 should apply to Question 4.3, however optimizing some parameters which occur in the proof, in order to get the best possible bound for this approach, can be more complicated in this case.

The next question is more of theoretic value, since in most of the practical applications of the sequence (6), t is chosen to be prime anyway (for example, see [45]). Still, to complete the picture, and maybe to gain more understanding about the **EC-NRG** we pose it here.

Question 4.4. Extend Theorem 4.3 to the case of points G of composite orders t.

Finally, we conclude with a question to which the method of (6) does not immediately apply (even for $N = 2^k$) and which we believe is harder than Questions 4.3 and 4.4.

Question 4.5. Obtain a nontrivial bound on the s-dimensional discrepancy of the points

$$\left(\frac{x\left(F_{\mathbf{a}}(n)\right)}{p}, \ldots, \frac{x\left(F_{\mathbf{a}}(n+s-1)\right)}{p}\right), \qquad n = 1, \ldots, N,$$

where $N < 2^k$.

4.4. *Discrepancy of EC-SSG*

The following result has been give in [18].

Theorem 4.4. *Assume that* $\mathbf{E}$ *is an elliptic curve over* $\mathbb{F}_p$ *where* p *is prime and let* $u(n)$ *be a linear recurrence sequence of elements of* $\mathbb{F}_2$ *of order* k *and period* τ*, whose characteristic polynomial* $Z^k + c_{k-1}Z^{k-1} + \ldots + c_1 Z + c_0 \in \mathbb{F}_2[Z]$ *is irreducible over* $\mathbb{F}_2$*. Then for any* $\delta > 0$*, for all but* $O(\delta p^k)$ *choices of* k *points* $G_1, \ldots, G_k \in \mathbf{E}(\mathbb{F}_p)$*, the discrepancy* $D_{G_1,\ldots,G_k}(N)$ *of the points*

$$\frac{x\left(S_n\right)}{p}, \qquad n = 1, \ldots, N,$$

where the sequence S_n *is given by* (7) *satisfies the bound*

$$D_{G_1,\ldots,G_k}(N) = O\left(\delta^{-1} \min\{N^{-1/2}, p^{-1/4}\} \log^3 p\right),$$

for every $N \le \tau$*.*

Certainly obtaining an analogue of Theorem 4.4 is a very important task, which is likely to require some new ideas.

Question 4.6. Estimate the s-dimensional discrepancy of the points

$$\left(\frac{x\left(S_n\right)}{p}, \ldots, \frac{x\left(S_{n+s-1}\right)}{p}\right), \qquad n = 1, \ldots, N,$$

where the sequence S_n is given by (7).

5. Linear Complexity

5.1. *Linear Complexity of EC-LCG*

The following estimate is a special partial case of a much more general result from [29].

Theorem 5.1. *Assume that* $\mathbf{E}$ *is an elliptic curve over* $\mathbb{F}_p$ *where* p *is prime Let* $G \in \mathbf{E}(\mathbb{F}_p)$ *be of prime order* t*. The linear complexity* $L(N)$ *of the*

sequence of $x(U_n)$, $n = 1, \ldots, N$, *where the sequence* U_n *given by* (4), *satisfies*

$$L(N) \ge \begin{cases} \min\{N/3, t/3\}, & \text{if } \mathcal{O} = U_0, \\ \min\{N/4, t/3\}, & \text{if } \mathcal{O} \in \langle G \rangle \oplus U_0, \\ \min\{N/3, t/2\}, & \text{otherwise,} \end{cases}$$

where $\langle G \rangle$ *is the subgroup of* $\mathbf{E}(\mathbb{F}_p)$ *generated by* G.

Probably the constants in the denominators can be improved slightly, but overall the bound is quite satisfactory.

5.2. *Linear Complexity of EC-PG*

Unfortunately for the **EC-PG** the lower bound, obtained in [40] is more modest.

Theorem 5.2. *Assume that* $\mathbf{E}$ *is an elliptic curve over* $\mathbb{F}_p$ *where* p *is prime and* t *is the order of* $G \in \mathbf{E}(\mathbb{F}_p)$. *Let* T *be the period of the sequence* (5). *The linear complexity* L *of the sequence of* $x(W_n)$, $n = 1, \ldots, 2T$, *where the sequence* W_n *given by* (4), *satisfies*

$$L \gg Tt^{-2/3}.$$

As we have mentioned in Section 1.4, the linear complexity of a periodic sequence of any periodic sequences of period T achieves it largest value at the interval of length $2T$. However shorter intervals are of interest too.

Question 5.1. Obtain a nontrivial bound lower bound on the the linear complexity of the sequence of $x(W_n)$, $n = 1, \ldots, N$, for $N \le 2T$.

5.3. *Linear Complexity of EC-NRG*

The following result is shown in [56].

Theorem 5.3. *Assume that* $\mathbf{E}$ *is an elliptic curve over* $\mathbb{F}_p$ *where* p *is prime Let* $G \in \mathbf{E}(\mathbb{F}_p)$ *be of prime order* t. *Suppose that* $\gamma > 0$ *and* k *are chosen to satisfy*

$$k \ge (2 + \gamma) \log l.$$

Then for any $\delta > 0$ *and sufficiently large* t, *the linear complexity* $L_{\mathbf{a}}$ *of the sequence of* $x(F_{\mathbf{a}}(n))$, $n = 0, \ldots, 2^k - 1$, *where the sequence* $F_{\mathbf{a}}(n)$ *given by* (6), *satisfies*

$$L_{\mathbf{a}} \gg \min\{t^{1/3-\delta}, t^{\gamma-3\delta} \log^{-2} t\}$$

for all except $O\left(t^{k-\delta}\right)$ *vectors* $\mathbf{a} \in (\mathbb{Z}/t\mathbb{Z})^k$.

Typically the bit length of p and l are of the same order as n. Thus

$$\log p \sim \log l \sim n.$$

In the most interesting case k is the bit length of p, that is, $k \sim \log p$. In this case Theorem 5.3 implies a lower bound on $L_{\mathbf{a}}$ which is exponential in k, if $t \le p^{1/2-\varepsilon}$ for some $\varepsilon > 0$. On the other hand the uniformity of distribution results of Theorem 4.3 are nontrivial for $t \ge p^{1/2+\varepsilon}$. Thus, unfortunately these results, characterizing different aspects of randomness in the **EC-NRG** do not overlap.

Question 5.2. Obtain a nontrivial bound lower bound on the the linear complexity of the sequence of $x\left(F_{\mathbf{a}}(n)\right)$, $n = 0, \ldots, 2^k - 1$, for $k \sim \log p$ and $t \ge p^{1/2}$.

Studying the linear complexity of the **EC NRG** in parts of the period is of ultimate interest as well.

Question 5.3. Obtain a nontrivial bound lower bound on the the linear complexity of the sequence of $x\left(F_{\mathbf{a}}(n)\right)$, $n = 0, \ldots, N$, for $N < 2^k$.

5.4. *Linear Complexity of EC-SSG*

Some results have been obtained in [50,51] however many basic questions about the linear complexity of the **EC-SSG** are widely open and definitely deserve more attention.

6. Alternative Approach

Here we outline an alternative approach to generating pseudorandom points on elliptic curves which is based on different ideas and has been intensively studied in the literature, see [1,33,38,39,41,44,58,59] and references therein.

We explain this construction in the simplest case of *Koblitz curves* $\mathbf{E}_a$ defined over $\mathbb{F}_2$ by an equation of the form

$$Y^2 + XY = X^3 + aX^2 + 1,$$

where $a \in \mathbb{F}_2$, which have been introduced in [33].

Let r be a sufficiently large positive integer. Then we fix a point $G \in \mathbf{E}_a(\mathbb{F}_{2^r})$ of order t.

We now consider the set of vectors

$$\mathcal{N}_k = \{(\nu_1, \ldots, \nu_k) \in \{0, \pm 1\}^k \mid \nu_j \nu_{j+1} = 0\},$$

that is, the set of vectors with coordinates 0 and ± 1 without two consecutive nonzero coordinates.

It is known, see [9], that

$$\#\mathcal{N}_k = \frac{4}{3}2^k + O(1). \tag{9}$$

It is not hard to show that every integer n has a representation of the form

$$n = \sum_{j=0}^{k-1} \nu_j 2^j$$

with some $\mathbf{n} = (\nu_0, \ldots, \nu_{k-1}) \in \mathcal{N}_k$, where $k = \log n + O(1)$, but it is more natural to index the points which we about to construct by vectors $\mathbf{n} \in \mathcal{N}_k$ rather than by the corresponding integers, since the points $G_{\mathbf{n}}$ can be arranged in a sequence by ordering the vectors $\mathbf{n} \in \mathcal{N}_k$ lexicographically.

Let σ be the *Frobenius endomorphism*, acting on points $(x, y) \in \mathbb{F}_{2^r}^2$ as

$$\sigma((x, y)) = (x^2, y^2).$$

Clearly, if $Q \in \mathbf{E}_a(\mathbb{F}_{2^r})$ then $\sigma(Q) \in \mathbf{E}_a(\mathbb{F}_{2^r})$ as well.

For $k \le r$ we consider the points

$$G_{\mathbf{n}} = \sum_{j=0}^{k-1} \nu_j \sigma^j(G), \qquad \mathbf{n} = (\nu_0, \ldots, \nu_{k-1}) \in \mathcal{N}_k. \tag{10}$$

We start with the remark that a similar construction can also be implemented with elements of the multiplicative group $\mathbb{F}_{2^r}^*$. That is, instead of the points (10), one can consider the points

$$g_{\mathbf{n}} = \prod_{j=0}^{k-1} \left(\sigma^j(g)\right)^{\nu_j}, \qquad \mathbf{n} = (\nu_0, \ldots, \nu_{k-1}) \in \mathcal{N}_k,$$

for some fixed point $g \in \mathbb{F}_{2^r}^*$. However, its full advantages can only be seen on elliptic curves. To demonstrate this, we observe the following:

- it does not involve any doubling of points, which is more expensive than addition of distinct points on elliptic curves (while in a finite field multiplication and squaring both cost the same);
- it involves subtraction which costs the same as addition (while in a finite field division is more expensive than multiplication).

We now however present the following result of [41], which give a partial characterization of the behaviour of these points.

Theorem 6.1. *Let $G \in \mathbf{E}_a(\mathbb{F}_{2^r})$ be of prime order t, where $a \in \mathbb{F}_2$. For any integers k and s with $1 \le s \le k$ and $2^s \le t/8$, and for every point $Q \in \mathbf{E}_a(\mathbb{F}_{2^r})$ the number of points $G_{\mathbf{n}}$ given by (10) with $G_{\mathbf{n}} = Q$, where $\mathbf{n} \in \mathcal{N}_k$, does not exceed $\#\mathcal{N}_{k-s}$.*

In particular, the bound of Theorem 6.1 implies that if $2^k < t/8$ then the points $G_{\mathbf{n}}$ are all distinct. For larger values of k choosing $s = \lfloor \log_2 t \rfloor - 3$, from (9) we conclude that for any point $Q \in \mathbf{E}_a(\mathbb{F}_{2^r})$ the number of representations $G_{\mathbf{n}} = Q$, with $\mathbf{n} \in \mathcal{N}_k$ is $O(2^k t^{-1})$.

The method of proofs uses some facts about resultants in the residue ring $\mathbb{Z}/t\mathbb{Z}$ and does not immediately extend to points G of composite order t. The condition of primality of t is quite adequate and in fact natural for cryptographic applications. Still, obtaining a more general results would certainly be of interest.

Question 6.1. Obtain an analogue of of Theorem 6.1 for points G of arbitrary order t.

On the other hand, there are extensions of Theorem 6.1 to more general fields and elements of the ideal class group of hyperelliptic curves, see [41] for more details.

Theorem 6.1 has been used in a substantial way in [42] where the distribution of the sequence (10) has been studied.

Finally, we conclude with the following general question.

Question 6.2. Study whether the ideas used in the **EC-LCG**, **EC-PG** and **EC-NRG** can be combined with the idea of using the Frobenius endomorphism and lead to new pseudorandom number generators of cryptographic and mathematical interest.

References

1. R. Avanzi, H. Cohen, C. Doche, G. Frey, T. Lange and K. Nguyen, *Elliptic and hyperelliptic curve crytography: Theory and practice*, CRC Press, 2005.
2. W. D. Banks, J. B. Friedlander, M. Garaev and I. E. Shparlinski, 'Double character sums over elliptic curves and finite fields,' *Pure and Appl. Math. Quart.*, **2** (2006), 179–197.
3. P. Beelen and J. Doumen, 'Pseudorandom sequences from elliptic curves', *Finite Fields with Applications to Coding Theory, Cryptography and Related Areas*, Springer-Verlag, Berlin, 2002, 37–52.
4. S. R. Blackburn, D. Gomez-Perez, J. Gutierrez and I. E. Shparlinski, 'Predicting the inversive generator', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **2898** (2003), 264–275.

5. S. R. Blackburn, D. Gomez-Perez, J. Gutierrez and I. E. Shparlinski, 'Predicting nonlinear pseudorandom number generators', *Math. Comp.*, **74** (2005), 1471–1494.
6. S. R. Blackburn, D. Gomez-Perez, J. Gutierrez and I. E. Shparlinski, 'Reconstructing noisy polynomial evaluation in residue rings', *J. of Algorithms*, (to appear).
7. I. Blake, G. Seroussi and N. Smart, *Elliptic curves in cryptography*, London Math. Soc., Lecture Note Series, **265**, Cambridge Univ. Press, 1999.
8. I. Blake, G. Seroussi and N. Smart, *Advances in elliptic curves in cryptography*, London Math. Soc., Lecture Note Series, **317**, Cambridge Univ. Press, 2005.
9. W. Bosma, 'Signed bits and fast exponentiation', *J. Théorie des Nombres Bordeaux*, **13** (2001), 27–41.
10. J. Boyar, 'Inferring sequences produced by pseudo-random number generators', *J. ACM*, **36** (1989), 129–141.
11. J. Boyar, 'Inferring sequences produces by a linear congruential generator missing low–order bits', *J. Cryptology*, **1** (1989) 177–184.
12. M. Caragiu, R. A. Johns and J. Gieseler, 'Quasi-random structures from elliptic curves', *J. Algebra, Number Theory and Appl.*, **6** (2006), 561–571.
13. O. Chevassut, P.-A. Fouque, P. Gaudry and D. Pointcheval, 'The twist-AUgmented technique for key exchange', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **3958** (2006), 410–426.
14. A. Conflitti and I. E. Shparlinski, 'On the multidimensional distribution of the subset sum generator of pseudorandom numbers', *Math. Comp.*, **73** (2004), 1005–1011.
15. S. Contini and I. E. Shparlinski, 'On Stern's attack against secret truncated linear congruential generators', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **3574** (2005), 52–60.
16. T. W. Cusick, C. Ding and A. Renvall, *Stream ciphers and number theory*, Elsevier, Amsterdam, 2003.
17. M. Drmota and R. Tichy, *Sequences, discrepancies and applications*, Springer-Verlag, Berlin, 1997.
18. E. El Mahassni, 'On the distribution of the elliptic subset sum generator of pseudorandom numbers', *Preprint*, 2007.
19. E. El Mahassni and I. E. Shparlinski, 'On the uniformity of distribution of congruential generators over elliptic curves', *Proc. Intern. Conf. on Sequences and their Applications, Bergen 2001*, Springer-Verlag, London, 2002, 257–264.
20. E. El Mahassni and I. E. Shparlinski, 'On the distribution of the elliptic curve power generator', *Preprint*, 2007.
21. R. R. Farashahi, B. Schoenmakers and A. Sidorenko, 'Efficient Pseudorandom Generators Based on the DDH Assumption', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **4450**, (2007), 426–441.
22. A. M. Frieze, J. Håstad, R. Kannan, J. C. Lagarias and A. Shamir, 'Reconstructing truncated integer variables satisfying linear congruences', *SIAM J. Comp.*, **17** (1988), 262–280.

23. J. von zur Gathen and I. E. Shparlinski, 'Predicting subset sum pseudorandom number generators', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **3357** (2005), 241–251.
24. D. Gomez-Perez, J. Gutierrez and Á. Ibeas, 'Attacking the Pollard generator', *IEEE Trans. Inform. Theory*, **52** (2006), 5518–5523.
25. G. Gong, T. A. Berson and D. A. Stinson, 'Elliptic curve pseudorandom sequence generators', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **1758** (2000), 34–49.
26. G. Gong and C. C. Y. Lam, 'Linear recursive sequences over elliptic curves', *Proc. Intern. Conf. on Sequences and their Applications, Bergen 2001*, Springer-Verlag, London, 2002, 182–196.
27. J. Gutierrez and Á. Ibeas, 'Inferring sequences produced by a linear congruential generator on elliptic curves missing high-order bits', *Designs, Codes and Cryptography*, (to appear)
28. S. Hallgren, 'Linear congruential generators over elliptic curves', *Preprint CS-94-143*, Dept. of Comp. Sci., Cornegie Mellon Univ., 1994, 1–10.
29. F. Hess and I. E. Shparlinski, 'On the linear complexity and multidimensional distribution of congruential generators over elliptic curves', *Designs, Codes and Cryptography*, **35** (2005), 111–117.
30. H. Hu, L. Hu and D. Feng, 'On a class of pseudorandom sequences from elliptic curves over finite fields *IEEE Trans. Inform. Theory*, **53** (2007), 2598–2605.
31. A. Joux and J. Stern, 'Lattice reduction: A toolbox for the cryptanalyst', *J. Cryptology*, **11** (1998), 161–185.
32. B. S. Kaliski, 'One-way permutations on elliptic curves', *J. Cryptology*, **3** (1991), 187–199.
33. N. Koblitz, 'CM curves with good cryptographic properties', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **576** (1992), 279–287.
34. D. R. Kohel and I. E. Shparlinski, 'Exponential sums and group generators for elliptic curves over finite fields', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **1838** (2000), 395–404.
35. H. Krawczyk, 'How to predict congruential generators', *J. Algorithms*, **13** (1992), 527–545.
36. L. Kuipers and H. Niederreiter, *Uniform distribution of sequences*, Wiley-Intersci., New York-London-Sydney, 1974.
37. J. C. Lagarias, 'Pseudorandom number generators in cryptography and number theory', *Proc. Symp. in Appl. Math.*, Amer. Math. Soc., Providence, RI, **42** (1990), 115–143.
38. T. Lange, *Efficient arithmetic on hyperelliptic curves*, PhD thesis, Universität Gesamthochschule Essen, 2001.
39. T. Lange, 'Koblitz Curve Cryptosystems', *Finite Fields and Their Applications* , (to appear).
40. T. Lange and I. E. Shparlinski, 'Certain exponential sums and random walks on elliptic curves', *Canad. J. Math.*, **57** (2005), 338–350.
41. T. Lange and I. E. Shparlinski, 'Collisions in fast generation of ideal classes and points on hyperelliptic and elliptic curves', *Appl. Algebra in Engin.,*

Commun. and Computing, **15** (2005), 329–337.
42. T. Lange and I. E. Shparlinski, 'Distribution of some sequences of points on elliptic curves', *J. Math. Cryptology*, **1** (2007), 1–11.
43. R. Lidl and H. Niederreiter, *Finite fields*, Cambridge University Press, Cambridge, 1997.
44. V. Müller, 'Fast multiplication on elliptic curves over small fields of characteristic two', *J. of Cryptology*, **11** (1998), 219–234.
45. M. Naor and O. Reingold, 'Number-theoretic constructions of efficient pseudo-random functions', *Proc 38th IEEE Symp. on Found. of Comp. Sci.*, IEEE, 1997, 458–467.
46. H. Niederreiter, *Random number generation and Quasi–Monte Carlo methods*, SIAM Press, 1992.
47. H. Niederreiter, 'Design and analysis of nonlinear pseudorandom number generators', *Monte Carlo Simulation*, A.A. Balkema Publishers, Rotterdam, 2001, 3–9.
48. H. Niederreiter and I. E. Shparlinski, 'Recent advances in the theory of nonlinear pseudorandom number generators', *Proc. Conf. on Monte Carlo and Quasi-Monte Carlo Methods, 2000*, Springer-Verlag, Berlin., 2002, 86–102.
49. H. Niederreiter and I. E. Shparlinski, 'Dynamical systems generated by rational functions', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **2643** (2003), 6–17.
50. R. A. Rueppel, *Analysis and design of stream ciphers*, Springer-Verlag, Berlin, 1986.
51. R. A. Rueppel, 'Stream ciphers', *Contemporary cryptology: The science of information integrity*, IEEE Press, NY, 1992, 65–134.
52. R. A. Rueppel and J. L. Massey, 'Knapsack as a nonlinear function', *IEEE Intern. Symp. of Inform. Theory*, IEEE Press, NY, 1985, 46.
53. I. E. Shparlinski, 'On the Naor–Reingold pseudo-random function from elliptic curves', *Appl. Algebra in Engin., Commun. and Computing*, **11** (2000), 27–34.
54. I. E. Shparlinski, *Cryptographic applications of analytic number theory*, Birkhauser, 2003.
55. I. E. Shparlinski, 'Orders of points on elliptic curves', *Affine Algebraic Geometry*, Amer. Math. Soc., 2005, 245–252.
56. I. E. Shparlinski and J. H. Silverman, 'On the linear complexity of the Naor–Reingold pseudo-random function from elliptic curves', *Designs, Codes and Cryptography*, **24** (2001), 279–289.
57. J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, Berlin, 1995.
58. N. P. Smart, 'Elliptic curve cryptosystems over small fields of odd characteristic', *Journal of Cryptology*, **12** (1999), 141–151.
59. J. Solinas, 'Efficient arithmetic on Koblitz curves', *Designs, Codes and Cryptography*, **19** (2000), 195–249.
60. A. Topuzoğlu and A. Winterhof, 'Pseudorandom sequences', *Topics in Geometry, Coding Theory and Cryptography*, Springer-Verlag, 2006, 135–166.

Symmetric Cryptography and Algebraic Curves

José Felipe Voloch

Department of Mathematics, University of Texas, Austin, TX 78712, USA
E-mail: voloch@math.utexas.edu
url: http://www.ma.utexas.edu/˜voloch

We discuss some applications of the theory of algebraic curves to the study of S-boxes in symmetric cryptography.

1. Introduction

A symmetric block cipher usually consists of several iterations, known as rounds, of the following operations on the input message: An $\mathbb{F}_2$-linear transformation (to "mix the bits"), a non-linear map (consisting of one or several S-boxes) and the $\mathbb{F}_2$-addition of part of the key. For our purposes an S-box is simply a map $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$. Two well-known attacks on such ciphers, differential and linear cryptanalysis, exploit situations in which an S-box is "close to $\mathbb{F}_2$-linear". There are two corresponding measures of nonlinearity for S-boxes, which we define below. These are closely related (see [3]).

For a function $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ we define

$$\delta(f) = \max_{\alpha \neq 0, \beta} \#\{x \in \mathbb{F}_{2^n} \mid f(x+\alpha) - f(x) = \beta\}$$

For any f, $\delta(f)$ is a positive even integer and if f is a polynomial of degree m then $\delta(f) \leq m-1$ unless f is an additive polynomial plus a constant. To defend against differential cryptanalysis one needs $\delta(f)$ to be small. A function f is said to be almost perfectly nonlinear (APN) if $\delta(f) = 2$. In this paper we will study the behaviour of $\delta(f)$ for polynomials f.

For a function $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ we define

$$\lambda(f) = \max_{\alpha \neq 0, \beta} |\#\{x \in \mathbb{F}_{2^n} \mid \mathrm{Tr}(\alpha f(x) + \beta x) = 0\} - 2^{n-1}|$$

For any f, $\lambda(f) \geq 2^{(n-1)/2}$ and if f is a polynomial of degree m which is not an additive polynomial plus a constant then $\lambda(f) \leq (m-1)2^{(n-1)/2}$.

To defend against linear cryptanalysis one needs $\lambda(f)$ to be small. The function f is said to be almost bent if $\lambda(f) = 2^{(n-1)/2}$. We will not discuss λ in this paper.

The S-box used by AES is $s : \mathbb{F}_{2^8} \to \mathbb{F}_{2^8}, s(x) = x^{-1}, x \neq 0, s(0) = 0$. We have that $\delta(s) = 4$. More generally, the same function on $\mathbb{F}_{2^n}$ has $\delta(s) = 2$ for n odd and $\delta(s) = 4$ for n even. (see [8]).

Other examples of APN functions are the Gold functions, $f(x) = x^{2^j+1}, (n, j) = 1$, and the Kasami functions, $f(x) = x^{4^j-2^j+1}, (n, j) = 1$. These are the only examples known of polynomials which are APN in $\mathbb{F}_{2^n}$ for infinitely many n and we conjecture that there are no others, up to a natural equivalence which we define in section 2.

Edel, Kyureghan and Pott [4] gave the following example of an APN function which is not equivalent in the natural sense alluded above to a monomial: $f(x) = x^3 + \omega x^{36}$ as a function on $\mathbb{F}_{2^{10}}$, where ω is a primitive cube root of unity. Byrne and McGuire [1] showed that this function is APN for only finitely many fields $\mathbb{F}_{2^n}$. We will give a new proof of this result as a consequence of a more general fact, stated in Theorem 3 below.

For monomials, the following is known. For an integer $m > 0$, define l to be the largest integer such that 2^l divides $m - 1$. Also, let $m' = (m - 1)/2^{l-1} + 1$ and $d = (m - 1, 2^l - 1) = ((m' - 1)/2, 2^l - 1)$. Then Jedlicka [7] proved that if $d < (m - 1)/2^l, m > 5$ then $f(x) = x^m$ can only be APN for finitely many fields $\mathbb{F}_{2^n}$. He also showed that $f(x) = x^{-m}, x \neq 0, f(0) = 0$, for $m \equiv 1(\mod 4), m > 5$ can only be APN for finitely many fields $\mathbb{F}_{2^n}$.

It is not hard to show that if a polynomial f is APN, then the curves $F_\alpha(x, y) = 0$ have very few rational points, where

$$F_\alpha(x, y) = (f(x + \alpha) + f(x) + f(y + \alpha) + f(y))/((x + y)(x + y + \alpha)).$$

If one of these curves has an absolutely irreducible factor defined over $\mathbb{F}_{2^n}$ and n is large enough, then Weil's theorem guarantees that the curve has enough points so that f is not APN. (See, e.g., [1] Theorem 4 for a similar argument).

In the work of Jedlicka (which extends work of Janwa, McGuire and Wilson [6]) and the work of Byrne and McGuire mentioned above, they show the existence of this absolutely irreducible factor by using intersection theory and a study of the singularities of the curve. See also the work of Férard and Rodier [5].

To study the values of $\delta(f)$ other than $\delta(f) = 2$ we need to study other curves in addition to $F_\alpha(x, y) = 0$.

Our main results are the following:

Theorem 1.1. *For a given integer $m > 4, m \equiv 0, 3(\mod 4)$ let $\delta_0 = m-1$ or $m-2$ according to whether m is odd or even. Then most polynomials f of degree m over $\mathbb{F}_{2^n}$ satisfy $\delta(f) = \delta_0$. More precisely,*

$$\lim_{n\to\infty} \frac{\#\{f \in \mathbb{F}_{2^n}[x] \mid \deg f = m, \delta(f) = \delta_0\}}{\#\{f \in \mathbb{F}_{2^n}[x] \mid \deg f = m\}} = 1$$

Theorem 1.2. *Let m be an integer $m > 4$ and $f(x) = x^m + a_1x^{m-1} + a_2x^{m-2} + \cdots$ be a polynomial over $\mathbb{F}_{2^n}$. Then $\delta(f)$ is strictly smaller than $m-1$ or $m-2$ according to whether m is odd or even, provided that one of the following holds:*

(i) $a_1 = 0$ *and* $m \equiv 0(\mod 4)$
(ii) $a_2 = 0$, *n is odd and* $m \equiv 5(\mod 8)$
(iii) $a_1^2 + a_2 = 0$, *n is odd and* $m \equiv 3(\mod 8)$

Theorem 1.3. *Let $f(x) = x^m + cx^r$, where $c \in \overline{\mathbb{F}}_2^*$, $3 \leq r < m$ are coprime integers, neither a power of two and such that $(m-1, r-1)$ is a power of two. Then F_α is irreducible in $\overline{\mathbb{F}}_2[x, y, \alpha]$. Consequently, if f is defined over $\mathbb{F}_{2^n}$, $\delta(f) > 2$ for n sufficiently large with respect to m.*

2. Proofs

We start with Theorem 1. Fix for the moment $\alpha \in \mathbb{F}_{2^n}^*$ and $m > 3$ odd. Let f be a polynomial in $\mathbb{F}_{2^n}[x]$ of degree at most m, then there exists a polynomial g in $\mathbb{F}_{2^n}[x]$ of degree at most $d = (m-1)/2$ such that $f(x+\alpha)+f(x) = g(x(x+\alpha))$. The function $L : f \mapsto g$ is linear and its kernel consists of the polynomials of the form $h(x(x+\alpha)), \deg h \leq d$. It follows that the kernel has dimension $d+1$ and since f varies on a space of dimension $m+1$, it follows that L is surjective. If m is even, then let $d = (m-2)/2$, then again $f(x+\alpha) + f(x) = g(x(x+\alpha)), \deg g \leq d$. The kernel of L is now of dimension $m/2+1$ and again it follows that L is surjective. Note that either way, d is odd by hypothesis.

We now define a polynomial f to be generic if $\deg f = m$ and if $L(f)$ is a polynomial of degree d which when viewed as a morphism from $\mathbb{P}^1$ to itself is such that above each affine branch point there is only one ramification point and the ramification degree of such points is 2. The condition on $L(f)$ defines a Zariski open dense set of the coefficients of $L(f)$ and since L is surjective, we get an open dense condition on the coefficients of f as well.

The genericity condition ensures that the geometric Galois group of the Galois closure of the map $L(f)$ is the symmetric group S_d. (This follows from a classical argument, see e.g. [2] Lemma 2.3 and the paragraph

following that lemma.) We now compute the Galois group of the Galois closure of the map f. If t is a coordinate on $\mathbb{P}^1$ and $u_0, \dots, u_{d-1}$ are the roots of $L(f)(u) = t$, then the roots of $f(x) = t$ are the solutions of $x_i^2 + \alpha x_i = u_i, i = 0, \dots, d-1$. For each place v of $F = \mathbb{F}_{2^n}(t, u_0, \dots, u_{d-1})$ above $t = \infty$ we have that u_0 has a simple pole and that $u_j = \zeta^{\sigma(j)} u_0 + O(1)$, where ζ is a primitive d-th root of unity. The map $v \mapsto \sigma$ gives a bijection between the places of F above $t = \infty$ and S_d.

Let J be a set of indices such that $\sum_{j\in J} u_j$ has no poles (i.e. is constant), then $\sum_{j\in J} \zeta^{\sigma(j)} = 0, \forall \sigma \in S_d$. If J is neither empty nor the whole of $\{0, \dots, d-1\}$, consider the last equation with σ the identity and also $\sigma = (j_0 j_1)$ where $j_0 \in J, j_1 \notin J$. It follows that $\zeta^{j_0} = \sum_{j\in J, j\neq j_0} \zeta^j = \zeta^{j_1}$, contradiction. Now, we have that $\sum_j u_j = b_1/b_0$, where $L(f) = b_0 x^d + b_1 x^{d-1} + \cdots$. It follows that the geometric Galois group of the Galois closure of the map f is an extension of S_d by $(\mathbb{Z}/2)^{d-1}$ and the arithmetic Galois group is either that or has an extra $\mathbb{Z}/2$ depending on whether $b_1/b_0 = y^2 + \alpha y, y \in \mathbb{F}_{2^n}$ or not. If the former is the case, then the curve corresponding to the Galois closure of the map f is defined over $\mathbb{F}_{2^n}$ and an application of Chebotarev's density theorem gives the existence of d distinct pairs $x_i, x_i + \alpha$ with $f(x_i) + f(x_i + \alpha) = \beta$ for some β and thus $\delta(f) = 2d = \delta_0$, if n is large enough.

Now, if $f(x) = x^m + a_1 x^{m-1} + a_2 x^{m-2} + \cdots$, then if $m \equiv 3(\mod 4)$

$$b_1/b_0 = (1 + \binom{d}{2})\alpha^2 + a_1\alpha + a_2,$$

whereas if $m \equiv 0(\mod 4)$,

$$b_1/b_0 = (1 + \binom{d}{2})\alpha^2 + (\alpha^3 + a_2\alpha + a_3)/a_1.$$

It is not hard to check (see the proof of Theorem 2, item (iii)) that for n large enough it will exist α such that $b_1/b_0 = y^2 + \alpha y, y \in \mathbb{F}_{2^n}$, unless $m \equiv 3(\mod 8)$, n is odd and $a_1^2 + a_2 = 0$. So, if we assume that f is generic and $a_1^2 + a_2 \neq 0$, we will be sure to find y for n sufficiently large. This completes the proof of Theorem 1.

We now prove Theorem 2.

For item (i), note that $\deg L(f) < d$, when $a_1 = 0$.

For item (ii), with notation as in the proof of Theorem 1, we find that $b_1/b_0 = \binom{d}{2}\alpha^2 + a_2 = \alpha^2$ under the current hypothesis, and this cannot be of the form $y^2 + \alpha y, y \in \mathbb{F}_{2^n}$ for n odd.

For item (iii), using the formulas for b_1/b_0 from the proof of Theorem 1 (which are valid for all monic polynomials f), we get that if all u_i are in

$\mathbb{F}_{2^n}$ then there exists $y \in \mathbb{F}_{2^n}$ such that

$$(1 + \binom{d}{2})\alpha^2 + a_1\alpha + a_2 = y^2 + \alpha y.$$

Thus $\mathrm{Tr}_{\mathbb{F}_{2^n}/\mathbb{F}_2}(((1 + \binom{d}{2})\alpha^2 + a_1\alpha + a_2)/\alpha^2) = 0$ but the trace is easily seen to be $\mathrm{Tr}_{\mathbb{F}_{2^n}/\mathbb{F}_2}(1) = 1$ under the assumptions of item (iii). This proves Theorem 2.

Finally, we prove Theorem 3. $F_\alpha(\alpha x, \alpha y) = \alpha^{r-2}(G\alpha^k + H)$, where $k = m - r$,

$$G = ((x+1)^m + x^m + (y+1)^m + y^m)/((x+y)(x+y+1))$$

and

$$H = c((x+1)^r + x^r + (y+1)^r + y^r)/((x+y)(x+y+1)).$$

We claim that it suffices to show that G and H are non-zero, have no common factor and that G/H is not an s-th power for any divisor $s > 1$ of k. Indeed, if these conditions hold, $G\alpha^k + H$ is irreducible as a polynomial in α so if it factors as a polynomial in α, x, y then one of the factors does not involve α and this contradicts the fact that G, H have no common factor. Once we know the polynomial is irreducible as a polynomial in α, x, y, the Lang-Weil estimate implies that, for n large enough, there exists $\alpha, x, y \in \mathbb{F}_{2^n}, x \neq y, y + \alpha, F_\alpha(x, y) = 0$, hence $\delta(f) > 2$.

Assume first that both m and r are odd. Then, $G(x, x) = (x+1)^{m-1} + x^{m-1}, H(x, x) = c((x+1)^{r-1} + x^{r-1})$ are clearly non-zero. Furthermore, the map $x \mapsto (x+1)/x$ establishes a bijection between the roots of $G(x, x)$ and the $(m-1)$-st roots of unity distinct from 1 and likewise for $H(x, x)$. It now follows that G and H have no common factor since $(m-1, r-1)$ is a power of 2 and thus G/H is not an s-th power for any divisor $s > 1$ of k, since $G(x, x)$ and $H(x, x)$ are of the form a polynomial with distinct roots raised to a power of two.

If either m or r is even, then since neither is a power of two, neither G nor H is identically zero. If say, $m = 2^u v$ is even but r, v are odd then $G = ((x+y)(x+y+1))^{2^u-1}K^{2^u}$ for some polynomial K with $K(x, x)$ non-zero and a similar argument as in the case m, r odd applies with K in place of G. Finally the case m odd, r even is obtained by reversing the roles of G and H.

3. Low degrees

We conclude with some consequences of our methods in the case of low degrees. First of all, note that $\delta(f)$ is unchanged if f is replaced by $f + h$

where h is an additive polynomial or if f is replaced by $f(ax+b)/c$ where, $a, b, c \in \mathbb{F}_{2^n}, ac \neq 0$. It is elementary that $\delta(f) = 2$ if $\deg f \leq 4$ unless f is an additive polynomial plus a constant.

In the case of $\deg f = 5$, the above allows us to reduce the problem to the consideration of two cases $f = x^5, x^5 + x^3$. Theorem 2 gives that $\delta(x^5) = 2$ for n odd and one can check that $\delta(x^5) = 4$ for $n > 2$ even. Theorem 3 gives that $\delta(x^5 + x^3) = 4$ for n large and in fact $n > 2$ suffices.

The cases of $\deg f \geq 6$ cannot be fully analysed just by referring to the theorems but some cases can still be fully analysed by our methods as follows.

If $\deg f = 6$ and $L(f)$ is a separable polynomial, the extension $F/\mathbb{F}_{2^n}(t)$, from the proof of Theorem 1, has to have Galois group S_2 and the proof of Theorem 1 gives $\delta(f) = 4$ unless $a_1 = a_3 = 0$. In the case $a_1 = a_3 = 0$, f is equivalent to x^6 (and $L(f)$ is inseparable) and $\delta(f) = \delta(x^3) = 2$.

The case of $\deg f = 7$. As in the proof of Theorem 1, the extension $F/\mathbb{F}_{2^n}(t)$ has Galois group S_3 or A_3 and likewise for the geometric Galois group. Since there is a place above infinity such that $u_j = \zeta^j u_0 + O(1)$ and ζ is a primitive root of unity, it follows directly that $\sum_{j\in J} u_j$ is non-constant unless $J = \emptyset$ or $J = \{0, 1, 2\}$. A calculation shows that $b_1/b_0 = a_1\alpha + a_2$ and this will be of the form $y^2 + \alpha y, y \in \mathbb{F}_{2^n}$ for some α, assuming n is large enough. If we show that the arithmetic and geometric Galois groups coincide then the proof of Theorem 1 gives that $\delta(f) = 6$. The only way the two groups can differ is if the former is S_3 but the latter is A_3. This can only happen if the quadratic polynomial $(L(f)(u) - L(f)(v))/(u - v)$ factors in a quadratic extension of $\mathbb{F}_{2^n}$ for all α and a messy but straightforward calculation shows that this polynomial is irreducible. So $\delta(f) = 6$ for all f of degree 7 if n is large enough.

Polynomials of degree 8 are equivalent to polynomials of lower degree. For degree 9 and beyond we do not have complete results.

As mentioned in the introduction, we conjecture that the only polynomials that are APN for infinitely many n are those equivalent to the Gold and Kasami functions. We also make the following conjecture:

Conjecture 3.1. *For a given integer $m > 4$, there exists $\varepsilon_m > 0$ such that for all sufficiently large n, if f is a polynomial of degree m over $\mathbb{F}_{2^n}$ for at least $\varepsilon_m 2^{2n}$ values of $\alpha \neq 0, \beta \in \mathbb{F}_{2^n}$, $\#\{x \in \mathbb{F}_{2^n} | f(x + \alpha) - f(x) = \beta\} = \delta(f)$.*

A corollary of this conjecture would be that picking $\alpha \neq 0, \beta \in \mathbb{F}_{2^n}$ at random gives a probabilistic polynomial time algorithm for computing

$\delta(f)$. The analogue of the conjecture does not hold for rational functions and already fails for x^{-1}.

References

1. E. Byrne and G. McGuire, On the non-existence of quadratic APN and crooked functions on finite fields, preprint, 2005, http://www.maths.may.ie/staff/gmg/pubs.html
2. A. O. Bender and O. Wittenberg, A Potential Analogue of Schinzel's Hypothesis for Polynomials with Coefficients in $\mathbb{F}_q[t]$, Int. Math. Res. Not. **36** (2005) 2237–2248.
3. F. Chabaud and S. Vaudenay, Links between differential and linear cryptanalysis, Advances in Cryptology-EUROCRYPT '94, A. De Santis, Ed., Lecture Notes in Computer Science, vol. 950, Springer-Verlag, New York, 1995, pp. 356–365.
4. Y. Edel, G. Kyureghyan, and A. Pott, A new APN function which is not equivalent to a power mapping preprint, http://arxiv.org/abs/math.CO/0506420, 2005.
5. E. Férard and F. Rodier, Nonlinarity of Boolean functions and hyperelliptic curves preprint, http://arxiv.org/abs/0705.1751, 2007.
6. H. Janwa, G. McGuire and R. M. Wilson, Double-error-correcting cyclic codes and absolutely irreducible polynomials over GF(2), Journal of Algebra **178** (1995) 665–676.
7. D. Jedlicka, Classifying APN Monomials, preprint 2005, http://eprint.iacr.org/2005/096.pdf, to appear in Finite Fields and Appl.
8. K. Nyberg, Differentially uniform mappings for cryptography. Advances in Cryptology - EUROCRYPT '93, T. Helleseth, Ed., Lecture Notes in Computer Science, vol. 765, Springer-Verlag, New York, 1994, pp. 55-64.

Galois invariant smoothness basis*

Jean-Marc Couveignes

Institut de Mathématiques de Toulouse,
Université de Toulouse et CNRS.
France.

Reynald Lercier

Centre d'Électronique de l'Armement,
35170 Bruz, France.

This text answers a question raised by Joux and the second author about the computation of discrete logarithms in the multiplicative group of finite fields. Given a finite residue field $\mathbf{K}$, one looks for a smoothness basis for $\mathbf{K}^*$ that is left invariant by automorphisms of $\mathbf{K}$. For a broad class of finite fields, we manage to construct models that allow such a smoothness basis. This work aims at accelerating discrete logarithm computations in such fields. We treat the cases of codimension one (the linear sieve) and codimension two (the function field sieve).

To Gilles Lachaud, on the occasion of his 60th birthday

1. Motivation

We look for finite fields that admit Galois invariant smoothness basis. It is known that such basis accelerate the calculation of discrete logarithms. We first recall this observation by Joux and Lercier in section 2 and we give a first example of this situation in section 3. We recall in section 4 the rudiments of Kummer and Artin-Schreier theories. These theories produce the known examples of such smoothness basis. We then show in section 5 that the only extensions admitting Galois invariant flags of linear spaces are given by those two theories. In section 6, we consider a more general setting: specialization of isogenies between algebraic groups. We deduce a first non trivial example of Galois invariant smoothness basis in section 7.

*Research supported by the French Délégation Générale pour l'Armement, Centre d'Électronique de l'Armement and by the Fonds National pour la Science (ACI NIM).

In the next section 8, we show that elliptic curves produce a range of such invariant basis, provided the degree of the field is not too large.

In section 9, we recall the principles of fast sieving algorithms (the number field sieve and the function field sieve). We show in section 10 that our approach can be adapted to these algorithms. A detailed example is given in section 11. We finish with a few remarks and questions about the relevance of our method.

2. A remark by Joux and Lercier

We recall in this section the principle of a simple algorithm for computing discrete logarithms in the multiplicative group of a finite field $\mathbb{F}_q$ where $q = p^d$ and $d \geq 2$. See [7] for a survey on discrete logarithm computation.

The finite field $\mathbb{F}_q$ is seen as a residue field $\mathbb{F}_p[X]/(A(X))$ where $A(X) \in \mathbb{F}_p[X]$ is a degree d unitary irreducible polynomial. We set $x = X \bmod A(X)$. Let k be an integer such that $0 \leq k \leq d-1$ and let $V_k \subset \mathbb{F}_q$ be the $\mathbb{F}_p$-vector space generated by $1, x, \ldots, x^k$. So $V_0 = \mathbb{F}_p \subset V_1 \subset \ldots \subset V_{d-1} = \mathbb{F}_q$ and $V_k \times V_l \subset V_{k+l}$ if $k + l \leq d - 1$.

One looks for multiplicative relations between elements of V_κ for some integer κ. For example, if one takes $\kappa = 1$, the relations we are looking for take the form

$$\prod_i (a_i + b_i x)^{e_i} = 1 \in \mathbb{F}_q \tag{1}$$

where the a_i and b_i lie in $\mathbb{F}_p$. We collect such relations until we obtain a basis of the $\mathbb{Z}$-module of relations between elements in V_κ.

How do we find relations like relation (1) ? Assume again $\kappa = 1$. The simplest form of the sieving algorithm picks random triplets (a_i, b_i, e_i) and computes the remainder $r(X)$ of the Euclidean division of $\prod_i (a_i + b_i X)^{e_i}$ by $A(X)$. So

$$r(X) \equiv \prod_i (a_i + b_i X)^{e_i} \bmod A(X)$$

where $r(X)$ is a more or less random polynomial in $\mathbb{F}_p[X]$ with degree $\leq d - 1$.

We hope $r(X)$ decomposes as a product of polynomials with degree smaller than or equal to $\kappa = 1$. If this is the case, we find $r(X) = \prod_j (a'_j + b'_j X)^{e'_j}$ and we obtain a relation

$$\prod_i (a_i + b_i x)^{e_i} \prod_j (a'_j + b'_j x)^{-e'_j} = 1$$

of the expected form. One says that V_κ is the smoothness basis.

Joux and Lercier notice in [3] that, if there exists an automorphism $\mathfrak{a}$ of $\mathbb{F}_q$ such that $\mathfrak{a}(x) = ux + v$ with $u,\ v \in \mathbb{F}_p$, then the action of $\mathfrak{a}$ on equation (1) produces another equation of the same kind. Since the efficiency of discrete logarithm algorithms depends on the number of such equations one can produce in a given amount of time, one wishes to know when such useful automorphisms exist. We also wonder how to generalize this observation.

We stress that $\mathfrak{a}$ acts both on equations and factors of the form $a_i + b_i x$. Rather than increasing the number of equations, such an action may be used to lower the number of factors involved in them. If $\mathfrak{a}$ is the n-th power of the Frobenius automorphism, we obtain for free

$$\mathfrak{a}(a + bx) = (a + bx)^{p^n} = v + a + ubx$$

So we can remove $v + a + ubx$ out of the smoothness basis and replace it everywhere by $(a+bx)^{p^n}$. This way, we only keep one element in every orbit of the Galois group acting on V_κ. As a consequence, the size of the linear system we must solve is divided by the order of the group generated by $\mathfrak{a}$. If $\mathfrak{a}$ generates the full Galois group of $\mathbb{F}_q/\mathbb{F}_p$, then the number of unknowns is divided by d, the degree of the finite field $\mathbb{F}_q$.

Our concern in this text is to find models for finite fields for which the automorphisms respect the special form of certain elements. For example, if the finite field is given as above, the elements are given as polynomials in x. Any element z of the finite field has a degree: This is the smallest integer k such that $z \in V_k$. The degree of $a_0 + a_1 x + \cdots + a_k x^k$ is thus k provided $0 \leq k < d$ and $a_k \neq 0$ (and by convention, $\deg 0 = 0$). The degree is sub-additive, $\deg(w \times z) \leq \deg(w) + \deg(z)$.

The question raised boils down to asking if this degree function is preserved by the automorphisms of $\mathbb{F}_q$. It is worth noticing that the interest of the degree function in this context comes from the following properties.

- The degree is sub-additive (and often even additive): The degree of the product of two non zero elements is the sum of the degrees of either elements provided this sum is $< d$.
- The degree sorts nicely the elements of $\mathbb{F}_q$: There are q^n elements of degree $< n$ if $1 \leq n \leq d$.
- There exists a factoring algorithm that decomposes some elements in $\mathbb{F}_q$ as products of elements with smaller degrees (*e.g.* with degree $\leq \kappa$). The density of such κ-smooth elements is not too small.

In this article, we look for such degree functions on finite fields having the extra property that they are Galois invariant: Two conjugate elements have the same degree.

3. A first example

Here is an example provided by Joux and Lercier. Take $p = 43$ and $d = 6$, so $q = 43^6$, and set $A(X) = X^6 - 3$ which is an irreducible polynomial in $\mathbb{F}_{43}[X]$. So $\mathbb{F}_q$ is seen as the residue field $\mathbb{F}_{43}[X]/(X^6 - 3)$.

One checks that $p = 43$ is congruent to 1 modulo $d = 6$, so $\phi(x) = x^{43} = (x^6)^7 \times x = 3^7 x = \zeta_6 x$ where $\zeta_6 = 3^7 = 37 \bmod 43$ is a primitive sixth root of unity. Since the Frobenius ϕ generates the Galois group, one can divide by 6 the size of the smoothness basis.

In the second example provided by Joux and Lercier (and coming from XTR of type T30) one takes $p = 370801$ and $d = 30$ with $A(X) = X^{30} - 17$. This time, p is congruent to 1 modulo $d = 30$ and $\phi(x) = x^p = x^{30 \times 12360} \times x = \zeta_{30} x$ where $\zeta_{30} = 17^{12360} \bmod p = 172960 \bmod p$. As a consequence, one can divide by 30 the size of the smoothness basis.

We are here in the context of Kummer theory. In the next section we recall the basics of this theory, that classifies cyclic extensions of $\mathbb{F}_p$ with degree d dividing $p - 1$. Artin-Schreier theory is the counterpart for cyclic p-extensions in characteristic p and we sketch it as well. Both theories are of very limited interest for our purpose. We shall need to consider the more general situation of an algebraic group with rational torsion.

4. Kummer and Artin-Schreier theories

The purpose here is to classify cyclic extensions of degree $d \geq 2$ of a field $\mathbf{K}$ with characteristic p in two simple cases.

- Kummer case: p is prime to d and $\mathbf{K}$ contains a primitive d-th root of unity;
- Artin-Schreier case: $d = p$.

Kummer theory. We follow Bourbaki [1, A V.84]. According to Kummer theory, if p is prime to d and $\mathbf{K}$ contains a primitive d-th root of unity, then every degree d cyclic extension of $\mathbf{K}$ is generated by a radical.

Assume $\mathbf{K}$ is embedded in some algebraic closure $\bar{\mathbf{K}}$. To every a in $\mathbf{K}^*/(\mathbf{K}^*)^d$ (which we may regard as an element in $\mathbf{K}^*$), we associate the field $\mathbf{L} = \mathbf{K}(a^{\frac{1}{d}})$ where $a^{\frac{1}{d}}$ is any root of $X^d - a$ in $\bar{\mathbf{K}}$.

The map $x \mapsto x^d$ is an epimorphism from the multiplicative group $\bar{\mathbf{K}}^*$

onto itself. The kernel of this epimorphism is the group of d-th roots of unity. The roots of $X^d - a$ lie in the inverse image of d by this epimorphism.

The field $\mathbf{K}(a^{\frac{1}{d}})$ may not be isomorphic to $\mathbf{K}[X]/(X^d - a)$. It is when a has order d in the group $\mathbf{K}^*/(\mathbf{K}^*)^d$. On the other hand, if a lies in $(\mathbf{K}^*)^d$ then $\mathbf{K}[X]/(X^d - a)$ is the product of d copies of $\mathbf{K}$.

Let's come back to the case when a has order d in $\mathbf{K}^*/(\mathbf{K}^*)^d$. The degree d extension $\mathbf{L}/\mathbf{K}$ is Galois since, if we set $b = a^{\frac{1}{d}}$, we have

$$X^d - a = (X - b)(X - b\zeta_d)(X - b\zeta_d^2)\dots(X - b\zeta_d^{d-1})$$

where ζ_d is a primitive d-th root of unity. The Galois group of $\mathbf{L}/\mathbf{K}$ is made of transformations $\mathfrak{a}_n : x \mapsto x\zeta_d^n$ and the map $n \mapsto \mathfrak{a}_n$ is an isomorphism from the group $\mathbb{Z}/d\mathbb{Z}$ onto $\mathrm{Gal}(\mathbf{L}/\mathbf{K})$.

To avoid distinguishing too many cases, one follows Bourbaki [1, A V.84]. Rather than a single element in $\mathbf{K}^*/(\mathbf{K}^*)^d$ one picks a subgroup H of $\mathbf{K}^*$ containing $(\mathbf{K}^*)^d$ and one forms the extension $\mathbf{K}(H^{\frac{1}{d}})$ by adding to $\mathbf{K}$ all d-th roots of all elements in H. To every automorphism $\mathfrak{a}$ in $\mathrm{Gal}(\mathbf{K}(H^{\frac{1}{d}})/\mathbf{K})$, one associates an homomorphism $\psi(\mathfrak{a})$ from $H/(\mathbf{K}^*)^d$ to the group $\mu_d(\mathbf{K})$ of d-th roots of unity. The homomorphism $\psi(\mathfrak{a})$ is defined by

$$\psi(\mathfrak{a}) : \theta \mapsto \frac{\mathfrak{a}(\theta^{\frac{1}{d}})}{\theta^{\frac{1}{d}}}$$

where $\theta^{\frac{1}{d}}$ is one of the d-th roots of θ. The map $\mathfrak{a} \mapsto \psi(\mathfrak{a})$ is an isomorphism from the $\mathrm{Gal}(\mathbf{K}(H^{\frac{1}{d}})/\mathbf{K})$ onto $\mathrm{Hom}(H/(\mathbf{K}^*)^d, \mu_d(\mathbf{K}))$. This presentation of Kummer theory constructs abelian extensions of $\mathbf{K}$ with exponent dividing d.

In the case we are interested in, the field $\mathbf{K} = \mathbb{F}_q$ is finite. Any subgroup H of $\mathbf{K}^*$ is cyclic. In order to have μ_d in $\mathbf{K}$, one assumes that d divides $q - 1$. We set $q - 1 = md$. The group $(\mathbf{K}^*)^d$ has order m. The quotient $\mathbf{K}^*/(\mathbf{K}^*)^d$ is cyclic of order d. It is natural to take $H = \mathbf{K}^*$. We find the unique degree d cyclic extension $\mathbf{L}$ of $\mathbf{K}$. It is generated by a d-th root of a generator a of $\mathbf{K}^*$.

Set $b = a^{\frac{1}{d}}$ and $\mathbf{L} = \mathbf{K}(b)$. The Galois group $\mathrm{Gal}(\mathbf{L}/\mathbf{K})$ is generated by the Frobenius ϕ and the action of ϕ on b is given by $\phi(b) = b^q$, so

$$\zeta = \frac{\phi(b)}{b} = b^{q-1} = a^m$$

is a d-th root of unity that depends on a. The map $a \mapsto \zeta$ is an isomorphism of $\mathbf{K}^*/(\mathbf{K}^*)^d$ onto $\mu_d(\mathbf{K})$ which is nothing but exponentiation by m.

The limitations of this construction are clear: It requires primitive d-th roots of unity in $\mathbf{K}$. Otherwise, one may jump to some auxiliary extension

$\mathbf{K}' = \mathbf{K}(\zeta_d)$ of $\mathbf{K}$, that may be quite large. One applies Kummer theory to this bigger extension and one obtains a degree d cyclic extension $\mathbf{L}'/\mathbf{K}'$. Descent can be performed using resolvants (see [6, Chapter III.4]) at a serious computational expense. We shall not follow this track.

Example. Coming back to the first example one finds $q = p = 43$, $p - 1 = 42$, $d = 6$, $m = 7$, $a = 3$ and $\phi(b)/b = a^m = 3^7 \bmod 43$.

Artin-Schreier theory. We follow Bourbaki [1, A V.88]. If p is the characteristic of $\mathbf{K}$, then any cyclic degree p extension of $\mathbf{K}$ is generated by the roots of a polynomial of the form

$$X^p - X - a = \wp(X) - a = 0$$

where $a \in \mathbf{K}$ and the expression $\wp(X) = X^p - X$ plays a similar role to X^d in Kummer theory. The map $x \mapsto \wp(x)$ defines an epimorphism from the additive group $\bar{\mathbf{K}}$ onto itself. The kernel of this epimorphism is the additive group of the prime field $\mathbb{F}_p \subset \bar{\mathbf{K}}$.

Let a be an element of $\mathbf{K}/\wp(\mathbf{K})$ (that we may see as an element of $\mathbf{K}$ in this class). One associates to it the extension field $\mathbf{L} = \mathbf{K}(b)$ where $b \in \wp^{-1}(a)$. If a has order p in $\mathbf{K}/\wp(\mathbf{K})$, the extension $\mathbf{L}/\mathbf{K}$ has degree p and is Galois since we have

$$X^p - X - a = (X - b)(X - b - 1)(X - b - 2)\dots(X - b - (p - 1)).$$

The Galois group is made of transformations of the form $\mathfrak{a}_n : x \mapsto x + n$ and the map $n \mapsto \mathfrak{a}_n$ is an isomorphism from the group $\mathbb{Z}/p\mathbb{Z}$ onto $\mathrm{Gal}(\mathbf{L}/\mathbf{K})$.

Again, if one wishes to construct all abelian extensions of $\mathbf{K}$ with exponent p one follows Bourbaki [1, A V.88]. One takes a subgroup H of $(\mathbf{K}, +)$ containing $\wp(\mathbf{K})$ and one forms the extension $\mathbf{K}(\wp^{-1}(H))$. To every automorphism $\mathfrak{a}$ in $\mathrm{Gal}(\mathbf{K}(\wp^{-1}(H))/\mathbf{K})$, one associates an homomorphism $\psi(\mathfrak{a})$ from $H/\wp(\mathbf{K})$ onto the additive group $\mathbb{F}_p$ of the prime field. The homomorphism $\psi(\mathfrak{a})$ is defined by

$$\psi(\mathfrak{a}) : \theta \mapsto \mathfrak{a}(c) - c$$

where c belongs to $\wp^{-1}(\theta)$, the fiber of $\wp$ above θ.

The map $\mathfrak{a} \mapsto \psi(\mathfrak{a})$ is an isomorphism from the Galois group $\mathrm{Gal}(\mathbf{K}(\wp^{-1}(H))/\mathbf{K})$ onto $\mathrm{Hom}(H/\wp(\mathbf{K}), \mathbb{F}_p)$.

In our case, the field $\mathbf{K} = \mathbb{F}_q$ is finite of characteristic p. We set $q = p^f$. The morphism $\wp : \mathbb{F}_q \to \mathbb{F}_q$ has kernel $\mathbb{F}_p$ and the quotient $\mathbb{F}_q/\wp(\mathbb{F}_q)$ has order p. The unique degree p extension $\mathbf{L}$ of $\mathbb{F}_q$ is generated by $b \in \wp^{-1}(a)$ where $a \in \mathbb{F}_q - \wp(\mathbb{F}_q)$. The Galois group $\mathrm{Gal}(\mathbf{L}/\mathbf{K})$ is generated by the

Frobenius ϕ and $\phi(b) - b$ belongs to $\mathbb{F}_p$. The map $a \mapsto \phi(b) - b$ is an isomorphism from $\mathbf{K}/\wp(\mathbf{K})$ onto $\mathbb{F}_p$.

Let us make this isomorphism more explicit. We have $\phi(b) = b^q$ where $q = p^f$ is the order of $\mathbf{K} = \mathbb{F}_q$. One computes

$$\phi(b) - b = b^q - b = (b^p)^{p^{f-1}} - b = (b + a)^{p^{f-1}} - b \text{ since } \wp(b) = b^p - b = a.$$

So $b^{p^f} - b = b^{p^{f-1}} - b + a^{p^{f-1}}$. Iterating, we obtain

$$\phi(b) - b = b^{p^f} - b = a + a^p + a^{p^2} + \cdots + a^{p^{f-1}}.$$

The isomorphism from $\mathbf{K}/\wp(\mathbf{K})$ onto the additive group $\mathbb{F}_p$ is nothing but the absolute trace.

Example. Take $p = 7$ and $f = 1$, so $q = 7$. The absolute trace of 1 is 1, so we set $\mathbf{K} = \mathbb{F}_7$ and $A(X) = X^7 - X - 1$ and we set $\mathbf{L} = \mathbb{F}_{7^7} = \mathbb{F}_7[X]/(A(X))$. Setting $x = X \bmod A(X)$, one has $\phi(x) = x + 1$.

5. Invariant linear spaces of a cyclic extension

Let us recall that the question raised in section 2 concerns the existence of automorphisms that stabilize a given smoothness basis. We saw that smoothness basis are usually made using flags of linear spaces. Therefore, one wonders if, for a given cyclic extension $\mathbf{L}/\mathbf{K}$, there exists $\mathbf{K}$-vector subspaces of $\mathbf{L}$ that are left invariant by the Galois group of $\mathbf{L}/\mathbf{K}$.

Let $d \geq 2$ be an integer and $\mathbf{L} = \mathbf{K}[X]/(X^d - r)$ a Kummer extension. For any integer k between 0 and $d-1$, let $L_k = \mathbf{K} \oplus \mathbf{K}x \oplus \cdots \oplus \mathbf{K}x^k$ be the $\mathbf{K}$-vector subspace generated by the $k+1$ first powers of $x = X \bmod X^d - r$. The L_k are invariant under Galois action since for $\mathfrak{a}$, a $\mathbf{K}$-automorphism of $\mathbf{L}$, there exists a d-th root of unity $\zeta \in \mathbf{K}$ such that

$$\mathfrak{a}(x) = \zeta x$$

and $\mathfrak{a}(x^k) = \zeta^k x^k$. One has a flag of $\mathbf{K}$-vector spaces, $V_0 = \mathbf{K} \subset V_1 \subset \cdots \subset V_{d-1} = \mathbf{L}$, respected by Galois action. So the "degree" function is invariant under this action. This is exactly what happens in the two examples of section 2. If the smoothness basis is made of irreducible polynomials of degree $\leq \kappa$, then it is acted on by the Galois group.

If now $\mathbf{L} = \mathbf{K}[X]/(X^p - X - a)$ is an Artin-Schreier extension, for every integer k between 0 and $p-1$, we call $V_k = \mathbf{K} \oplus \mathbf{K}x \oplus \cdots \oplus \mathbf{K}x^k$ the $\mathbf{K}$-vector space generated by the $k+1$ first powers of $x = X \bmod X^p - X - a$. The V_k

are globally invariant under Galois action. Indeed, if $\mathfrak{a}$ is a $\mathbf{K}$-automorphism of $\mathbf{L}$, then there is a $n \in \mathbb{F}_p$ such that $\mathfrak{a}(x) = x + n$, so

$$\mathfrak{a}(x^k) = (x+n)^k = \sum_{0 \leq \ell \leq k} \binom{k}{\ell} n^{k-\ell} x^\ell.$$

We find again a flag of $\mathbf{K}$-vector spaces, $V_0 = \mathbf{K} \subset V_1 \subset \cdots \subset V_{p-1} = \mathbf{L}$, that is fixed by Galois action. This time, the Galois action is no longer diagonal but triangular. For cyclic extensions of degree a power of p, Witt-Artin-Schreier theory also produces a flag of Galois invariant vector spaces. See the beginning of Lara Thomas's thesis [8] for an introduction with references.

One may wonder if Galois invariant flags of vector spaces exist for other cyclic field extensions. Assume $\mathbf{L}/\mathbf{K}$ is a degree d cyclic extension where d is prime to the characteristic p. Let ϕ be a generator of the Galois group $C =< \phi >= \mathrm{Gal}(\mathbf{L}/\mathbf{K})$. According to the normal basis theorem [4, Theorem 13.1.], there exists a w in $\mathbf{L}$ such that

$$(w, \phi(w), \phi^2(w), \ldots, \phi^{d-1}(w))$$

is a $\mathbf{K}$-basis of $\mathbf{L}$. Therefore $\mathbf{L}$, as a $\mathbf{K}[C]$-module, is isomorphic to the regular representation. The order d of C being prime to the characteristic, the ring $\mathbf{K}[C]$ is semi-simple according to Maschke theorem [4, Theorem 1.2.]. The characteristic polynomial of ϕ acting on the $\mathbf{K}$-vector space $\mathbf{L}$ is $X^d - 1$. This is a separable polynomial in $\mathbf{K}[X]$.

To every $\mathbf{K}$-irreducible factor $f(X) \in \mathbf{K}[X]$ of $X^n - 1$, there corresponds a unique irreducible characteristic subspace $V_f \subset \mathbf{L}$, invariant by ϕ. The characteristic polynomial of ϕ restricted to V_f is f. According to Schur's lemma [4, Proposition 1.1.], any $\mathbf{K}[C]$-submodule of $\mathbf{L}$ is a direct sum of some V_f.

Assume there exists a complete flag of $\mathbf{K}$-vector spaces, each invariant by ϕ, $V_0 = \mathbf{K} \subset V_1 \subset \cdots \subset V_{d-1} = \mathbf{L}$, where V_k has dimension k. Then all irreducible factors of $X^d - 1$ must have degree 1. So $\mathbf{K}$ contains primitive roots of unity and we are in the context of Kummer theory. To every Galois invariant flag, there corresponds an order on d-th roots of unity (or equivalently on the associated characteristic spaces in $\mathbf{L}$). There are $d!$ such flags.

The flags produced by Kummer theory are of the following form:

$$V_1 \subset V_1 \oplus V_\zeta \subset V_1 \oplus V_\zeta \oplus V_{\zeta^2} \subset \ldots$$
$$\subset V_1 \oplus V_\zeta \oplus V_{\zeta^2} \oplus \cdots \oplus V_{\zeta^{d-2}} \subset V_1 \oplus V_\zeta \oplus V_{\zeta^2} \oplus \cdots \oplus V_{\zeta^{d-2}} \oplus V_{\zeta^{d-1}}$$

where ζ is a primitive d-th root of unity and V_ζ is $V_{X-\zeta}$, the eigenspace associated to ζ.

Among the $d!$ flags that are ϕ-invariants, only $\varphi(d)$ come from Kummer theory. They correspond to the $\varphi(d)$ primitive roots of unity. These flags enjoy a multiplicative property: If $k \geq 0$ and $l \geq 0$ and $k + l \leq d - 1$, then $V_k \times V_l \subset V_{k+l}$.

The conclusion of this section is thus rather negative. If we want to go further than Kummer theory, we cannot ask for Galois invariant flags of vector subspaces.

6. Specializing isogenies between commutative algebraic groups

Kummer and Artin-Schreier theories are two special cases of a general situation that we now describe. Our aim is to produce nice models for a broader variety of finite fields.

Let $\mathbf{K}$ be a field and $\mathbf{G}$ a commutative algebraic group over $\mathbf{K}$. Let $T \subset \mathbf{G}(\mathbf{K})$ be a non trivial finite group of $\mathbf{K}$-rational points in $\mathbf{G}$ and let

$$I : \mathbf{G} \to \mathbf{H}$$

be the quotient isogeny of $\mathbf{G}$ by T. Let $d \geq 2$ be the cardinality of T which is also the degree of I. Assume there exists a $\mathbf{K}$-rational point a on $\mathbf{H}$ such that $I^{-1}(a)$ is irreducible over $\mathbf{K}$. Then every point $b \in \mathbf{G}(\bar{\mathbf{K}})$ such that $I(b) = a$ defines a cyclic degree d extension $\mathbf{L}$ of $\mathbf{K}$: We set $\mathbf{L} = \mathbf{K}(b)$ and we notice that the geometric origin of this extension results in a nice description of $\mathbf{K}$-automorphisms of $\mathbf{L}$.

Let t be a point in T and let $\oplus_{\mathbf{G}}$ stand for the addition law in the algebraic group $\mathbf{G}$. Let $\oplus_{\mathbf{H}}$ stand for the addition law in $\mathbf{H}$. We denote by $0_{\mathbf{G}}$ the unit element in $\mathbf{G}$ and $0_{\mathbf{H}}$ the one in $\mathbf{H}$. The point $t \oplus_{\mathbf{G}} b$ verifies

$$I(t \oplus_{\mathbf{G}} b) = I(t) \oplus_{\mathbf{H}} I(b) = 0_{\mathbf{H}} \oplus_{\mathbf{H}} a = a.$$

So $t \oplus_{\mathbf{G}} b$ is Galois conjugated to b and all conjugates are obtained that way from all points t in T. So we have an isomorphism between T and $\mathrm{Gal}(\mathbf{L}/\mathbf{K})$, which associates to every $t \in T$ the residual automorphism

$$b \in I^{-1}(a) \mapsto b \oplus_{\mathbf{G}} t.$$

Now, assuming the geometric formulae that describe the translation $P \mapsto P \oplus_{\mathbf{G}} t$ in $\mathbf{G}$ are simple enough, we obtain a nice description of the Galois group of $\mathbf{L}$ over $\mathbf{K}$.

Kummer and Artin-Schreier theories provide two illustrations of this general geometric situation.

The algebraic group underlying Kummer theory is the multiplicative group $\mathbf{G}_m$ over the base field $\mathbf{K}$. The isogeny I is the multiplication by d:

$$I = [d] : \mathbf{G}_m \to \mathbf{G}_m.$$

One can see the group $\mathbf{G}_m$ as a sub-variety of the affine line $\mathbb{A}^1$ with z-coordinate. The inequality $z \neq 0$ defines the open subset $\mathbf{G} \subset \mathbb{A}^1$. The origin $0_{\mathbf{G}}$ has coordinate $z(0_{\mathbf{G}}) = 1$. The group law is given by

$$z(P_1 \oplus_{\mathbf{G}_m} P_2) = z(P_1) \times z(P_2).$$

Here we have $\mathbf{H} = \mathbf{G} = \mathbf{G}_m$ and the isogeny I can be given in terms of the z-coordinates by

$$z(I(P)) = z(P)^d.$$

Points in the kernel of I have for z-coordinates the d-th roots of unity. The inverse image by I of a point P in $\mathbf{G}$ is made of d geometric points having for z-coordinates the d-th roots of $z(P)$. Translation by an element t of the kernel of I, $P \mapsto P \oplus_{\mathbf{G}_m} t$, can be expressed in terms of z-coordinates by

$$z(P \oplus_{\mathbf{G}_m} t) = z(P) \times \zeta$$

where $\zeta = z(t)$ is the d-th root of unity associated by z to the d-torsion point t.

As far as Artin-Schreier theory is concerned, the underlying algebraic group is the additive group $\mathbf{G}_a$ over the base field $\mathbf{K}$, identified with the affine line $\mathbb{A}^1$ over $\mathbf{K}$. A point P on $\mathbf{G}_a$ is given by its z-coordinate. The origin $0_{\mathbf{G}}$ has coordinate $z(0_{\mathbf{G}}) = 0$ and the group law is given by

$$z(P_1 \oplus_{\mathbf{G}_a} P_2) = z(P_1) + z(P_2).$$

The degree p isogeny I is $\wp : \mathbf{G}_a \to \mathbf{G}_a$, given in terms of z-coordinates by

$$z(\wp(P)) = z(P)^p - z(P).$$

Here again $\mathbf{H} = \mathbf{G}$. The z-coordinates of points in the kernel of $\wp$ are the elements of the prime field $\mathbb{F}_p$. The inverse image by I of a point P in $\mathbf{G}$ is made of p geometric points whose z-coordinates are the p roots of the equation $X^p - X = z(P)$. Translation by an element t in the kernel of I, $P \mapsto P \oplus_{\mathbf{G}_a} t$, can be expressed in terms of z-coordinates by

$$z(P \oplus_{\mathbf{G}_a} t) = z(P) + \tau \text{ where } \tau = z(t) \in \mathbb{F}_p \ .$$

7. A different example

We plan to apply the generalities in the previous section to various algebraic groups. We guess every commutative algebraic group may bring its contribution to the construction of Galois invariant smoothness basis. Since we look for simple translation formulae, we expect the simplest algebraic groups to be the most useful. We start with the most familiar algebraic groups (after $\mathbf{G}_m$ and $\mathbf{G}_a$): These are the dimension 1 tori. Let $\mathbf{K}$ be a field with characteristic different from 2 and let D be a non zero element in $\mathbf{K}$. Let $\mathbb{P}^1$ be the projective line with projective coordinates $[U, V]$. Let $u = \frac{U}{V}$ be the associated affine coordinate. We denote by $\mathbf{G}$ the open subset of $\mathbb{P}^1$ defined by the inequality

$$U^2 - DV^2 \neq 0.$$

To every point P of $\mathbf{G}$, we associate its u-coordinate, possibly infinite but distinct from $\sqrt{D}$ and $-\sqrt{D}$. The unit element in $\mathbf{G}$ is the point $0_{\mathbf{G}}$ with projective coordinates $[1, 0]$ and u-coordinate ∞. For $P_1 \neq 0_{\mathbf{G}}$ and $P_2 \neq 0_{\mathbf{G}}$, the addition law is given by

$$u(P_1 \oplus_{\mathbf{G}} P_2) = \frac{u(P_1)u(P_2) + D}{u(P_1) + u(P_2)} \text{ and } u(\ominus_{\mathbf{G}} P_1) = -u(P_1).$$

We now assume that $\mathbf{K} = \mathbb{F}_q$ is a finite field and $D \in \mathbb{F}_q^*$ is not a square in $\mathbb{F}_q$. The group $\mathbf{G}(\mathbb{F}_q)$ has order $q + 1$ and the corresponding values of u lie in $\mathbb{F}_q \cup \{\infty\}$. The Frobenius endomorphism, $\phi : \mathbf{G} \to \mathbf{G}$, $[U, V] \to [U^q, V^q]$, is nothing but multiplication by $-q$. Indeed, let P be a point with projective coordinates $[U, V]$. The projective coordinates of $R = [q]P$ are the coordinates in $(1, \sqrt{D})$ of

$$(U + V\sqrt{D})^q = U^q - \sqrt{D}V^q$$

because D is not a square in $\mathbb{F}_q$. So R has coordinates $[U^q, -V^q]$ and it is the inverse of $\phi(P)$.

We pick an integer $d \geq 2$ such that the d-torsion $\mathbf{G}[d]$ is $\mathbb{F}_q$-rational. This is equivalent to the condition that d divides $q+1$. We set $q+1 = md$. Let I be the multiplication by d isogeny, $I = [d] : \mathbf{G} \to \mathbf{G}$, with kernel the cyclic group $\mathbf{G}[d]$ of order d. The quotient $\mathbf{G}(\mathbb{F}_q)/I(\mathbf{G}(\mathbb{F}_q)) = \mathbf{G}(\mathbb{F}_q)/\mathbf{G}(\mathbb{F}_q)^d$ is cyclic of order d.

Let a be a generator of $\mathbf{G}(\mathbb{F}_q)$ and b a geometric point in the fiber of I above a. Let $u(b)$ be the u-coordinate of b and set $\mathbf{L} = \mathbf{K}(u(b))$. This is a degree d extension of $\mathbf{K} = \mathbb{F}_q$. So $\mathbf{L} = \mathbb{F}_{q^d}$.

The Galois group of $\mathbb{F}_{q^d}/\mathbb{F}_q$ is isomorphic to $\mathbf{G}[d]$: For any $\mathfrak{a} \in \mathrm{Gal}(\mathbb{F}_{q^d}/\mathbb{F}_q)$, the difference $\mathfrak{a}(b) \ominus_{\mathbf{G}} b$ is in $\mathbf{G}[d]$ and the pairing

$$(\mathfrak{a}, a) \mapsto \mathfrak{a}(b) \ominus_{\mathbf{G}} b$$

defines an isomorphism of $\mathrm{Gal}(\mathbb{F}_{q^d}/\mathbb{F}_q)$ onto $\mathrm{Hom}(\mathbf{G}(\mathbb{F}_q)/(\mathbf{G}(\mathbb{F}_q))^d, \mathbf{G}[d])$.

Here $\mathrm{Gal}(\mathbb{F}_{q^d}/\mathbb{F}_q)$ is cyclic of order d generated by the Frobenius ϕ. The pairing (ϕ, a) equals $\phi(b) \ominus_{\mathbf{G}} b$. Remember that $\phi(b) = [-q]b$ in $\mathbf{G}$. So

$$(\phi, a) = [-q-1]b = [-m]a. \tag{2}$$

We obtain an exact description of Galois action on $I^{-1}(a)$. It is given by translations of the form $P \mapsto P \oplus_{\mathbf{G}} t$ with $t \in \mathbf{G}[d]$. If we denote by τ the affine coordinate of t and by u the coordinate of P then the action is given by

$$u \mapsto \frac{\tau u + D}{u + \tau},$$

which is rather nice since it is a rational linear transform.

We form the polynomial

$$A(X) = \prod_{b \in I^{-1}(a)} (X - u(b))$$

annihilating the u-coordinates of points in the inverse image of a by I. This is a degree d polynomial with coefficients in $\mathbf{K} = \mathbb{F}_q$. It is irreducible in $\mathbb{F}_q[X]$ because a generates $\mathbf{G}(\mathbb{F}_q)$. We have $\mathbf{L} = \mathbf{K}[X]/(A(X)) = \mathbb{F}_{q^d}$.

The exponentiation formulae in $\mathbf{G}$ give the explicit form of $A(X)$. One has

$$(U + \sqrt{D}V)^d =$$

$$\sum_{0 \le 2k \le d} \binom{d}{2k} U^{d-2k} V^{2k} D^k + \sqrt{D} \sum_{1 \le 2k+1 \le d} \binom{d}{2k+1} U^{d-2k-1} V^{2k+1} D^k.$$

So,

$$u([d]P) = \frac{\sum_{0 \le 2k \le d} u(P)^{d-2k} \binom{d}{2k} D^k}{\sum_{1 \le 2k+1 \le d} u(P)^{d-2k-1} \binom{d}{2k+1} D^k}.$$

And

$$A(X) = \sum_{0 \le 2k \le d} X^{d-2k} \binom{d}{2k} D^k - u(a) \sum_{1 \le 2k+1 \le d} X^{d-2k-1} \binom{d}{2k+1} D^k.$$

We set $x = X \bmod A(X)$. Since every $\mathbb{F}_q$-automorphism of $\mathbb{F}_{q^d}$ transforms x into a linear rational fraction of x, it is natural to define for every integer k such that $k \geq 0$ and $k < d$ the subset

$$V_k = \{\frac{u_0 + u_1 x + u_2 x^2 + \cdots + u_k x^k}{v_0 + v_1 x + v_2 x^2 + \cdots + v_k x^k} \mid (u_0, \ldots, u_k, v_0, \ldots, v_k) \in \mathbf{K}^{2k+2}\} .$$

One has $\mathbb{F}_q = V_0 \subset V_1 \subset \cdots \subset V_{d-1} = \mathbb{F}_{q^d}$ and the V_k are Galois invariant. Further, it is clear that $V_k \times V_l \subset V_{k+l}$ provided $k + l \leq d - 1$. Again we find a flag of Galois invariant subsets of $\mathbf{L} = \mathbb{F}_{q^d}$. But these subsets are no longer vector spaces.

If we define the degree of an element of $\mathbf{L}$ to be the smallest integer k such that V_k contains this element, then the degree is Galois invariant and sub-additive, $\deg(wz) \leq \deg(w) + \deg(z)$. The degree this times takes values between 0 and $\lceil \frac{d-1}{2} \rceil$. It is a slightly less informative function than in the Kummer or Artin-Schreier cases (it takes twice less values).

Example. Take $p = q = 13$ and $d = 7$. So $m = 2$. Let $D = 2$ which is not a square in $\mathbb{F}_{13}$. We look for some $a = U + \sqrt{2}V$ such that $U^2 - 2V^2 = 1$ and a has order $p + 1 = 14$ in $\mathbb{F}_{13}(\sqrt{2})^*$. For example $U = 3$ and $V = 2$ are fine. The u-coordinate of $3 + 2\sqrt{2}$ is $u(a) = \frac{3}{2} = 8$. One can write the polynomial

$$A(X) = X^7 + 3X^5 + 10X^3 + 4X - 8(7X^6 + 5X^4 + 6X^2 + 8).$$

Formula (2) predicts the Frobenius action. We set $t = [-m]a = [-2]a$ so $u(t) = 4$ and Frobenius operates by translation by t, so $X^p = \frac{4X+2}{X+4} \bmod A(X)$.

So we have made a small progress: We can now treat extensions of $\mathbb{F}_q$ of degree dividing $q + 1$. Unfortunately this condition is just as restrictive (though different) as the one imposed by Kummer theory. What do we do if the degree does not divide $q + 1$ nor $q - 1$?

We must diversify the algebraic groups we use. The next family to consider is made of elliptic curves.

8. Residue fields of divisors on elliptic curves

We now specialize the computations in section 6 to the case where $\mathbf{G}$ is an elliptic curve. Take $\mathbf{K} = \mathbb{F}_q$ a finite field for which we want to construct a degree $d \geq 2$ extension where d is prime to the characteristic p of $\mathbb{F}_q$. Here $\mathbf{G} = E$ is an ordinary elliptic curve over $\mathbb{F}_q$. We denote by ϕ the Frobenius

endomorphism of E. Let $\mathfrak{i}$ be an invertible ideal in the endomorphism ring $\mathrm{End}(E)$ of E. Assume $\mathfrak{i}$ divides $\phi - 1$ and $\mathrm{End}(E)/\mathfrak{i}$ is cyclic of order $d \geq 2$. So $E(\mathbb{F}_q)$ contains a cyclic subgroup $T = \mathrm{Ker}\,\mathfrak{i}$ of order d.

Let $I : E \to F$ be the degree d cyclic isogeny with kernel T. The quotient $F(\mathbb{F}_q)/I(E(\mathbb{F}_q))$ is isomorphic to T. Take a in $F(\mathbb{F}_q)$ such that $a \bmod I(E(\mathbb{F}_q))$ generates this quotient. The fiber $I^{-1}(a)$ is an irreducible divisor. This means that the d geometric points above a are defined on a degree d extension $\mathbf{L}$ of $\mathbf{K}$ and permuted by Galois action. We denote by $B = I^{-1}(a)$ the corresponding prime divisor.

Since $\mathbf{L}$ is the residue extension of E at B, we can represent elements of $\mathbf{L}$ in the following way: If f is a function on E with polar divisor disjoint to B, we denote by $f \bmod B \in \mathbf{L}$ the residue of f at B.

For f a function in $\mathbb{F}_q(E)$, the degree of f is the number of poles of f counted with multiplicities. For every $k \geq 0$ we call $\mathcal{F}_k$ the set of degree $\leq k$ functions in $\mathbb{F}_q(E)$, having no pole at B. We denote by V_k the corresponding set of residues in $\mathbf{L}$,

$$V_k = \{f \bmod B | f \in \mathcal{F}_k\}.$$

We have $V_0 = V_1 = \mathbf{K} \subset V_2 \subset \cdots \subset V_d = \mathbf{L}$ (Riemann-Roch) and $V_k \times V_l \subset V_{k+l}$. It is clear also that $\mathcal{F}_k$ is Galois invariant since composition by a translation from T does not affect the degree of a function. Therefore V_k is invariant under the action of $\mathrm{Gal}(\mathbf{L}/\mathbf{K})$.

If we want to test whether an element z of $\mathbf{L}$ is in V_k, we look for a function f in $\mathcal{F}_k$ such that $f = z \pmod B$. This is an interpolation problem which is hardly more difficult than in the two previous cases (polynomials for Kummer and rational fractions for the torus). We look for f as a quotient of two homogeneous forms of degree $\lceil \frac{k+1}{3} \rceil$, which can be done with linear algebra.

One can choose a smoothness basis consisting of all elements $f \bmod B$ in V_κ for a given κ. Factoring an element $z = f \bmod B$ of $\mathbf{L}$ boils down to factoring the divisor of f as a sum of prime divisors of degree $\leq \kappa$.

What conditions are sufficient for an elliptic curve to exist with all the required properties ? Since the number of $\mathbb{F}_q$-rational points on the elliptic curve is divisible by d, the size q of the field cannot be too small, that is

$$q + 2\sqrt{q} + 1 > d.$$

Assume d is odd and there exists a squarefree multiple D of d such that $D \not\equiv 1 \bmod p$ and

$$q + 1 - 2\sqrt{q} < D < q + 1 + 2\sqrt{q}.$$

There exists an ordinary elliptic curve E over $\mathbb{F}_q$ having D rational points over $\mathbb{F}_q$ and trace $t = q+1-D$. The ring $\mathbb{Z}[\phi]$ is integrally closed locally at every odd prime dividing D. The larger ring $\mathrm{End}(E)$ has the same property. The ideal $(\phi - 1)$ of $\mathrm{End}(E)$ has a unique degree d factor $\mathfrak{i}$. The quotient $\mathrm{End}(E)/\mathfrak{i}$ is cyclic and $\mathfrak{i}$ is invertible in $\mathrm{End}(E)$.

Given q and ϕ (a quadratic integer) as above, one can find an elliptic curve $E/\mathbb{F}_q$ by exhaustive search or using complex multiplication theory.

Example. Let $p = q = 11$, and $d = D = 7$, so $t = 5$ and $\phi^2 - 5\phi + 11 = 0$. The elliptic curve E with equation $y^2 + xy = x^3 + 2x + 8$ has complex multiplication by $\mathbb{Z}[\frac{\sqrt{-19}+1}{2}]$. The discriminant of $\mathbb{Z}[\phi]$ is -19, so $\mathrm{End}(E) = \mathbb{Z}[\phi]$. The ideal $\mathfrak{i} = (\phi - 1)$ is invertible and its kernel T is the full group of $\mathbb{F}_q$-rational points on E. The kernel of the degree 7 isogeny $I : E \to F$ is the group of rational points on E and for any non zero $a \in F(\mathbb{F}_{11})$, the fiber $B = I^{-1}(a)$ is irreducible.

9. Sieving algorithms and surfaces

There exists a family of algorithms for factoring integers and computing discrete logarithms that rely on intersection theory on algebraic or arithmetic surfaces. These algorithms are known as *the number field sieve*, *the function field sieve*, etc. The core of these algorithms is illustrated on the front page of the book [5]. In this section, we present the ideas underlying this family of algorithms in a rather general setting. This will help us to describe our construction in the next section 10. The sieving algorithm invented by Joux and Lercier in [2] for computing discrete logarithms will serve as a nice illustration for these ideas.

Let $\mathbb{F}_p$ be the field with p elements where p is prime. Let $\mathcal{S}$ be a smooth projective reduced, absolutely irreducible surface over $\mathbb{F}_p$. Let $\mathcal{A}$ and $\mathcal{B}$ be two absolutely irreducible curves on $\mathcal{S}$. Let $\mathcal{I}$ be an irreducible sub-variety of the intersection $\mathcal{A} \cap \mathcal{B}$. We assume that $\mathcal{A}$ and $\mathcal{B}$ meet transversely at $\mathcal{I}$ and we denote by d the degree of $\mathcal{I}$. The residue field of $\mathcal{I}$ is $\mathbb{F}_p(\mathcal{I}) = \mathbb{F}_q$ with $q = p^d$.

We need a pencil (linear or at least algebraic and connected) of effective divisors on $\mathcal{S}$. We denote it by $(D_\lambda)_{\lambda \in \Lambda}$ where Λ is the parameter space.

We fix an integer κ and we look (at random) for divisors D_λ in the pencil, such that both intersection divisors $D \cap \mathcal{A}$ and $D \cap \mathcal{B}$ are disjoint to $\mathcal{I}$ and κ-smooth (they split as sums of effective $\mathbb{F}_q$-divisors of degree $\leq \kappa$).

We define an equivalence relation $\equiv_{\mathcal{I}}$ on the set of divisors on $\mathcal{S}$ not meeting $\mathcal{I}$: We say $D \equiv_{\mathcal{I}} 0$ if and only if D is the divisor of a function f

and f is constant modulo $\mathcal{I}$. The equivalence classes for this relation are parameterized by points in some algebraic group denoted $\mathrm{Pic}(\mathcal{S},\mathcal{I})$. This algebraic group is an extension of $\mathrm{Pic}(\mathcal{S})$ by a torus $T_{\mathcal{I}}$ of dimension $d-1$.

One similarly defines the algebraic groups $\mathrm{Pic}(\mathcal{A},\mathcal{I})$ and $\mathrm{Pic}(\mathcal{B},\mathcal{I})$. These are generalized jacobians of $\mathcal{A}$ and $\mathcal{B}$ respectively. The natural (restriction) morphisms $\mathrm{Pic}(\mathcal{S},\mathcal{I}) \to \mathrm{Pic}(\mathcal{A},\mathcal{I})$ and $\mathrm{Pic}(\mathcal{S},\mathcal{I}) \to \mathrm{Pic}(\mathcal{B},\mathcal{I})$ induce the identity on the torus $T_{\mathcal{I}}$.

Let N be an integer that kills the three groups $\mathrm{Pic}^0(\mathcal{S})(\mathbb{F}_p)$, $\mathrm{Pic}^0(\mathcal{A})(\mathbb{F}_p)$, and $\mathrm{Pic}^0(\mathcal{B})(\mathbb{F}_p)$. Let λ and μ be two parameters in Λ corresponding to the divisors D_λ and D_μ in our pencil. We assume that $D_\lambda \cap \mathcal{A}$, $D_\mu \cap \mathcal{A}$, $D_\lambda \cap \mathcal{B}$, and $D_\mu \cap \mathcal{B}$ are smooth and disjoint to $\mathcal{I}$.

Let $D_\lambda \cap \mathcal{A} = \sum \mathfrak{a}_i$, $D_\mu \cap \mathcal{A} = \sum \mathfrak{b}_j$, $D_\lambda \cap \mathcal{B} = \sum \mathfrak{c}_k$ and $D_\mu \cap \mathcal{B} = \sum \mathfrak{d}_l$ be decompositions as sums of effective divisors on $\mathcal{A}$ and $\mathcal{B}$ with degree $\leq \kappa$. The divisor $D_\lambda - D_\mu$ is algebraically equivalent to zero and $N(D_\lambda - D_\mu)$ is principal.

Let f be a function on $\mathcal{S}$ with divisor $N(D_\lambda - D_\mu)$. We fix a smooth divisor X on $\mathcal{A}$ (resp. Y on $\mathcal{B}$) with degree 1. For every i and j, let α_i and β_j be functions on $\mathcal{A}$ with divisors $N(\mathfrak{a}_i - \deg(\mathfrak{a}_i)X)$ and $N(\mathfrak{b}_j - \deg(\mathfrak{b}_j)X)$. Similarly, for every k and l, let γ_k and δ_l be functions on $\mathcal{B}$ with divisors $N(\mathfrak{c}_k - \deg(\mathfrak{c}_k)Y)$ and $N(\mathfrak{d}_l - \deg(\mathfrak{d}_l)Y)$. There exist two multiplicative constant c and c' in $\mathbb{F}_p^*$ such that

$$f \equiv c.\frac{\prod_i \alpha_i}{\prod_j \beta_j} \equiv c'.\frac{\prod_k \gamma_k}{\prod_l \delta_l} \bmod \mathcal{I}.$$

This congruence can be regarded as a relation in the group $T_{\mathcal{I}}(\mathbb{F}_p) = \mathbb{F}_q^*/\mathbb{F}_p^*$. The factors in the first fraction belong to the smoothness basis on the $\mathcal{A}$ side: They are residues modulo $\mathcal{I}$ of functions on $\mathcal{A}$ with degree $\leq \kappa$. Similarly, the factors in the second fraction belong to the smoothness basis on the $\mathcal{B}$ side: They are residue modulo $\mathcal{I}$ of functions on $\mathcal{B}$ with degree $\leq \kappa$.

Joux and Lercier take $\mathcal{S}/\mathbb{F}_p$ to be $\mathcal{S} = \mathbb{P}^1 \times \mathbb{P}^1$ the product of $\mathbb{P}^1$ with itself over $\mathbb{F}_p$. To avoid any confusion we call $\mathcal{C}_1 = \mathbb{P}^1/\mathbb{F}_p$ the first factor and $\mathcal{C}_2 = \mathbb{P}^1/\mathbb{F}_p$ the second factor. Let O_1 be a rational point on $\mathcal{C}_1$ and $\mathcal{U}_1 = \mathcal{C}_1 - O_1$. Let x be an affine coordinate on $\mathcal{U}_1 \sim \mathbb{A}^1$. We similarly choose $O_2, \mathcal{U}_2$ and y an affine coordinate on $\mathcal{U}_2$.

They choose $\mathcal{A}$ to be the Zariski closure in $\mathcal{S}$ of the curve in $\mathcal{U}_1 \times \mathcal{U}_2$ with equation $y = f(x)$ where f is a polynomial with degree d_f in $\mathbb{F}_p[x]$. As for $\mathcal{B}$, they choose the Zariski closure in $\mathcal{S}$ of the curve with equation $x = g(y)$ where g is a polynomial with degree d_g in $\mathbb{F}_p[y]$.

The Néron-Severi group of a product of two smooth algebraically irreducible projective curves is $\mathbb{Z}$ times $\mathbb{Z}$ times the group of homomorphisms between the jacobians of the two curves. See [9, Mumford's appendix to Chapter VI]. The Hurwitz formula for the intersection of two classes is also given in this appendix.

Here the Néron-Severi group of $\mathcal{S}$ is $\mathbb{Z} \times \mathbb{Z}$. The algebraic equivalence class of a divisor D is given as its bidegree $(d_x(D), d_y(D))$ where $d_x(D) = D.(\mathcal{C}_1 \times O_2)$ and $d_y(D) = D.(O_1 \times \mathcal{C}_2)$. The intersection form is given by the formula

$$D.E = d_x(E)d_y(D) + d_x(D)d_y(E).$$

The bidegree of $\mathcal{A}$ is $(d_f, 1)$ and the bidegree of $\mathcal{B}$ is $(1, d_g)$. So $\mathcal{A}.\mathcal{B} = 1 + d_f d_g$ and the intersection of $\mathcal{A}$ and $\mathcal{B}$ is made of the point $O_1 \times O_2$ and the $d_f d_g$ points of the form $(\alpha, f(\alpha))$ where α is one of the $d_f d_g$ roots of $g(f(x)) - x$.

Let $h(x)$ be a simple irreducible factor of the later polynomial and let d be its degree. We take $\mathcal{I}$ to be the zero dimensional and degree d corresponding variety. The residue field $\mathbb{F}_p(\mathcal{I})$ is finite of order q where $q = p^d$.

To finish with, we need a pencil of effective divisors $(D_\lambda)_{\lambda \in \Lambda}$ on $\mathcal{S}$. It is standard to take for Λ the set of polynomials λ in $\mathbb{F}_p[x, y]$ with given bidegree (u_x, u_y) where u_x and u_y are chosen according to p and q. The corresponding divisor D_λ to λ is the Zariski closure of the zero set of λ. It has bidegree (u_x, u_y) too.

We fix an integer κ and look for divisors D_λ such that the two intersection divisors $D_\lambda \cap \mathcal{A}$ and $D_\lambda \cap \mathcal{B}$ are disjoint to $\mathcal{I}$ and κ-smooth. For example, if $\lambda(x, y)$ is a polynomial in x and y, the intersection of D_λ and $\mathcal{A}$ has degree $d_f u_y + u_x$. Its affine part is given by the roots of the polynomial $\lambda(x, f(x)) = 0$. Similarly, the intersection of D_λ and $\mathcal{B}$ has degree $u_y + u_x d_g$. Its affine part is given by the roots of the polynomial $\lambda(g(y), y)) = 0$. Joux and Lercier explain how to choose u_x, u_y and κ according to p and d.

10. Finite residue fields on elliptic squares

In this section we try to conciliate the generic construction in section 9 and the ideas developed in section 8. We would like the automorphisms of $\mathbb{F}_p(\mathcal{I})$ to be induced by automorphisms of the surface $\mathcal{S}$. So let E be an ordinary elliptic curve over $\mathbb{F}_p$ and let $\mathfrak{i}$ be an invertible ideal in the endomorphism ring $\mathrm{End}(E)$. We assume that $\mathfrak{i}$ divides $\phi - 1$ and $\mathrm{End}(E)/\mathfrak{i}$ is cyclic of order $d \geq 2$. So $E(\mathbb{F}_q)$ contains a cyclic subgroup $T = \mathrm{Ker}\,\mathfrak{i}$ of order d. Let

$I : E \to F$ be the quotient by $\mathrm{Ker}\,\mathfrak{i}$ isogeny and let $J : F \to E$ be such that $\phi - 1 = J \circ I$.

We take for $\mathcal{S}$ the product $E \times E$ and to avoid any confusion, we call E_1 the first factor and E_2 the second factor. Let O_1 be the origin on E_1 and O_2 the origin on E_2.

We use again the description of the Néron-Severi group of a product of two curves as given in [9, Appendix to Chapter VI]. This time, the Néron-Severi group of $\mathcal{S}$ is $\mathbb{Z} \times \mathbb{Z} \times \mathrm{End}(E)$. The class (d_1, d_2, ξ) of a divisor D consists of the bidegree and the induced isogeny. More precisely, d_1 is the intersection degree of D and $E_1 \times O_2$, d_2 is the intersection degree of D and $O_1 \times E_2$, and ξ is the homomorphism from E_1 to E_2 induced by the correspondence associated with D.

Let α and β be two endomorphisms of E and let a and b be two $\mathbb{F}_p$-rational points on E. We take $\mathcal{A}$ to be the inverse image of a by the morphism from $E \times E$ to E that maps (P, Q) onto $\alpha(P) - Q$. Let $\mathcal{B}$ be the inverse image of b by the morphism from $E \times E$ onto E that sends (P, Q) onto $P - \beta(Q)$.

Assume $1 - \beta\alpha = \phi - 1$. The intersection of $\mathcal{A}$ and $\mathcal{B}$ consists of points (P, Q) such that $(\phi - 1)(P) = b - \beta(a)$ and $Q = \alpha(P) - a$.

We choose a and b such that there exists a point c in $F(\mathbb{F}_p)$ generating $F(\mathbb{F}_p)/I(E(\mathbb{F}_p))$ and satisfying $J(c) = b - \beta(a)$. Then the intersection between $\mathcal{A}$ and $\mathcal{B}$ contains an irreducible component $\mathcal{I}$ of degree d.

The class of $\mathcal{A}$ is $(\alpha\bar{\alpha}, 1, \alpha)$. Indeed, the first coordinate of this triple is the degree of the projection $\mathcal{A} \to E_2$ onto the second component, that is the number of solutions in P to $\alpha(P) = Q + a$ for generic Q. This is the degree $\alpha\bar{\alpha}$ of α. The second coordinate of this triple is the degree of the projection $\mathcal{A} \to E_1$ onto the first component, that is the number of solutions in Q to $Q = \alpha(P) - a$ for generic P. This is 1. The third coordinate is the morphism in $\mathrm{Hom}(E_1, E_2)$ induced by the correspondence $\mathcal{A}$. This is clearly α. In the same way, we prove that the class of $\mathcal{B}$ is $(1, \beta\bar{\beta}, \bar{\beta})$.

Now let D be a divisor on $\mathcal{S}$ and (d_1, d_2, ξ) its class in the Néron-Severi group. The intersection degree of D and $\mathcal{A}$ is thus

$$D.\mathcal{A} = d_1 + d_2\alpha\bar{\alpha} - \xi\bar{\alpha} - \bar{\xi}\alpha \tag{3}$$

and similarly

$$D.\mathcal{B} = d_1\beta\bar{\beta} + d_2 - \xi\bar{\beta} - \bar{\xi}\beta. \tag{4}$$

We are particularly interested in the case where α and β have norms of essentially the same size (that is the square root of the norm of $\phi - 2$). We

then obtain a similar behavior as the algorithm in section 9 with an extra advantage: The smoothness bases on both $\mathcal{A}$ and $\mathcal{B}$ are Galois invariant.

Indeed, let $f_{\mathcal{A}}$ be a function with degree $\leq \kappa$ on $\mathcal{A}$. A point on $\mathcal{A}$ is a couple (P,Q) with $Q = \alpha(P) - a$. So the projection on the first component $\Pi_1 : E_1 \times E_2 \to E_1$ is an isomorphism. There is a unique function f_1 on E_1 such that $f_{\mathcal{A}} = f_1 \circ \Pi_1$. Assume now that (P,Q) is in $\mathcal{I} \subset \mathcal{A}$. Then $f_\alpha(P,Q) = f_1(P)$ is an element of the smoothness basis on $\mathcal{A}$. We observe that $f_1(P)^p = f_1(\phi(P)) = f_1(P+t)$ where t is in the kernel T of $\mathfrak{i}$. So $f_1(P)^p$ is the value at P of $f_1 \circ \tau_t$ where $\tau_t : E_1 \to E_1$ is the translation by t. Since $f_1 \circ \tau_t$ and f_1 have the same degree, the value of $f_1 \circ \tau_t$ at P is again an element in the smoothness basis.

That way, one can divide by d the size of either smoothness basis on $\mathcal{A}$ and $\mathcal{B}$.

As in section 9 we need a pencil of divisors on $\mathcal{S}$ with small class in the Néron-Severi group. We choose small values for (d_1, d_2, ξ) that minimize the expressions in Eq. (3) and Eq. (4) under the three contraints $d_1 \geq 1$, $d_2 \geq 1$ and

$$d_1 d_2 \geq \xi\bar{\xi} + 1. \tag{5}$$

We look for effective divisors in the algebraic equivalence class $\mathfrak{c} = (d_1, d_2, \xi)$. Recall O_1 is the origin on E_1 and O_2 the origin on E_2. The graph $\mathcal{G} = \{(P,Q) | Q = -\xi(P)\}$ of $-\xi : E_1 \to E_2$ is a divisor in the class $(\xi\bar{\xi}, 1, -\xi)$. The divisor $\mathcal{H} = -\mathcal{G} + (d_1 + \xi\bar{\xi})O_1 \times E_2 + (d_2+1)E_1 \times O_2$ is in $\mathfrak{c}$. We compute the linear space

$$\mathcal{L}(-\mathcal{G} + (d_1 + \xi\bar{\xi})O_1 \times E_2 + (d_2+1)E_1 \times O_2)$$

using the (restriction) exact sequence

$$\begin{aligned} 0 \to \mathcal{L}_S(-\mathcal{G} + (d_1 + \xi\bar{\xi})O_1 \times E_2 + (d_2+1)E_1 \times O_2) \\ \to \mathcal{L}_{E_1}((d_1 + \xi\bar{\xi})O_1) \otimes \mathcal{L}_{E_2}((d_2+1)O_2) \to \mathcal{L}_{\mathcal{G}}(\Delta) \end{aligned}$$

where Δ is the divisor on $\mathcal{G}$ given by the intersection with

$$(d_1 + \xi\bar{\xi})O_1 \times E_2 + (d_2+1)E_1 \times O_2.$$

This divisor has degree $d_1 + \xi\bar{\xi} + (d_2+1)\xi\bar{\xi}$, so the dimension of the right hand term in the sequence above is equal to this number.

On the other hand, the middle term has dimension $(d_1 + \xi\bar{\xi})(d_2+1)$, that is strictly bigger than the dimension of the right hand term, because of Inequality (5). So the linear space on the left is non zero and the divisor class is effective. Inequality (5) is a sufficient condition for effectivity.

In practice, one computes a basis for $\mathcal{L}_{E_1}((d_1+\xi\bar{\xi})O_1)$ and a basis for $\mathcal{L}_{E_2}((d_2+1)O_2)$ and one multiplies the two basis (one takes all products of one element in the first basis with one element in the second basis.) This produces a basis for $\mathcal{L}_{E_1}((d_1+\xi\bar{\xi})O_1)\otimes\mathcal{L}_{E_2}((d_2+1)O_2)$.

One selects enough (more than $d_1+\xi\bar{\xi}+(d_2+1)\xi\bar{\xi}$) points $(A_i)_i$ on $\mathcal{G}$ and one evaluates all functions in the above basis at all these points. A linear algebra calculation produces a basis for the subspace of $\mathcal{L}_{E_1}((d_1+\xi\bar{\xi})O_1)\otimes\mathcal{L}_{E_2}((d_2+1)O_2)$ consisting of functions that vanish along $\mathcal{G}$. For every function ϕ in the later subspace, the divisor of zeroes of ϕ contains $\mathcal{G}$ and the difference $(\phi)_0-\mathcal{G}$ is an effective divisor in the linear equivalence class of $\mathcal{H}$.

We have thus constructed a complete linear equivalence class inside $\mathfrak{c}$. To find the other linear classes in $\mathfrak{c}$, we remind that $E\times E$ is isomorphic to its Picard variety. So it suffices to replace $\mathcal{H}$ in the above calculation by $\mathcal{H}+E_1\times Z_2-E_1\times O_2+Z_1\times E_2-O_1\times E_2$ where Z_1 and Z_2 run over $E_1(\mathbb{F}_p)$ and $E_2(\mathbb{F}_p)$ respectively.

11. Experiments

In this section, we give a practical example of the geometric construction of section 10. We perform a discrete logarithm computation in $\mathbb{F}_{61^{19}}$. In such a field, Joux and Lercier algorithm would handle a factor basis of irreducible polynomials of degree 2 over $\mathbb{F}_{61}$, in two variables. Such a factor basis would have about 3600 elements. It turns out that in this case we can reduce the factor basis to only 198 elements using the ideas given in the previous section.

Initialization phase. We set $p=61$ and consider the plane projective elliptic curve E over $\mathbb{F}_p$ with equation $Y^2Z=X^3+20XZ^2+21Z^3$. It is ordinary with trace $t=-14$. The ring generated by the Frobenius ϕ has discriminant -48. The full endomorphism ring of E is the maximal order in the field $\mathbb{Q}(\sqrt{-3})$.

Let β be the degree 3 endomorphism of E given by

$$\begin{aligned}\beta:\quad & E\to E\,,\\ (x\ :y\ :1)&\mapsto \left(\tfrac{20\,x^3+36\,x^2+35\,x+40}{(x+7)^2}\ :y\tfrac{58\,x^3+59\,x^2+12\,x+21}{(x+7)^3}\ :1\right).\end{aligned}$$

We check $\beta^2=-3$ and we fix an isomorphism between $\mathrm{End}(E)\otimes\mathbb{Q}$ and $\mathbb{Q}(\sqrt{-3})\subset\mathbb{C}$ by setting $\beta=\sqrt{-3}$. The Frobenius endomorphism is $\phi=-7+2\sqrt{-3}$.

Let α be the degree 4 endomorphism defined by $\alpha = 1 + \beta = 1 + \sqrt{-3}$. It can be given explicitly by

$$(x : y : 1) \mapsto \left(\frac{49\,x^4+28\,x^3+55\,x^2+53\,x+27}{(x+25)(x+27)^2} : y\frac{38\,x^5+37\,x^4+30\,x^3+49\,x^2+9\,x+46}{(x+25)^2(x+27)^3} : 1\right).$$

The endomorphism $I = 1 - \beta\alpha$ has degree 19 and divides $\phi - 1$. The kernel of I consists of the following 19 rational points,

$$\begin{aligned}\operatorname{Ker} I = \{&(0:1:0),(11:\pm 13:1),(14:\pm 19:1),(21:\pm 8:1),(35:\pm 15:1),\\ &(40:\pm 10:1),(41:\pm 10:1),(45:\pm 27:1),(48:\pm 2:1),(51:\pm 23:1)\}.\end{aligned}$$

Let $\mathcal{S} = E \times E$. We call $E_1 = E$ the first factor and $E_2 = E$ the second one. If P and Q are independent generic points on E, then (P, Q) is a generic point on $\mathcal{S}$. Let a on E be the point with coordinates $(52 : 24 : 1)$. Let $\mathcal{A} \subset \mathcal{S}$ be the curve with equation $\alpha(P) - Q = a$. Let b on E be the point with coordinates $(1 : 46 : 1)$. Let $\mathcal{B} \subset \mathcal{S}$ be the curve with equation $P - \beta(Q) = b$. The numerical class of $\mathcal{A}$ is $(4, 1, 1+\sqrt{-3})$ and the numerical class of $\mathcal{B}$ is $(1, 3, -\sqrt{-3})$. Note that $b - \beta(a) = (57 : 11 : 1)$ is of order 38 and generates $E(\mathbb{F}_p)$ modulo the image of I.

Call $\mathcal{I}$ the intersection $\mathcal{A} \cap \mathcal{B}$. It consists of points (P, Q) such that $(1 - \beta\alpha)(P) = b - \beta(a)$, $Q = \alpha(P) - a$ and thus $(\alpha\beta - 1)(Q) = a - \alpha(b)$. In terms of the affine coordinates (x_1, y_1) of P and (x_2, y_2) of Q, this reads

$$x_1 = \frac{\left(44\,{x_2}^4 + 12\,{x_2}^3 + 9\,{x_2}^2 + 46\,x_2 + 40\right) y_2}{{x_2}^6 + 34\,{x_2}^5 + 41\,{x_2}^4 + 47\,{x_2}^3 + 7\,{x_2}^2 + 14\,x_2 + 58} + \frac{{x_2}^6 + 26\,{x_2}^5 + 25\,{x_2}^3 + 41\,{x_2}^2 + 19\,x_2 + 6}{{x_2}^6 + 34\,{x_2}^5 + 41\,{x_2}^4 + 47\,{x_2}^3 + 7\,{x_2}^2 + 14\,x_2 + 58}, \quad (6)$$

$$y_1 = \frac{\left(11\,{x_2}^7 + 2\,{x_2}^6 + 50\,{x_2}^5 + 59\,{x_2}^4 + 57\,{x_2}^3 + 30\,{x_2}^2 + 4\,x_2 + 14\right) y_2}{{x_2}^9 + 51\,{x_2}^8 + 7\,{x_2}^7 + 32\,{x_2}^6 + 56\,{x_2}^5 + 48\,{x_2}^4 + 26\,{x_2}^3 + 49\,{x_2}^2 + 18\,x_2 + 41} + \frac{46\,{x_2}^9 + 54\,{x_2}^8 + 2\,{x_2}^7 + 4\,{x_2}^6 + 52\,{x_2}^5 + 17\,{x_2}^4 + 60\,{x_2}^3 + 41\,{x_2}^2 + 48\,x_2 + 21}{{x_2}^9 + 51\,{x_2}^8 + 7\,{x_2}^7 + 32\,{x_2}^6 + 56\,{x_2}^5 + 48\,{x_2}^4 + 26\,{x_2}^3 + 49\,{x_2}^2 + 18\,x_2 + 41}, \quad (7)$$

or alternatively, x_2, y_2 can be given as functions of degree 8 and degree 12 in x_1, y_1.

The projection of $\mathcal{I}$ on E_1 (resp. E_2) yields a place $\mathcal{P}$ (resp. $\mathcal{Q}$) of degree 19 defined in the affine coordinates (x, y) by the equations

$$\begin{aligned}\mathcal{P} = &\\ &({x_1}^{19}+60\,{x_1}^{18}+25\,{x_1}^{17}+21\,{x_1}^{16}+23\,{x_1}^{15}+22\,{x_1}^{14}+49\,{x_1}^{13}+38\,{x_1}^{12}+30\,{x_1}^{11}+57\,{x_1}^{10}+\\ &\quad 3\,{x_1}^9 + 15\,{x_1}^8 + 26\,{x_1}^7 + 17\,{x_1}^6 + 45\,{x_1}^5 + 30\,{x_1}^4 + 48\,{x_1}^3 + 55\,{x_1}^2 + 18\,x_1 + 35,\\ &y_1 + 12\,{x_1}^{18} + 38\,{x_1}^{17} + 5\,{x_1}^{16} + {x_1}^{15} + 45\,{x_1}^{14} + 42\,{x_1}^{13} + 18\,{x_1}^{12} + 34\,{x_1}^{11} + 39\,{x_1}^{10} +\\ &\quad 59\,{x_1}^9 + 16\,{x_1}^8 + 18\,{x_1}^7 + 16\,{x_1}^6 + 36\,{x_1}^5 + 11\,{x_1}^4 + 9\,{x_1}^3 + 48\,{x_1}^2 + 59\,x_1 + 8),\end{aligned}$$

$$\mathcal{Q} = (x_2{}^{19}+25\,x_2{}^{18}+34\,x_2{}^{17}+46\,x_2{}^{16}+16\,x_2{}^{15}+14\,x_2{}^{14}+58\,x_2{}^{13}+52\,x_2{}^{12}+39\,x_2{}^{11}+48\,x_2{}^{10}+18\,x_2{}^{9}+56\,x_2{}^{8}+41\,x_2{}^{7}+40\,x_2{}^{6}+11\,x_2{}^{5}+33\,x_2{}^{4}+55\,x_2{}^{3}+14\,x_2{}^{2}+5\,x_2+56,\ y_2+42\,x_2{}^{18}+40\,x_2{}^{17}+23\,x_2{}^{16}+41\,x_2{}^{15}+14\,x_2{}^{14}+12\,x_2{}^{13}+30\,x_2{}^{12}+50\,x_2{}^{11}+33\,x_2{}^{10}+33\,x_2{}^{9}+60\,x_2{}^{8}+15\,x_2{}^{7}+54\,x_2{}^{6}+13\,x_2{}^{5}+17\,x_2{}^{4}+31\,x_2{}^{3}+50\,x_2{}^{2}+52\,x_2+3)\,.$$

The residue fields of these two places are isomorphic (both being degree 19 extensions of $\mathbb{F}_{61}$). We fix an isomorphism between these two residue fields by setting

$$x_2 \mapsto 2\,x_1{}^{18}+57\,x_1{}^{17}+21\,x_1{}^{16}+10\,x_1{}^{15}+54\,x_1{}^{14}+35\,x_1{}^{13}+45\,x_1{}^{12}+27\,x_1{}^{11}+41\,x_1{}^{10}+55\,x_1{}^{9}+27\,x_1{}^{8}+36\,x_1{}^{7}+29\,x_1{}^{6}+50\,x_1{}^{5}+44\,x_1{}^{4}+18\,x_1{}^{3}+38\,x_1{}^{2}+51\,x_1+18\,. \tag{8}$$

Fixing this isomorphism is equivalent to choosing a geometric point in $\mathcal{I}$.

Sieving phase. We are now going to look for "smooth" functions on $\mathcal{S}$. We first explain what we mean by smooth in this context. Let $\varepsilon(x_1, y_1, x_2, y_2)$ be a function on $\mathcal{S}$. We assume ε does not vanish at $\mathcal{I}$. Let $\Pi_1 : \mathcal{S} = E_1 \times E_2 \to E_1$ be the projection on the first factor. The restriction of Π_1 to $\mathcal{A}$ is a bijection. So we can define a point on $\mathcal{A}$ by its coordinates (x_1, y_1). Let $\Pi_2 : \mathcal{S} = E_1 \times E_2 \to E_2$ be the projection on the second factor. The restriction of Π_2 to $\mathcal{B}$ is a bijection. So we can define a point on $\mathcal{B}$ by its coordinates (x_2, y_2).

Let $\varepsilon_1(x_1, y_1)$ (resp. $\varepsilon_2(x_2, y_2)$) be the restriction of ε to $\mathcal{A}$ (resp. $\mathcal{B}$). For example $\varepsilon_2(x_2, y_2)$ is obtained by substituting x_1, y_1 as functions in x_2, y_2 in ε thanks to Eq. (6) and Eq. (7).

The function ε is said to be smooth if the divisors of ε_1 and ε_2 both contain only places of small degree κ. In our example, we choose $\kappa = 2$. Let us remark at this point that thanks to the isomorphism given by Eq. (8), the reduction modulo $\mathcal{P}$ of ε_1 is equal to the reduction modulo $\mathcal{Q}$ of ε_2, and this yields an equality in $\mathbb{F}_{61^{19}}$.

To every non-zero function on $\mathcal{S}$, one can associate a linear pencil of divisors. We define the linear (resp. numerical) class of the function to be the linear (resp. numerical) class of the divisor of its zeroes (or poles).

We shall be firstly interested in functions ε with numerical class $(1, 0, 0)$. An effective divisor in these classes is $c \times E_2$ where c is a place of degree 1 on E_1 and it is not difficult to see that the intersection degrees of such a divisor with $\mathcal{A}$ and $\mathcal{B}$ are 1 and 3. Functions with numerical class $(2, 0, 0)$ are obtained in the same way.

We found similarly functions ε in the class $(0, 1, 0)$, derived from divisors $E_1 \times c$. The intersection degrees are now 4 and 1. Functions with numerical

class $(0,2,0)$ are obtained in the same way too. More interesting, the class $(1,1,1)$ containing the divisors with equation $P = Q + c$, yields intersection degrees 3 and 4.

We finally consider the class $(2,2,1)$ which is, by far, much larger than the previous classes. The intersection degrees are 8 and 8. To enumerate functions in this class, we first build a basis for the linear space associated to divisors of degree 3 on both E_1 and E_2. For instance, let us consider $\mathcal{L}_{E_1}(3\,O_1)$ and $\mathcal{L}_{E_2}(3\,O_2)$, basis of which are given by $\{1, x_1, y_1\}$ and $\{1, x_2, y_2\}$. We then determinate that a basis for the subspace of $\mathcal{L}_{E_1}(3\,O_1) \otimes \mathcal{L}_{E_2}(3\,O_2)$, consisting of functions that vanish along the graph $\mathcal{G} = \{(P,Q), Q = -P\}$, is given by $\{y_1\,x_2 + x_1\,y_2, y_1 + y_2, x_1 - x_2\}$. An exhaustive enumeration of functions of the form $y_1\,x_2 + x_1\,y_2 + \lambda(y_1 + y_2) + \mu(x_1 - x_2)$, with $\lambda, \mu \in \mathbb{F}_p$ yields useful equations.

We give examples of such relations in Fig. 1 (page 165).

Linear algebra phase. With our smoothness choice, the factor basis is derived from places of degree one and two. Since we prefer functions to divisors, the factor basis will contain the reduction modulo $\mathcal{P}$, resp. $\mathcal{Q}$, of functions the divisors of which are equal to $76(x_1 + \alpha, y_1 + \beta) - 76(1/x_1, y_1/x_1^2)$, resp. $76(x_2 + \alpha, y_2 + \beta) - 76(1/x_2, y_2/x_2^2)$ (remember that in our example $\#E(\mathbb{F}_p) = 76$). In this setting, the evaluation at $\mathcal{P}$ or $\mathcal{Q}$ of any smooth function can be easily written as a product of elements of the factor basis.

It is worth recalling that the action of the Frobenius ϕ on the reduction of a function modulo $\mathcal{P}$ or $\mathcal{Q}$ is equal to the reduction of a function, the poles and the zeros of which are translated by one specific point of $\operatorname{Ker} I$. In our example, this point is $F_1 = (11 : 48 : 1)$ for the reduction modulo $\mathcal{P}$ and $F_2 = (45 : 34 : 1)$ for the reduction modulo $\mathcal{Q}$. For instance, let us consider a function g_0 the divisor of which is equal to $76(x_1 + 41, y_1 + 8) - 76(1/x_1, y_1/x_1^2)$. Let us now consider a function g_6 which corresponds to $(-41 : -8 : 1) + 6F_1$, that is a function with divisor equal to $76(x_1 + 45, y_1 + 17) - 76(1/x_1, y_1/x_1^2)$. We have then $\overline{g_6} = c.\overline{g_0}^{p^6}\overline{f}^{1+p+p^2+p^3+p^4+p^5}$ for some $c \in \mathbb{F}_p$, where f is a function the divisor of which is equal to $76\,F_1 - 76(1/x_1, y_1/x_1^2)$.

Thanks to this observation, we can thus divide by 19 the size of the factor basis, at the expense in the linear algebra phase of entries equal to sums of powers of p. We finally have 4 meaningful places of degree 1 and 92 meaningful places of degree 2 on each side, that is a total of 196 entries in our factor basis. Of course, under the Galois conjugations, most of the relations obtained in the sieving phase are redundant, but it does

Class	div ε_1	div ε_2
$(1,0,0)$	$(x_1+43, y_1+33)-(x_1+13, y_1+59)$	$(x_2^2+x_2+52, y_2+10\,x_2+37)+(x_2+12, y_2+35)-(x_2+2, y_2+20)-(x_2^2+26\,x_2+39, y_2+5\,x_2+27)$
$(2,0,0)$	$(x_1^2+56\,x_1+34, y_1+22\,x_1+52)-2\,(x_1+13, y_1+59)$	$(x_2^2+37\,x_2+53, y_2+42\,x_2+58)+(x_2^2+12\,x_2+19, y_2+52\,x_2+43)+(x_2^2+41\,x_2+29, y_2+33\,x_2+41)-2\,(x_2+2, y_2+20)-2\,(x_2^2+26\,x_2+39, y_2+5\,x_2+27)$
$(0,1,0)$	$(x_1^2+4\,x_1+12, y_1+55\,x_1+47)+(x_1^2+45\,x_1+31, y_1+19\,x_1+23)-(x_1+42, y_1+60)-(x_1+36, y_1+15)-(x_1^2+60\,x_1+25, y_1+36\,x_1+26)$	$(x_2+43, y_2+33)-(x_2+13, y_2+59)$
$(0,2,0)$	$(x_1^2+26\,x_1+12, y_1+12\,x_1+32)+(x_1^2+48\,x_1+6, y_1+59)+(x_1^2+53\,x_1+56, y_1+42\,x_1+56)+(x_1^2+3\,x_1+38, y_1+17\,x_1+36)-2\,(x_1+42, y_1+60)-2\,(x_1+36, y_1+15)-2\,(x_1^2+60\,x_1+25, y_1+36\,x_1+26)$	$(x_2^2+24\,x_2+39, y_2+37\,x_2+27)-2\,(x_2+13, y_2+59)$
$(1,1,1)$	$(x_1+2, y_1+41)+(x_1^2+26\,x_1+39, y_1+56\,x_1+34)-(x_1^2+48\,x_1+6, y_1+2)-(x_1+52, y_1+25)$	$(x_2+17, y_2+21)+(x_2^2+57\,x_2+11, y_2+33\,x_2)+(x_2+55, y_2+33)-(x_2^2+49\,x_2+42, y_2+26)-(x_2^2+3\,x_2+4, y_2+30\,x_2+20)$
$(2,2,2)$	$(x_1^2+25\,x_1+42, y_1+5\,x_1+13)+(x_1^2+30\,x_1+19, y_1+52\,x_1+42)+(x_1^2+59\,x_1+30, y_1+28\,x_1+22)-2\,(x_1^2+48\,x_1+6, y_1+2)-2\,(x_1+52, y_1+25)$	$(x_2^2+30\,x_2+21, y_2+50\,x_2+52)+(x_2^2+41\,x_2+8, y_2+54\,x_2+58)+(x_2^2+32\,x_2+20, y_2+34\,x_2+28)+(x_2^2+42\,x_2+49, y_2+29\,x_2+51)-2\,(x_2^2+49\,x_2+42, y_2+26)-2\,(x_2^2+3\,x_2+4, y_2+30\,x_2+20)$
$(2,2,1)$	$(x_1+24, y_1+33)+(x_1+25, y_1)+(x_1+35, y_1)+(x_1+60, y_1+46)+(x_1^2+33\,x_1+43, y_1+3\,x_1+34)+(x_1^2+53\,x_1+53, y_1+24\,x_1+33)-(x_1+1, y_1)-(x_1+54, y_1+4)-(x_1^2+17\,x_1+19, y_1+41\,x_1+21)-(x_1^2+51\,x_1+53, y_1+44\,x_1+31)-(x_1^2+55\,x_1+38, y_1+38\,x_1+58)$	$(x_2+3, y_2+42)+(x_2^2+7\,x_2+20, y_2+33\,x_2+46)+(x_2^2+38\,x_2+12, y_2+58\,x_2+6)+(x_2^2+42\,x_2+35, y_2+7\,x_2+41)-(x_2+1, y_2)-(x_2+11, y_2+42)-(x_2+16, y_2+34)-(x_2^2+26\,x_2+12, y_2+49\,x_2+29)-(x_2^2+47\,x_2+5, y_2+7\,x_2+14)$
$(2,2,1)$	$(x_1+10, y_1+23)+(x_1+20, y_1+x_1+30)+(x_1+29, y_1+1)+(x_1+41, y_1+x_1+33)+(x_1^2+6\,x_1+17, y_1+25\,x_1+16)+(x_1^2+25\,x_1+12, y_1+25\,x_1+47)-(x_1+1, y_1)-(x_1+54, y_1+4)-(x_1^2+17\,x_1+19, y_1+41\,x_1+21)-(x_1^2+51\,x_1+53, y_1+44\,x_1+31)-(x_1^2+55\,x_1+38, y_1+38\,x_1+58)$	$(x_2+29, y_2+60)+(x_2+36, y_2+15)+(x_2^2+15\,x_2+58, y_2+41\,x_2+39)+(x_2^2+23\,x_2+2, y_2+33\,x_2+7)+(x_2^2+44\,x_2+33, y_2+35\,x_2+28)-(x_2+1, y_2)-(x_2+11, y_2+42)-(x_2+16, y_2+34)-(x_2+50, y_2+13)-(x_2^2+26\,x_2+12, y_2+49\,x_2+29)-(x_2^2+47\,x_2+5, y_2+7\,x+14)$

Fig. 1. Some relations collected in the sieving phase.

not really matter since it is not difficult to reduce the sieving phase to the only meaningful relations.

We have

$$61^{19} - 1 = 2^2 \cdot 3 \cdot 5 \cdot 229 \cdot 607127818287731321660577427051.$$

We performed the linear algebra modulo the largest factor of $61^{19} - 1$, that is the 99-bit integer 607127818287731321660577427051. This gives us the discrete logarithm in basis $f \bmod \mathcal{I}$ of any element in the smoothness basis. For instance, if g is any function such that $\operatorname{div} g = 76\,(x_1^2 + 37\,x_1 + 54, y_1 + 41\,x_1 + 16) - 152(1/x_1, y_1/x_1^2)$, we find that

$$g^{2^2 \cdot 3 \cdot 5 \cdot 229} = (f^{2^2 \cdot 3 \cdot 5 \cdot 229})^{471821537021905592692223848756}.$$

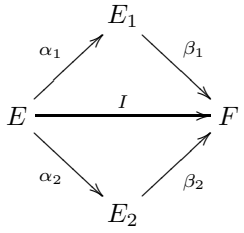
12. Generalization and limitations

The construction in section 10 can and should be generalized.

Let E be again an ordinary elliptic curve over $\mathbb{F}_p$ and let $\mathfrak{i}$ be an invertible ideal in the endomorphism ring $\operatorname{End}(E)$. We assume that $\mathfrak{i}$ divides $\phi - 1$ and $\operatorname{End}(E)/\mathfrak{i}$ is cyclic of order $d \geq 2$. Let F be the quotient of E by the kernel T of $\mathfrak{i}$ and $I : E \to F$ the quotient isogeny.

The integer d belongs to the ideal $\mathfrak{i}$. Let u and v be two elements in $\mathfrak{i}$ such that $d = u + v$ and $(u) = \mathfrak{i}\mathfrak{a}_1\mathfrak{b}_1$ and $(v) = \mathfrak{i}\mathfrak{a}_2\mathfrak{b}_2$ where $\mathfrak{a}_1$, $\mathfrak{b}_1$, $\mathfrak{a}_2$, $\mathfrak{b}_2$ are invertible ideals in $\operatorname{End}(E)$. We deduce the existence of two elliptic curves E_1 and E_2 and four isogenies α_1, β_1, α_2, β_2, such that $\beta_1\alpha_1 + \beta_2\alpha_2 = I$.

We represent all these isogenies on the (non commutative) diagram below.



We set $\mathcal{S} = E_1 \times E_2$. As for $\mathcal{A}$ we choose the image of $(\alpha_1, \alpha_2) : E \to \mathcal{S}$. And $\mathcal{B}$ is the inverse image of f by $\beta_1 + \beta_2 : \mathcal{S} \to F$ where f generates the quotient $F(\mathbb{F}_p)/I(E(\mathbb{F}_p))$. The intersection of $\mathcal{A}$ and $\mathcal{B}$ is the image by (α_1, α_2) of $I^{-1}(f) \subset E$. We choose u and v such that $\mathfrak{a}_1$, $\mathfrak{b}_1$, $\mathfrak{a}_2$, and $\mathfrak{b}_2$, have norms close to the square root of d.

This construction is useful when the norm of $\mathfrak{i}$ is much smaller than the norm of $\phi - 1$. So we managed to construct Galois invariant smoothness basis for a range of finite fields. Our constructions go beyond the classical Kummer case. They are efficient when the degree d is either below $4\sqrt{q}$ or in the interval $]q + 1 - 2\sqrt{q}, q + 1 + 2\sqrt{q}[$.

References

1. N. Bourbaki. *Algèbre, chapitre V*. Masson, 1981.
2. A. Joux and R. Lercier. The function field sieve is quite special. *Lecture Notes in Comput. Sci.*, 2369:431–445, 2002.
3. A. Joux and R. Lercier. The function field sieve in the medium prime case. *Lecture Notes in Comput. Sci.*, 4004:254–270, 2006.
4. S. Lang. *Algebra*. Addison-Wesley, 1984.
5. A.K. Lenstra and H.W. Lenstra. *The development of the number field sieve*. Number 1554 in Lecture Notes in Mathematics. Springer, 1993.
6. G. Malle and B.H. Matzat. *Inverse Galois Theory*. Springer, 1999.
7. A. M. Odlyzko. Discrete logarithms: The past and the future. *Designs, Codes, and Cryptography*, 19:129–145, 2000.
8. L. Thomas. *Arithmétique des extensions d'Artin-Schreier-Witt*. Thèse de l'Université Toulouse 2, 2006.
9. P. Zariski. *Algebraic surfaces, supplemented edition*. Springer, 1971.

Fuzzy pairings-based CL-PKC

Mikko Kiviharju

FDF Tech. Research Centre, P.B. 10
Riihimaki, FIN-11311,Finland
E-mail : mikko.kiviharju@mil.fi

Elliptic curves over finite fields are currently the most useful instantiation of bilinear mappings used extensively in todays public-key cryptography, especially in the Boneh-Franklin identity-based cryptosystem. Biometrics and identity-based cryptography (IBC) are a natural fit as they both aim to bind identity to security methods. However, biometrics are by nature fuzzy. For IB-schemes there exists a solution by Sahai and Waters (2004). Additionally, Al-Riyami and Paterson presented in 2003 a mixture of IBC and conventional PKI, called certificateless public-key cryptography (CL-PKC). In this paper we present, how CL-PKC schemes loan an important element to IBC and how fuzziness can be applied to this scheme as well. The construction uses the key-management architecture by Al-Riyami and Paterson to the polynomial-based approach of Sahai and Waters. The basic system consists of a supersymmetric elliptic curve group (over a sufficiently large finite field), with applied Shamir secret sharing. Encryption and decryption are performed as in Sahai-Waters scheme, with public elliptic curve group elements and one-time random exponents. The extension proposed in this paper guarantees that only the identity part of the encryption is fuzzy, whereas the public revocable part varies too much to allow an exact match. Applications of fuzzy CL-PKC (FC-PKC) include biometric systems with identity revocation. The purpose of the applications based on the construction presented in this paper is to layer the key-management architecture of a biometric system in such a way that the user is aware of the identity only, and the encryption module will contain the revocable part of the public key.

Keywords: identity-based cryptography, elliptic curve, certificateless cryptography, fuzzy cryptography.

1. Introduction

Identity-based cryptography (IBC) was introduced as early as 1984 [18] to tackle the problem of associating the identity to the public key. There were many incomplete schemes [10,13,17,18] that provided signature or key-agreement functionality. The main problem of creating a private key from

a given arbitrary public key, however, remained elusive until 2001, when Boneh and Franklin introduced in their seminal paper the first practical IBC-scheme, based on elliptic curve groups over finite fields.

Boneh and Franklins main contribution was to realize that the MOV-weakness [15] in certain EC groups meant that the decisional Diffie-Hellman problem (DDH) became easy but the computational Diffie-Hellman (CDH) still remained hard (as far as anyone knew), and that this could be exploited to achieve IBC.

The Boneh & Franklin IBC-scheme, and indeed many others, for example hierarchical schemes, consider a digital identity. In the real world, however, the real identity of the entity needs to be bound somehow to the digital identity. There is a large spectrum of access control and identity-management methodologies to cope with this binding. Generally they are divided into three: something you have, something you know and something you are. The first two categories can instantly be translated to a digital identity by tokens (i.e. smart cards) or information about the identity attributes (i.e. address, email), possibly augmented by passwords. However, the last category, which usually translates to biometrics, is not so straightforward to transform into a digital identity.

Biometrics have several attributes that seem unconventional to cryptological processing. Two main considerations are that they are fuzzy, and ultimately public. Fuzziness means that consecutive measurements do not produce identical results, but are nevertheless close to each other as measured in some metric space. The public nature of biometrics states that biometrics should not be used as private or secret keys lightly. The basic premise of biometrics is that they are hard to reproduce not their secrecy.

IBC and biometrics have some common areas that indicate they are a natural fit biometrics offers the natural key-storage with some wiggle room (fuzziness), whereas IBC offers the cryptological foundation to use that storage securely.

One way of dealing with fuzziness in IBC was presented by Sahai and Waters in 2004 [16]. It is a solution that implements fuzziness for bilinear-mappings-based IB-schemes: decrypting with one private key messages encrypted with two different (but not too different) public keys.

IBC has so far had two elusive open problems: randomizing the generation of the private key, and revoking ID-based public keys. These are interlinked such that if the private key generation could be randomized, it would be a simple matter to revoke the key as well. The dilemma is communicating this information to other users, and revoke the key as offline as

possible. There are two main types of solutions for revocation: push-type (revocation lists, protocol-based solutions and zero-knowledge proofs) and pull-type, where encryption and decryption are mediated, and the mediator is part of the key-management architecture, for example access control with reference monitors.

To help the revocation and dependency on a single trusted authority, Al-Riyami and Paterson presented in 2003 [1] an attractive mixture of IBC and conventional PKI, called certificateless public-key cryptography (CL-PKC). In this scheme the burden of carrying the trusted authority (CA in PKI, key-generation center or KGC here) information with the message is replaced with encrypting the KGC information into each message. This dispenses some of the PKI infrastructure, but not all of it. It also introduces other public pieces of information than just the identity, but this extra piece allows (basically) revocation and randomization.

Applications of fuzzy CL-PKCs (FC-PKC) include biometric systems with identity revocation. We will give a recommendation of layering the key-management architecture of such an application to achieve smooth Human-Computer Interaction. The key architecture should, in essence, contain the revocable part in a separate, user-specific secure device (usually a token), which can be replaced like an ordinary key.

1.1. *Our Contributions*

In this paper we combine and modify the two approaches in [1] and [16]. We show, how fuzziness can be applied to the CL-PKC like randomization technique as well. The construction uses the key-management architecture by Al-Riyami and Paterson combined with the polynomial-based approach of Sahai and Waters. The modification has two main parts:

- architectural modification to CL-PKC key management: we bind the randomized public key tighter to the identity with the help of a bilinear mapping function in order to avoid the denial-of-decryption attack mentioned by Susilo et at. in [14].
- randomization of the fuzzy IBC public and private key.

The basic system consists of a supersymmetric elliptic curve group (ECG) over a sufficiently large finite field, with applied Shamir secret sharing [19]: The ECG points are selected based on a coefficient calculated from a polynomial. This polynomial has a sufficiently high degree, such that it can be interpolated correctly at a point only if a sufficient number of points are provided. However, the degree of the polynomial is not as high as the

number points required to encode the identity string. This provides error-correction, or fuzziness.

The Al-Riyami-Paterson CL-PKC distributes responsibility of key-generation both to the key-management entity and user. We simplify this structure, since the goal is randomized key-generation more than distributed trust, and the latter can be achieved by other means as well. We note, however, that our construction does not limit the key generation places: the randomness can be added later by the user as well, as it will not require knowledge of the master key, only access to the secret and public keys.

In effect, our scheme does not attempt to dispense with the problem of key-escrow, but does address the problem of denial-of-decryption attack [14] apparent in the CL-PKC. This is accomplished by a simple change in the randomized public-key construction in the CL-PKC public key generation procedure, applied to the polynomial approach by Sahai and Waters.

The fuzzy IBC schemes user private-key generation is randomized, but this serves only to protect against collusion of several legitimate users to open an unauthorized message and does not play a part in key revocation (the polynomial can be changed, but any old polynomial issued for the identity will not stop working). It does not provide extra security for the master key either, since the encryption and decryption processes do not use naked master key anyway.

Encryption and decryption are performed as in Sahai-Waters scheme, with public elliptic curve group elements and one-time random exponents. The extension proposed in this paper guarantees that only the identity part of the encryption is fuzzy: the revocable part is not seen as points in the polynomial (as the identity is) but as attributes to the construction of the points.

The security of the system is based on a modified bilinear DDH (BLDDH), which in this case means that for an elliptic curve group (over a finite field) generator P, it is hard to distinguish between tuples

$$< aP, bP, cP, e(P,P)^{abc^{-1}} > \tag{1}$$

and

$$< aP, bP, cP, e(P,P)^{z} > \tag{2}$$

for a randomly chosen z. This is the same assumption as in the Sahai-Waters security model, since the construction follows closely the fuzzy IBC construction.

The scheme presented here is not without applications either: we will outline a two-factor encryption approach intended for biometric architectures with hierarchical IBC. In these architectures our basic work of setting fuzziness and revocable randomness into independent components becomes apparent, and the main motivation for the scheme.

1.2. *Related Work*

The problematics related to handling biometrics was nicely introduced in conjunction of randomness extractors for fuzzy samples in fuzzy extractors by Dodis in [7]. First practical identity-based cryptography schemes were given by Boneh, Franklin and Boyen in [3] and [4]. Excellent hierarchical schemes with IBC are for example those in [8].

Fuzzy approaches in IBC were given shortly after Boneh & Franklin gave their seminal result, by Sahai and Waters in [16], improved later by Baek, Susilo and Zhou in [2], and a biometric signature scheme was given by Burnett, Byrne, Dowling and Duffy in [5].

Revocation with IBC has been a topic of a number of IBC papers, but solutions are never quite satisfactory. The hierarchical work in [12] and the ticketing-based system in conjunction with CBIS-security [6] are good candidates for establishing revocation functionality in IBC.

As mentioned in [14], there are also related schemes to CL-PKC, such as self-certified public keys [10] certificate-based encryption [9], and self-generated certificates [14]. The schemes [10] and [9] are not exactly identity-based, and [14] proved to be less accessible to the Sahai-Waters approach of error-correction.

1.3. *Organization of the Paper*

This paper starts with an introduction listing our contributions and related work. Preliminaries are given in chapter 2, including definitions of bilinear maps and the relevance of elliptic curves to them, complexity assumptions and a short look to the scheme extended here: the fuzzy IBC. The main part of this paper is in chapter 3, where our FC-PKC scheme is constructed. We then discuss the security of the scheme in chapter 4, the applications and some future work in chapter 5. Chapter 6 concludes the paper.

2. Preliminaries

2.1. *Bilinear Maps and Elliptic Curves*

Identity-based cryptography schemes rely heavily on bilinear groups and functions. We assume here the use of elliptic curve groups and the modified bilinear pairing, as in [3], but essentially the only requirement for the scheme is an existing and efficient bilinear map. We review here shortly the necessary properties of the groups and maps.

Given two cyclic groups $\mathbb{G}_1$(additive) and $\mathbb{G}_2$(multiplicative) of prime order and a generator P for $\mathbb{G}_1$, we call $\mathbb{G}_1$ a bilinear group, if there exists a bilinear mapping $e : \mathbb{G}_1 \times \mathbb{G}_1 \longrightarrow \mathbb{G}_2$, such that

$$\forall(a, b \in \mathbb{Z} : e(aP, bP) = e(P, P)^{ab} \tag{3}$$

and

$$e(P, P) \neq 1_{\mathbb{G}_2} \tag{4}$$

For the purposes of this paper, $\mathbb{G}_1$ is the ECG, and $\mathbb{G}_2$ the underlying field.

Bilinear maps have two useful properties that make them ideal for IBC:

(1) The DDH problem becomes easy, but CDH remains hard
(2) As far as anyone knows, bilinear maps are not computationally feasible to invert

The first property forms the main reason to use bilinear maps: it allows manipulation of the exponents / coefficients without extracting them. This manipulation in turn enables combining two asymmetric keys such that they cancel each other out, and retrieving the message. The manipulation does not reveal the secret parts of the key, however, and this is where the second part plays an important role. The complexity assumptions presented here would not hold, if bilinear maps were efficiently invertible.

2.2. *Complexity Assumptions*

2.2.1. *BLDDH versions*

Our scheme assumes certain complexity-theoretic conjectures about calculation with exponents. They are both defined under bilinear groups and bilinear maps. They are extensively used in IBC, starting from [3] and moving up to [16] with the modified version.

Decisional Bilinear Diffie-Hellman Assumption (BLDDH):

Given a bilinear group $\mathbb{G}_1$, the "mapping function target set" (or codomain-) group $\mathbb{G}_2$, a bilinear mapping $e : \mathbb{G}_1 \times \mathbb{G}_1 \longrightarrow \mathbb{G}_2$, a generator $P \in \mathbb{G}_1$ and four random integers $a, b, c, z \in \mathbb{Z}_p$, no polynomial-time algorithm can distinguish between tuples $< aP, bP, cP, e(P,P)^{abc} >$ and $< aP, bP, cP, e(P,P)^{z} >$ with more than a negligible advantage.

Modified Decisional Bilinear Diffie-Hellman Assumption (mBLDDH):
Given a bilinear group $\mathbb{G}_1$, its codomain group $\mathbb{G}_2$, a bilinear mapping $e : \mathbb{G}_1 \times \mathbb{G}_1 \longrightarrow \mathbb{G}_2$, a generator $P \in \mathbb{G}_1$ and four random integers $a, b, c, z \in \mathbb{Z}_p$, no polynomial-time algorithm can distinguish between tuples $< aP, bP, cP, e(P,P)^{abc^{-1}} >$ and $< aP, bP, cP, e(P,P)^{z} >$ with more than a negligible advantage.

The assumptions have so far proven to be valid, but they have not been studied as extensively as for example DDH and CDH. Indeed, some recent results with elliptic curves [20] suggest that BLDDH-assumptions are not on as secure grounds as DDH or CDH. On the other hand the complexity assumptions are based on bilinear maps: if Weil or Tate pairings are shown to be insecure, the IBC constructions themselves remain valid, although without an efficient implementation base.

2.3. *Fuzzy IBC*

The FC-PKC scheme we present here, is an extension of the fuzzy IBC scheme in [16]. For later comparison, we will review the main elements of the scheme.

2.3.1. *Algorithms*

The fuzzy IBC [16] consists of four randomized algorithms: Setup, KeyGeneration, Encryption and Decryption. They operate on (multiplicative) bilinear groups $\mathbb{G}_1$ and $\mathbb{G}_2$ with a bilinear map e between them. Additionally the authors of [16] define a Lagrange coefficient

$$\Delta_{i,S}(x) = \prod_{j \in S, j \neq i} \frac{x-j}{i-j}, \text{ where } i \in \mathbb{Z}_p, S \subset \mathbb{Z}_p \tag{5}$$

The Lagrange coefficient is needed for interpolating the private key polynomial in decryption. The algorithms are defined as follows:

Setup(d):

(1) Define a universe $\mathcal{U}$ of elements. Associate them with a proper set $\mathbb{Z}_p$: 1,...,$|\mathcal{U}|$ *(mod p)*, the order p being the same as the bilinear group order
(2) Choose master key as $t_1, .., t_{|\mathcal{U}|}, y \in_U \mathbb{Z}_p$
(3) Publish the group descriptions, generators and the public key

$$T_1 = g^{t_1} \tag{6}$$

$$\vdots$$

$$T_{|\mathcal{U}|} = g^{t_{|\mathcal{U}|}}, \tag{7}$$

$$Y = e(g, g)^y \tag{8}$$

(4) Use d as the parameter describing, how many common elements must two identities have, in order to be defined as being "close" to each other

KeyGeneration:

(1) Choose a random polynomial q over $\mathbb{Z}_p$ such that its degree is $d - 1$ and $q(0) = y$.
(2) Given an identity $\omega \subseteq \mathcal{U}$ set the private key as $D = \{D_i\}_{i\in\omega} = \{g^{q(i)/t_i}\}_{i\in\omega}$

Encryption:

(1) Given a public key $\omega' \subseteq \mathcal{U}$ and a message $M \in \mathbb{G}_2$, choose $s \in_U \mathbb{Z}_p$
(2) Publish the ciphertext as $E = < \omega', E' = MY^s, \{E_i = T_i^s\}_{i\in\omega'} >$

Decryption:

(1) Given a ciphertext E and two identities $\omega, \omega' \subseteq \mathcal{U}, |\omega \cap \omega'| \geq d$, choose an arbitrary $S \subset |\omega \cap \omega'|, |S| = d$.
(2) Decrypt the ciphertext as $M = E'/\prod_{i\in S}(e(D_i, E_i))^{\Delta_{i,S}(0)}$.

2.3.2. *Security Model*

The security model used in fuzzy IBC is called a fuzzy selective ID security [16]. The model consists of a game (a slight modification of the selective ID game presented in [3]) and a statement of probabilities on winning the game.

Game: Fuzzy Selective ID

Init. Define two parties, an adversary $\mathcal{A}$ and a challenger $\mathcal{C}$. $\mathcal{A}$ declares an identity ω ,which he is to be challenged on. The common elements limit d is defined.

Setup. $\mathcal{C}$ runs the setup phase of the cryptographic scheme, and forwards the public parameters to $\mathcal{A}$.

Phase 1. $\mathcal{A}$ can query for private keys for a number of identities γ_i, such that always $|\gamma_i \cap \omega| < d$.

Challenge. $\mathcal{A}$ submits two messages, M_0 and M_1 of equal length. $\mathcal{C}$ flips a fair coin, and with the result, b, encrypts M_b. The resulting ciphertext is passed to $\mathcal{A}$.

Phase 2. Repeat Phase 1.

Guess. $\mathcal{A}$ outputs a guess b' of b.

Advantage: Defined from the perspective of $\mathcal{A}$ as: $P(b' = b) - 1/2$.

Definition of fuzzy selective ID security [16] is as follows: A scheme is secure in the Fuzzy Selective-ID model of security if all polynomial-time adversaries have at most a negligible advantage in the game "Fuzzy Selective ID".

3. The Construction of FC-PKC

The Fuzzy Certificateless Public-Key Cryptography works on the same mathematical background as fuzzy identity-based cryptography. However, the architecture and key management differ (see figure 1). The basic modification to fuzzy IBC is to randomize the public and private key generation. Unfortunately this destroys the natural association between the identity and the public key, but we add a "self-generated" certificate to the public key. The public key binds together the master key, identity and randomization information with bilinear mappings. Performing the certificate check is a simple calculation of a pairing function, and by getting an accepted result the verifier can trust that the public key is fresh, that it belongs to the identity intended, and that it has been issued by the key generation center (KGC) of that particular user. In addition, the public key follows the principles of fuzzy encryption, in that it has been defined element by element. This type of definition allows the verifier perform a "fuzzy" check: he can select a smaller number of elements to check than is actually the full identity resulting in a slightly incomplete or fuzzy check.

The security assumptions and the encryption and decryption procedures are identical to the fuzzy IBC. However, we work explicitly with

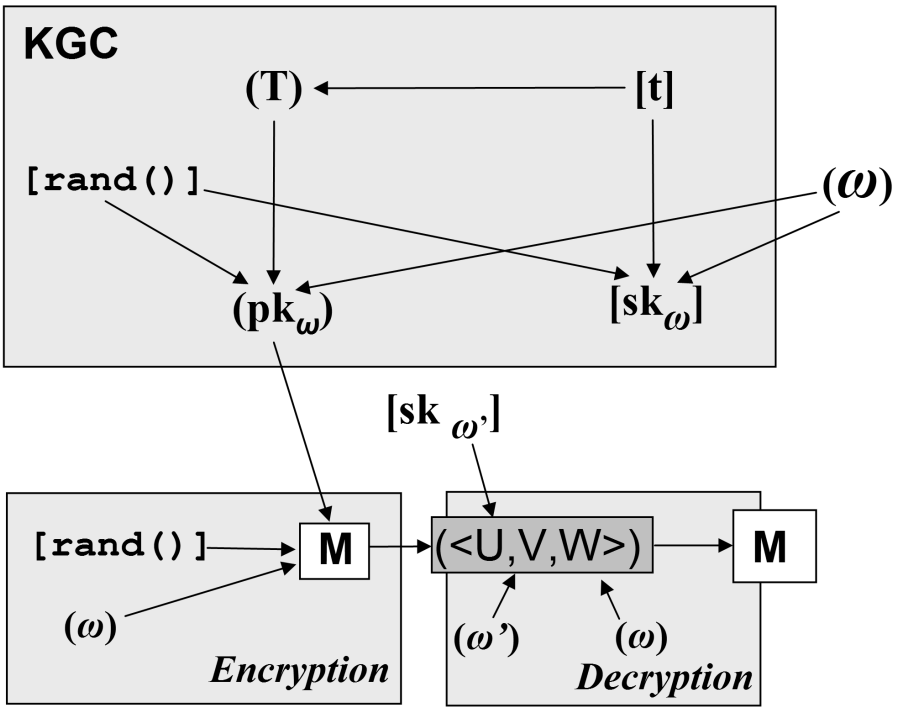


Fig. 1. FC-PKC key management and operations. Public values are in parenthesis, private values in square brackets.

elliptic curve groups (ECG), so we adopt the notation from EC groups and CL-PKC. We describe here six randomized operations reminiscent of [1]: Setup, Set-User-Random-Value, Set-Private-Key, Set-Public-Key, Encrypt and Decrypt. In CL-PKC, the Extract-Partial-Private-Key and Set-Private-Key are two separate algorithms, which we combine here, but the reason is only architectural: the algorithms can as well be separated between the KGC and the user. Here we assume (for simplicity) that the user random value is generated and integrated to the public and private keys at the KGC.

As the architecture and high-level operations are similar to the certificateless schemes, we adopt here name conventions from Al-Riyami and Paterson [1]. However, for the detailed operations we use symbols and terminology from Sahai and Waters [16], with emphasis on ECGs and thus using notation from additive group operations, e.g. exponentiation is shown as point multiplication.

Setup. At setup, the system parameters are selected at the KGC. These include the algebraic structures, operational parameters and secret and public keys. FC-PKC operates mostly on supersingular elliptic curve groups, which allows modified Weil or Tate pairings as bilinear mappings. The construction does not preclude the use of other bilinear groups, but ECGs are by far the most efficient and widely studied groups with bilinear properties. We use such prime-order ECG and denote it by $\mathbb{G}_1$, and one of its generators by P. Bilinear mappings use the underlying field of the ECGs as their codomain, usually $\mathbb{Z}_p$. The codomain field of the mapping is denoted by $\mathbb{G}_2$, and the bilinear mapping between them is defined as $e : \mathbb{G}_1 \times \mathbb{G}_1 \longrightarrow \mathbb{G}_2$.
The operational parameters are the security parameter $\mathcal{K}$ and the common elements limit d. The common elements limit means the number of elements two identities need to have in common, before they are considered the same under the metric. This is stated as a fixed constant for simplicity see future work for extended considerations.
The identities are here described as in [16], as subsets of some universe $\mathcal{U}$. The scheme becomes easier to present with this simplification, and actual implementations can use the theory irrespective of the definition of the universe. Here, $\mathcal{U}$ is taken as the first elements of $\mathbb{Z}_p/\{0\}$ such that $|\mathcal{U}| < p$.
As the error-correction is performed with the help of interpolating polynomials, the fuzzy scheme require the use of Lagrange coefficients, as in [16]. The coefficient is defined as:

$$\Delta_{i,K}(x) = \prod_{j \in K, j \neq i} \frac{x-j}{i-j}, \text{ where } i \in \mathbb{Z}_p, K \subset \mathbb{Z}_p. \tag{9}$$

The system will generate a master key, one independent random integer from $\mathbb{Z}_p/\{0\}$ for each of the elements in $\mathcal{U}$. It should be ensured that the field order is significantly larger than the universe size to prevent the exhaustion of random number space. These integers serve to protect and encode the fuzzy identity. In addition, an extra independent random integer is selected from $\mathbb{Z}_p/\{0\}$ to act as the actual encryption base. In total, the master key will consist of a tuple $t = t_1, ..., t_{|\mathcal{U}|}, y \in_U \mathbb{Z}_p/\{0\}$. The exclusion of the zero element guarantees that the master key elements will have an inverse, which is an essential requirement in decryption and calculating user secret keys.
The KGC public key (or system public encryption parameters) elements

are the same as the master key elements protected elementwise by ECG multiplication:

$$T = < \{T_i\}_{i \in \mathcal{U}}, Y > \tag{10}$$

$$T_i = t_i P \tag{11}$$

$$Y = e(P, P)^y \tag{12}$$

Note that the KGC public key has elements from both $\mathbb{G}_1$ and $\mathbb{G}_2$. This is due to the different purposes of the key elements: the latter part is used for encryption while the first parts are used in error correction.

Set-User-Random-Value. This algorithm selects $|\mathcal{U}|$ random values $r_i, ..., r_{|\mathcal{U}|}$ from $\mathbb{Z}_p/\{0\}$ to be used in the public and private keys. The exclusion of the zero-element again exists to guarantee, that each random value will have an inverse in $\mathbb{G}_2$. The algorithm is executed at KGC for each user. Note that the KGC must generated random values for each of the elements in $\mathcal{U}$ per user. The reasons for this are as follows:

- At encryption it cannot be known beforehand, which of the fuzzy identitys elements match and which do not. If the element set is defined as generally as here, and the universe is not very large, this is the only way to guarantee that encryptor has access to all the necessary public keys. For large universe constructions, more restrictions need to be placed for the elements along with another metric. The restrictions are needed to limit the number of random values.
- To be able to handle error-correction and randomization together mathematically, the random values for each element need to be the same in encryption and in decryption.
- The assurance of revocation of all fuzzy keys along with the basic identity used in the private key is not possible, if the random values for other elements are not treated the same way as those of the basic identity.

Set-Private-Key. Setting the user private key is done at the KGC by default in this scheme. Setting the private key incorporates selecting a random polynomial (KGC use only) for each user, and calculating the actual private key with the help of the polynomial, user randomization and the master key.

The polynomial is defined by $q(x), x \in \mathbb{Z}_p, deg(p) = d - 1, q(0) = y$. Note that the zero-value is fixed to provide a necessary evaluation point

for the interpolation. The hidden polynomial exists to guard against colluding users to combine their valid keys to obtain additional key material. It does not help in revocation, which can easily be seen from the encryption part: the encryption is performed in the same manner irrespective of the polynomial. Hence it cannot be revoked, although a decrypting party must have knowledge of some polynomial used for that user at some point in time. The per-user random polynomial has the same scope as the per-user random value: all (close-by) identities of the same user work under the same polynomial.

An identity is needed to define a private key. In this scheme whatever identity the user presents at the private key generation time is taken as his/her identity. Later, during the (fuzzy) encryption phase, any identity that is close enough to the defined identity is accepted and usable. There is many-to-one correspondence between public keys and the private key.

The actual private key is then simply a set of differently multiplied ECG generators:

$$sk = \{q(i)(r_i t_i)^{-1} P\}_{i \in \omega} \tag{13}$$

It should be noted that the inverses of r_i and t_i exist and can be calculated, since the elements are chosen to be from $\mathbb{Z}_p/\{0\}$. Also, the identity is encoded as the points in the polynomial and coefficient selections of the ECG element. This type of encoding is necessary to enable indepence of the randomization part (affecting the key) and the fuzzy part (the identity).

If the key management is to be closer to CL-PKC, this algorithm can be split to two: Partial-Private-Key-Extract (executed at the KGC) and Set-User-Private-Key (executed at the user device). In this case also Set-User-Random-Value and Set-Public-Key need to be executed at the user, although they will not change in content. The Partial-Private-Key-Extract will operate like Set-Private-Key, except it will not use the per-user random value as a coefficient to the ECG point. This part is added in the user process, given the partial private key elements.

Set-Public-Key. The public key consists of two parts $< PK, ce >$: the actual key (PK) used for encryption and decryption (depending on the application) and the self-generated certificate ce, with which the sender can ensure that PK actually belongs to the identity claimed.

The public key is of the form

$$\begin{aligned} pk &= < PK, ce > \\ &= < \{PK_i\}, \{ce_i\} >_{i \in \mathcal{U}} \\ &= < \{r_i t_i P\}, \{r_i P\} >_{i \in \mathcal{U}} \end{aligned} \tag{14}$$

The certificate is checked at the beginning of each encryption operation (see Encrypt).
Public key is calculated either directly at the KGC or partly by the KGC (the partial private key in the more CL-PKC-like key management) and partly by the user (adding the randomization). Public key also needs to publish the information for the whole universe, since the public key is the sole source of information on the random values that will be available to other users.

Encrypt. The encryption is performed by any user. It is a fuzzy and randomized algorithm, using a per-message nonce. Encryption has two phases, presented below.

(1) Public-key check. Given the ECG generator P, KGC public key T, an approximated user id ω' and user public key $< PK, ce >$ check if

$$\begin{aligned} & e(P, PK) = e(ce, T). \text{ This requires that} \\ \forall (i \in \omega') : & e(P, r_i t_i P) = e(r_i P, t_i P) \end{aligned} \tag{15}$$

If the equations are valid, continue to step 2, otherwise output error.

(2) Encryption. Given an approximated user id ω', plaintext $M \in \mathbb{G}_2$, a per-message nonce $s \in_U \mathbb{Z}_p$, output $C =< U, V, W >$, where

$$\begin{aligned} U &= \omega' \\ V &= MY^s \\ W &= \{sPK_i\}, i \in \omega' \end{aligned} \tag{16}$$

The first phase of the encryption is intended to protect against the Denial-of-Decryption (DoD) attack, where an adversary replaces the public key of an identity with another valid identity [14]. In CL-PKC, although the adversary cannot decrypt the message, neither can the receiver, and the sender has no way of checking the binding between the identity and the public key. Our randomization technique enables this checking without exposing any secret information.

The pk-check needs to be done only against the elements in ω' . This is sufficient, if revocation changes all the values $r_i, i \in \mathcal{U}$. If an adversary can use some old user random values from previous, exposed keys, he/she is able to gain extra information.
The encryption phase protects the message (which is assumed to be encoded to a field element in $\mathbb{G}_2$) by a one-time permutation key of the KGC public key Y. The nonce is in turn encoded to the user public key, based on his/her approximated user ID. The "unfuzzification" is left for the decrypting party encryption phase merely inputs as much information as needed, which is possible only if the whole element universe is randomized (this will not be the case with specific implementations and large universes).
Since there is no way to know, which exact identity was used to encrypt the message, ω' must be included in the message as well.

Decrypt. Decryption performs the major work in the system, as the process has the main responsibility of error correction and extracting the message under the protection. It also will have to perform two steps. Given ciphertext $C =< U, V, W >\in \mathcal{U} \times \mathbb{G}_2 \times \mathbb{G}_1^{|U|}$, common elements bound d, user private key $sk = \{sk_i\}_{i\in\omega}$ and the basic identity ω , Decrypt will

(1) Perform check on the number of common elements in the identities (whether sufficient information exists): If $|\omega \cap U| < d$, output error, otherwise continue to step 2.
(2) Decrypt using any sufficiently large set of the element universe such that they are all common to both of the basic and approximated identities:
   - Select $K \subset (\omega \cap U), |K| = d$
   - Output $M = V(\prod_{i\in K}(e(sk_i, W_i))^{\Delta_{i,K}(0)})^{-1}$

To see that the decryption is well defined and actually recovers the original message, we write M as:

$$
\begin{aligned}
&= V(\prod_{i\in K}(e(sk_i, W_i))^{\Delta_{i,K}(0)})^{-1} \\
&= Me(P,P)^{ys}(\prod_{i\in K}(e(q(i)(r_it_i)^{-1}P, sr_it_iP))^{\Delta_{i,K}(0)})^{-1} \\
&= Me(P,P)^{ys}(\prod_{i\in K}(e(P,P)^{sq(i)})^{\Delta_{i,K}(0)})^{-1}
\end{aligned}
$$

$$
\begin{aligned}
&= Me(P,P)^{ys}(e(P,P)^{s\sum_{i\in K} q(i)\prod_{j\in K, j\neq i}\frac{-j}{i-j}})^{-1} \\
&= Me(P,P)^{ys}(e(P,P)^{sq(0)})^{-1} \\
&= Me(P,P)^{ys}(e(P,P)^{-sy} \\
&= M
\end{aligned} \tag{17}
$$

The first two equations only expand the notations, while the third one shows how the KGC master key and the user random value cancel each other out. The fourth and fifth equation perform the "unfuzzification" process with a Lagrange interpolation polynomial: the polynomial $q(x)$, which is only known to the KGC, is interpolated at zero with the help of the common elements in the identities and the polynomial values encrypted in the user private key. The final extraction is performed, when the polynomial at zero equates to a part of the KGC master key, and cancels out the nonce.

The role of the polynomial can be seen from the decryption procedure as well: assume there are two polynomials, $q_1(x)$ and $q_2(x)$, and a message M encrypted with identical identities and public keys. Now it is obvious that as long as the both polynomials evaluate to y at zero, and the decryptor has in possession a private key with the correct KGC master key elements and random values, the message can be opened with them both. However, two users cannot combine parts of their elements to form new (and possibly valid) identities even if they used the same random value sets, because the polynomials are generated separately by the KGC for each user, and the actual values of the polynomials are protected by the master key and CDH assumption for ECGs. To enable revocation, the random values need to be changed as well.

Revocation is accomplished simply by rerunning Set-User-Random-Value, Set-Private-Key and Set-Public-Key.

4. Notes on the Security of FC-PKC

The FC-PKC scheme security relies on the hardness of the modified bilinear computational Diffie-Hellman problem (mBLCDH), same as mBLDDH except that instead of making the decision, one would actually have to produce the calculation in the exponent.

It is trivial to see that if an adversary is able to perform an inversion of the exponent, the private key is left only to the security of the per-user but all-time valid polynomial, effectively breaking the scheme. The reverse implication is conjectured to apply as well, due to the similarity of the FC-

PKC and fuzzy IBC. The different use of the public key and addition of an extra random value, however, do not enable us to extend to proof by Sahai and Waters trivially, so it is left for future research. The security model itself seems to be a little different from mBLDDH.

The check performed to the public key does not reveal extra information either (given BLDDH and mBLDDH); this is because the relevant decisional information present is already used for the validity check all the combinations attainable from the bilinear map and the public information used in the cryptosystem are already public information. Unless the adversary is able to extract a ECG discrete logarithm or inverse the exponent, the public key constructions protect the embedded secret values.

5. Applications and Future Work

The intended application area of the scheme is for biometric encryption. Biometric architectures require fuzzy constructions, but security requirements dictate that full keys should be revocable. In biometric authentication schemes, two-factor approaches are often implemented. In two-factor schemes, the authentication information is divided between identity (usually biometrics) and some secret knowledge (a token or a password). FC-PKC is constructed to fulfill the basic requirements of such an application. The actual architecture of such a two-factor system is left for future work, but we give here the basics of how FC-PKC can be applied to form a two-factor biometric system.

Suppose the user carries with him a token, which encodes his public key (and possibly relevant information of his KGC, or even the KGC public key) for his currently chosen random values. If the token encodes random (and mostly meaningless) values for all of the elements of the used universe, as suggested in the algorithm Set-Public-Key, the user of the token has very little use for the token without the corresponding identity. This is because for an identity ω an adversary would have to guess a subset $\omega' \subset \mathcal{U}$, such that $|\omega \cap \omega'| \geq d$. The probability for this is

$$Pr\{|\omega \cap \omega'| \geq d\} = \sum_{k=d}^{|\omega|} \frac{\binom{|\omega|}{k}\binom{|\mathcal{U}|-|\omega|}{|\omega|-k}}{\binom{|\mathcal{U}|}{|\omega|}} \tag{18}$$

This attains a maximum where d equals about half of the identity. If about 10% error margin is allowed, the probability decreases with a high-degree polynomial as the size of the universe grows, assuming the universe

is at least an order of magnitude larger than the identity.

It should be noted that FC-PKC is not by itself a two-factor scheme, since the relation from the public key to the identity is currently symmetrical and trivial. Architectural considerations, such as hierarchical identities, can make the relation asymmetric and less trivial: as an example the public key can be fetched based on a name, and selecting the key parts based on a lower-level identity, such as a fingerprint. A full biometric architecture with hierarchical identities (and corresponding hierarchical IB cryptography) is left for future considerations. Such architectures can use FC-PKC as the basis for public-key biometric authentication schemes as well.

For implementation, a bilinear map is needed. There are two common bilinear map types used: (modified) Weil pairing (that satisfies both conditions of the bilinear map) and Tate pairing. In a recent paper by Granger, Page and Smart [11] Tate pairings are found to be more efficient in all cases of the underlying ECG field $\mathbb{F}_{p^k}$ parameters, so it is natural to recommend Tate pairings for the implementations. On another implementation-related note, the FC-PKC scheme makes some simplifications. One of them is using a somewhat vague description of the element universe. We acknowledge that an implementation will need to consider things such as

- arbitrarily long identity strings
- investigating whether $\mathbb{Z}_p$ is the best possible field for the element universe
- mappings from $\{0,1\}^*$ to a suitable field

The implementation details are considered outside the scope of this paper, and left for future research.

The final significant part of future research is the security proof and the security model As mentioned in the security considerations, for future research is also left the security proof for the direction that given an algorithm breaking FC-PKC, one can construct, in a polynomial time, an algorithm breaking the mBLDDH.

6. Conclusions

In this paper we have developed an error-correcting and self-certifying public-key cryptosystem based on identities and bilinear mappings. This construction called FC-PKC uses Shamir secret sharing and a modified key management architecture and user randomness from CL-PKC.

The FC-PKC is conjectured to rely on the mBLDDH problem for its security, which is due to its similarity to the fuzzy IBC construction using

the polynomial interpolation method. The main contributions of the fuzzy IBC are to combine error correction to key revocation in an independent way, which in turn leads to a certain class of applications.

The applications are aimed at a practical biometric framework where key revocation is a meaningful concept, and the whole framework is as dependent on token-based encryption as it is on biometrics based encryption.

FC-PKC seems to us as a key ingredient in a biometric framework, which enjoys the benefits of a conventional cryptographic security architecture, nevertheless fully embracing biometric properties to enable a more natural human-computer interaction (HCI).

References

1. S.S. Al-Riyami and K. Paterson, *Certificateless public key cryptography*. Proc. of Asiacrypt 2003, LNCS 2894, Springer-Verlag, pp.452473, 2003.
2. J. Baek, W. Susilo and J. Zhou, *New Constructions of Fuzzy Identity-Based Encryption, Full version* , http://www1.i2r.a-star.edu.sg/jsbaek/.
3. D. Boneh and M. Franklin, *Identity-Based Encryption from the Weil Pairing*, SIAM J. of Computing, vol. 32, no. 3, pp.586-615, 2003.
4. D. Boneh and X. Boyen, *Efficient Selective-ID Secure Identity Based Encryption Without Random Oracles*, Proc. of Eurocrypt 2004, LNCS 3027, Springer-Verlag, pp.223238, 2004.
5. A. Burnett, F. Byrne, T. Dowling, A. Duffy, *A Biometric Identity Based Signature Scheme*, International of Journal of Network Security, 2007, In press.
6. C. Candolin and M. Kiviharju: *A Roadmap Towards Content Based Information Security*, Proc. of the 6th European Conference on Information Warfare and Security (ECIW). Shrivenham, UK 2 - 3 July 2007, pp.37-46, 2007.
7. Y. Dodis, L. Reyzin and A. Smith, *Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data*, Proc. of Eurocrypt 2004, LNCS 3027, Springer-Verlag, pp.523540, 2004.
8. C. Gentry and A. Silverberg, *Hierarchical ID-based Cryptography*, Proc. of Asiacrypt 2002, LNCS 2501, Springer-Verlag, pp.548-566, 2002.
9. C. Gentry, *Certificate based encryption and the certificate revocation problem*, Proc. of Eurocrypt 2003, LNCS 2656, Springer-Verlag, pp.272293, 2003.
10. M. Girault, *Self-certified public keys*, Proc. of Eurocrypt 1991, LNCS 547, Springer-Verlag, pp.490-497, 1992.
11. R. Granger, D. Page, and N.P. Smart, *High Security Pairing-Based Cryptography Revisited*, http://eprint.iacr.org/2006/.
12. Y. Hanaoka, G. Hanaoka, J. Shikata and H. Imai, *Identity-Based Hierarchical Strongly Key-Insulated Encryption and Its Application*, Proc. of Asiacrypt05, LNCS 3788, Springer-Verlag, pp.495-514, 2005.
13. A. Joux, *A one-round protocol for tripartite Diffie-Hellman*, Proc. of Algorithmic Number Theory Symposium, Leiden, The Netherlands, pp.385-394, 2000.

14. J. Liu, M. Au and W. Susilo, *Self-Generated-Certificate Public Key Cryptography and Certificateless Signature / Encryption Scheme in the Standard Model*, 2007 ACM Symposium on Information, Computer and Communications Security (ASIACCS'07), pp. 273 - 283, 2007.
15. A. Menezes, T. Okamoto and S. Vanstone, *Reducing Elliptic Curve Logarithms in a Finite Field*, IEEE Transactions on Information Theory, vol. IT-39, no.5, pp. 1639-1646, 1993.
16. A. Sahai and B. Waters, *Fuzzy Identity Based Encryption*, Proc. of Eurocrypt 2005, LNCS 3494, Springer-Verlag, pp.457473, 2005.
17. R. Sakai, K. Ohgishi and M. Kasahara, *Cryptosystems based on pairing*, The 2000 Symposium on Cryptography and Information Security, Okinawa, Japan, Jan. 26-28, 2000.
18. A. Shamir, *Identity-based cryptosystems and signature schemes*, Proc. of Crypto 1984, LNCS 196, Springer-Verlag, pp.47-53, 1985.
19. A. Shamir, *How to Share a Secret*, Communications, ACM 22(11), pp.612-613, 1979.
20. E. Verheul, *Evidence that XTR is more secure than supersingular elliptic curve cryptosystems*, Journal of Cryptology, 17 (2004), IACR, 277-296, 2004.

Trace Zero Varieties over Fields of Characteristic 2 for Cryptographic Applications

Roberto M. Avanzi

Faculty of Mathematics
Horst Görtz Institute for IT Security
Ruhr University Bochum – Germany.
E-mail : Roberto.Avanzi@ruhr-uni-bochum.de

Emanuele Cesena

Dipartimento di Matematica
Università degli Studi RomaTRE
Rome – Italy.
E-mail : cesena@mat.uniroma3.it

Given a low genus hyperelliptic curve defined over a finite field $\mathbb{F}_q$, a Trace Zero Variety is a specific subgroup of the group of the divisor classes on the curve which are rational over a small degree extension $\mathbb{F}_{q^n}$ of the definition field. Trace Zero Varieties (TZV) are interesting for cryptographic applications since they enjoy properties that can be exploited to achieve fast arithmetic and group construction.

In this paper we consider TZV over fields of characteristic two. We also underline the similarities and the differences between TZV in even and odd characteristic. We provide experimental results to show their performance advantages, and find that they provide an efficient alternative to ordinary curves. Furthermore TZV of elliptic curves over degree 5 extensions come out as one of the most efficient groups, suitable to build cryptographic systems based on DLP: at the same group size, they are almost three times faster than elliptic curves.

1. Introduction

The *Discrete Logarithm Problem* (DLP) in groups arising from geometric constructions is nowadays an established family of primitives for the design of secure cryptosystems. A common choice of group is the rational point group of the Jacobian variety of a hyperelliptic curve of low genus. As early as 1985 Miller [36] and Koblitz [26] independently proposed to use elliptic curves, which are the hyperelliptic curves of genus one. Shortly thereafter

Koblitz [28] proposed to use Jacobians of hyperelliptic curves of higher genus.

In 1998 Frey [17,18] suggested to use Trace Zero Varieties because their properties can be exploited to implement fast arithmetic, while the hardness of the DLP in them can be reduced to the hardness of the DLP in other known groups, which are used in practice.

Roughly speaking, if C is a hyperelliptic curve of genus g defined over a finite field of q elements $\mathbb{F}_q$, the *trace zero (sub)variety of C over a field extension of degree n* is a subgroup G of the Jacobian variety $J_C(\mathbb{F}_{q^n})$ of C over $\mathbb{F}_{q^n}$, that is isomorphic to the quotient group $J_C(\mathbb{F}_{q^n})/J_C(\mathbb{F}_q)$. It is a codimension one subvariety of the Weil descent of $J_C(\mathbb{F}_{q^n})$ on $\mathbb{F}_q$.

The group used for cryptographic applications is a subgroup G_0 of G of large prime order ℓ and of small index in G. Since we can assume that G_0 is the only subgroup of order ℓ of G, the Frobenius endomorphism σ of the Jacobian induced by the relative Frobenius automorphism of the field extension $\mathbb{F}_{q^n}/\mathbb{F}_q$ acts on G_0 as a group automorphism. Since G_0 is cyclic, σ a scalar multiplier, *i.e.* for each $D \in G_0$ we have $\sigma(D) = sD$ for some $s \in \mathbb{Z}$ (that depends on G_0). Similarly to the setting of Koblitz curves [29,45], the Frobenius endomorphism can be used to speed up the scalar multiplication in G_0. Another advantage that Trace Zero Varieties share with Koblitz curves is that it is quite easy (from a computational point of view) to determine the cardinality of the rational point group.

Naumann [40] and Blady [8] consider TZV of elliptic curves over extension fields of degree 3 ($n = 3$); Weimerskirch [48] analyses the case for extension fields of degree 5; finally Lange [30,31] builds TZV from the Jacobian variety of hyperelliptic curves of genus two, over extension fields of degree 3. From a cryptographic point of view, these are the only relevant cases, since for higher extension or higher genus the groups obtained are susceptible to Weil descent attacks [31]. Avanzi and Lange [4] compare the performance of these three kinds of TZV over fields of odd characteristic. In our paper we compare the same three types of TZV defined over binary fields, underlining similarities and main differences between TZV defined over fields of even and odd characteristic.

In Section 2 the main concepts about low genus hyperelliptic curves are recalled. Section 3 is devoted to the definition and main properties of Trace Zero Varieties. Security issues are briefly addressed in Section 4. Section 5 describes our implementation techniques. Examples of curves with cryptographically good orders are given in an Appendix. These are the curves used to get the experimental results reported and discussed in Section 7.

We conclude in Section 8.

2. Hyperelliptic curves

In this paper $\mathbb{F}_q$ denotes the Galois field of order q, where we shall in fact restrict ourselves to the case $q = 2^k$, even when some results are more general.

There are several books on the subject of elliptic and hyperelliptic curves. For reference material within the perspective of cryptographic of applications we refer to [3]. We mention here only a few facts.

A *hyperelliptic curve* C of genus g over $\mathbb{F}_q$ having an $\mathbb{F}_q$-rational Weierstraßpoint is a non singular curve defined by the equation

$$y^2 + h(x)y = f(x), \qquad f \text{ monic}, \ \deg f = 2g+1, \ \deg h \leq g \ .$$

A hyperelliptic curve of genus 1 is called an *elliptic curve*.

For every finite extension $L/\mathbb{F}_q$, the *Jacobian variety* $J_C(L)$ of C over L is isomorphic (as group) to the ideal class group over L. Hence we can represent the elements of $J_C(L)$ by a pair of polynomials with coefficients in L (Mumford's representation, [34]) and compute the group law using Cantor's algorithm [12,28].

The *Frobenius endomorphism* σ operates on a element of $J_C(L)$, represented by a pair of polynomials $[u, v]$, by raising each coefficient of u and v to the power of q. It satisfies the characteristic polynomial:

$$\chi(T) = T^{2g} + a_1 T^{2g-1} + \cdots + a_g T^g + + \cdots + a_1 q^{g-1} T + q^g \ , \qquad (1)$$

where $a_i \in \mathbb{Z}$. The Hasse–Weil theorem states that from the complex roots τ_i of $\chi(T)$ we can obtain the group order over any extension:

$$|J_C(\mathbb{F}_{q^n})| = \prod_{i=1}^{2g} (1 - \tau_i^n) \ ,$$

in particular $|J_C(\mathbb{F}_q)| = \chi(1)$.

2.1. *Ordinary elliptic curves over binary fields*

A nonsingular, ordinary (non supersingular) elliptic curve E over a binary field $\mathbb{F}_q$ can always be brought to the following Weierstraß form

$$y^2 + xy = x^3 + ax^2 + b, \qquad \text{with} \quad a \in \mathbb{F}_2, \ b \in \mathbb{F}_q \ . \qquad (2)$$

2.2. *Ordinary genus 2 HEC over binary fields*

In [10,13,32], ordinary genus 2 hyperelliptic curves over binary fields with a fixed rational point at infinity are classified in terms of the degree of the polynomial $h(x)$. In order to be able to use the Lange–Stevens' doubling formulæ [32] we shall consider the so-called curves of Type II (cf. [32] or § 14.5 of [3]), which have an equation of the form

$$y^2 + xy = x^5 + f_3 x^3 + \varepsilon x^2 + f_0, \qquad \varepsilon \in \mathbb{F}_2,\ f_0, f_3 \in \mathbb{F}_q \ . \tag{3}$$

3. Trace zero varieties

Let C be a hyperelliptic curve over a finite field $\mathbb{F}_q$, given by equation (3); let $J_C(\mathbb{F}_{q^n})$ be the Jacobian variety of C and let $D \in J_C(\mathbb{F}_{q^n})$.

We can formally define the *trace* of D, as for fields:

$$\mathrm{Tr}(D) = D + \sigma(D) + \cdots + \sigma^{n-1}(D) \ .$$

It is clearly an endomorphism of $J_C(\mathbb{F}_{q^n})$. The *trace zero (sub)variety* (TZV) of $J_C(\mathbb{F}_{q^n})$ is the set:

$$G = \{D \in J_C(\mathbb{F}_{q^n}) \mid \mathrm{Tr}(D) = \boldsymbol{O}\} \ ,$$

where $\boldsymbol{O}$ is the neutral element in $J_C(\mathbb{F}_{q^n})$. Being the kernel of the trace endomorphism, G is a group. It is a codimension 1 subvariety of the Weil descent of $J_C(\mathbb{F}_{q^n})$ on $\mathbb{F}_q$.

Now, the intersection of G and $J_C(\mathbb{F}_q)$ is formed by the n-torsion elements of $J_C(\mathbb{F}_q)$. Let $v = \gcd(n, |J_C(\mathbb{F}_q)|)$. If $v = 1$ then this intersection is empty and we can compute the group order as follows:

$$|G| = \frac{|J_C(\mathbb{F}_{q^n})|}{|J_C(\mathbb{F}_q)|} \ .$$

If $v \neq 1$ then $G \cap J_C(\mathbb{F}_q)[n] \neq \emptyset$, but, being interested in subgroups of large prime order, we could still take a subgroup of G not intersecting the n-torsion. However, we can restrict ourselves to the case $v = 1$ in what follows.

From the cryptographic point of view, only the following cases are relevant: $g = 1,\ n = 3$; $g = 1,\ n = 5$; $g = 2,\ n = 3$ (cf. Section 4). The following proposition (cf. [4]) gives us the group orders and useful bounds.

Proposition 3.1. *For the group order of trace zero varieties in the considered cases we have*

(1) For $g = 1$ and $n = 3$:

$$|G| = q^2 - q(1 + a_1) + a_1^2 - a_1 + 1$$

and

$$|G| \leq q^2 + 2q^{3/2} + 3q + 2q^{1/2} + 1 \ ; \tag{4}$$

(2) For $g = 1$ and $n = 5$:

$$|G| = q^4 - (a_1 + 1)q^3 + (a_1 + 1)^2 q^2 \ + \big(5a_1 - (a_1 + 1)^3\big)q - \big(5a_1(a_1^2 + a_1 + 1) - (a_1 + 1)^4\big)$$

and

$$|G| \leq q^q + 2q^{7/2} + 3q^3 + 4q^{5/2} + 5q^2 + 4q^{3/2} + 3q + 2q^{1/2} + 1 \ ; \tag{5}$$

(3) For $g = 2$ and $n = 3$:

$$|G| = q^4 - a_1 q^3 + (a_1^2 + 2a_1 - a_2 - 1)q^2 \ + (-a_1^2 - a_1 a_2 + 2a_1)q + a_1^2 + a_2^2 - a_1 a_2 - a_1 - a_2 + 1$$

and

$$|G| \leq q^4 + 4q^{7/2} + 10q^3 + 16q^{5/2} + 19q^2 + 16q^{3/2} + 10q + 4q^{1/2} + 1 \ . \tag{6}$$

Here, the integers a_1, a_2 are coefficients of the characteristic polynomial of the Frobenius endomophism (1)*.*

We can already see an advantage in using TZV: to count the number of points on such a variety we need to determine the characteristic polynomial of a curve defined over a smaller field than we would have if we considered plain elliptic and hyperelliptic curves with similar group size. This makes counting points on trace zero varieties much faster and allows for a faster generation of cryptographically suitable curves. For this there are several efficient methods, for example the AGM (arithmetic-geometric mean) method by Mestre [35].

For cryptographic applications, we want to work in a subgroup G_0 of G of large prime order ℓ (that may be G itself).

3.1. *Arithmetic in G_0*

Arithmetic in G_0 is performed using formulæ for the whole group $J_C(\mathbb{F}_{q^n})$. We may however speed up the scalar multiplication by making use of the Frobenius operation σ. It is still an open problem to find explicit (and faster) formulæ for TZV.

If G_0 is generated by D of order ℓ prime, then $\sigma(D) = sD$, for some integer s. Explicitly (see for instance [4] for proofs):

(1) For $g = 1$ and $n = 3$: $s = \dfrac{q-1}{1-a_1} \bmod \ell$;

(2) For $g = 1$ and $n = 5$: $s = \dfrac{q^2 - q - a_1^2 q + a_1 q + 1}{q - 2a_1 q + a_1^3 - a_1^2 + a_1 - 1} \bmod \ell$;

(3) For $g = 2$ and $n = 3$: $s = -\dfrac{q^2 - a_2 + a_1}{a_1 q - a_2 + 1} \bmod \ell$.

Using this result we want to replace any scalar multiplication mD ($|m| \leq \ell/2$) with the computation of

$$m_0 D + m_1 \sigma(D) + \cdots + m_{n-1}\sigma^{n-1}(D) \ ,$$

where $m_i = O(\ell^{1/(n-1)}) = O(q^g)$. In this way we are reducing the size of the scalars, and hence the number of "doubling" operations in a double-and-add scalar multiplication algorithm.

Since in cryptographic algorithms m is usually random, we can achieve the reduction following two ways: generate a random integer m, then split it in opportune m_i; or directly generate some m_i, by making sure to avoid collisions, *i.e.* different sequences of m_i give different elements of G_0.

The second approach has the advantage to save the time necessary to split the scalar, but in some context (for instance for digital signature verification) it is not applicable. Hence we are going to describe both.

3.1.1. *Scalar splitting*

Given an integer m with, without loss of generality, $|m| \leq \frac{\ell}{2} = \frac{|G_0|}{2}$, we want to compute some m_i of bounded size, such that:

$$mD = m_0 D + m_1 \sigma(D) + \cdots + m_{n-1}\sigma^{n-1}(D) \ .$$

First we write $m = n_0 + k_1 q^g$, with $|n_0| \leq q^{g/2}$. Then we use the fact that σ operate in G_0 as multiplication by s and the following relations modulo the group size ℓ:

$$\chi(s) \equiv 0 \bmod \ell \quad \text{and} \quad s^{n-1} + \cdots + s + 1 \equiv 0 \bmod \ell \ .$$

Proceeding in this direction we can expand m as desired, bounding each m_i to $O(q^g)$.

Theorem 3.1. *For the three cases we consider, there exists an efficient technique for expressing a scalar m in the form $m \equiv \sum_{i=0}^{n-2} m_i s^i \bmod \ell$ where $m_i = O(q^g)$. We have:*

(1) For $g = 1$ and $n = 3$, we have $|m_i| < 4q$ if $k \geq 7$;
(2) For $g = 1$ and $n = 5$, we have $|m_i| < 2q$ if $k \geq 17$;
(3) For $g = 2$ and $n = 3$, we have $|m_i| < 4q^2$ if $k \geq 23$.

Proof. The proof is essentially the same as in odd characteristic, as reported in [4], and so we limit here ourselves to sketch the approach for $g = 1$, $n = 3$ and to briefly remark the differences in the other two cases.

Let $g = 1$ and $n = 3$. Write first $m = n_0 + k_1 q$, with $|n_0| \leq q/2$. Now use the fact that $\chi(s) \equiv 0 \bmod \ell$, i.e. that $s^2 + a_1 s + q \equiv 0 \bmod \ell$ to get: $m \equiv n_0 + k_1(-a_1 s - s^2) \bmod \ell$. The term $-a_1 k_1$ is $O(q^{3/2})$, which is still too big, hence we need to reduce again: $-a_1 k_1 = n_1 + k_2 q$, with $|n_1| \leq q/2$. Then

$$m \equiv n_0 + n_1 s + k_2(-a_1 s^2 - s^3) - k_1 s^2 \bmod \ell \ .$$

All the coefficients are now $O(q)$.

At this time, we need to reduce the terms in s^2 and s^3, using the trace relation $s^2 + s + 1 \equiv 0 \bmod \ell$. We obtain:

$$m \equiv m_0 + m_1 s \mod \ell \ ,$$

with $m_0 = n_0 + k_1 - k_2 + a_1 k_2 = O(q)$ and $m_1 = n_1 + k_1 + a_1 k_2 = O(q)$. We may now explicitly bound m_0, m_1 using the estimation on the group size ℓ, given by (4), and finally obtain that $|m_0|$ and $|m_1|$ are both lower than $4q$ if $q > 73$, i.e. $k \geq 7$.

For $g = 1$, $n = 5$ we continue to reduce $n_i \equiv -a_1 k_i - k_{i-1} \bmod q$, $k_{i+1} := \frac{-a_1 k_i - k_{i-1} - n_i}{q}$, for $1 \leq i \leq 5$ (assuming $k_0 = 0$); The relation $\chi(s) \equiv 0 \bmod \ell$ is the same as above, and to reduce the powers of s greater than 4 the trace zero relation is $s^4 + s^3 + s^2 + s + 1 \equiv 0 \bmod \ell$. Proceeding as in the case $g = 1$, $n = 3$, we get a relation $m \equiv m_0 + m_1 s + m_2 s^2 + m_3 s^3 \bmod \ell$ where $m_1 < 91q$. In order to get the sharp bounds stated in the theorem, we reduce m_1 again once $m_1 = m_1' + k_2'(-a_1 s - s^2)$ and thus we consider the coefficients $m_0, m_1', m_2 - a_1 k_2'$ and $m_3 - k_2'$.

For $g = 2$, $n = 3$, since the constant term of the characteristic polynomial is q^2 we reduce as in the previous cases, but by taking quotients and rests by q^2. In order to get the sharp bounds m_1 must be reduced again as in the case $g = 1$, $n = 5$. □

As already mentioned, to construct the m_i one computes quotients and takes remainders modulo q^g, and for trace zero varieties in even character-

istic this amounts to simple bit shifting or masking operations, which are very efficient.

3.1.2. *Multi–scalar generation*

We consider now the technique of beginning with $(n-1)$-tuples of scalars, instead of deriving them from a single scalar. To avoid collisions, *i.e.* in order to ensure that different tuples $(m_0, ..., m_{n-2})$ and $(m'_0, ..., m'_{n-2})$ yield different elements of G_0, i.e. $\sum_{i=0}^{n-2} m_i s^i \not\equiv \sum_{i=0}^{n-2} m'_i s^i \bmod \ell$, we use the following theorem to bound the size of the scalars. The proofs of the three cases are immediate adaptations of the proofs found respectively in [40],[4] and [30], for the case where the definition field has odd characteristic.

Theorem 3.2. *Let D be a generator of G_0. Then the r^{n-1} elements $r_0 D + \cdots + r_{n-2}\sigma^{n-2}(D)$ are pairwise distinct for $r_i < r$, where:*

(1) For $g = 1$ and $n = 3$,

$$r := \min\left\{\frac{\ell}{q-a_1}, \frac{q-1}{\gcd(q-1, a_1-1)}\right\} \ ;$$

(2) For $g = 1$ and $n = 5$,

$$r := \min\left\{\frac{\ell}{(1+q+|a_1|q)M}, \frac{|q^2 - a_1^2 q + a_1 q - q + 1|}{\gamma}\right\} \ ,$$

where $M = \max\left\{|q^2 - a_1^2 q + 3a_1 q - 2q - a_1^3 + a_1^2 - a_1 + 2|,\ |q^2 - a_1^2 q - a_1 q + a_1^3 - a_1^2 + a_1|\right\}$ and $\gamma = \gcd(q^2 - a_1^2 q - q + 1, 2a_1 q - q - a_1^3 + a_1^2 - a_1 + 1)$;

(3) For $g = 2$ and $n = 3$,

$$r := \min\left\{\frac{\ell}{M}, \frac{q^2 - a_2 + a_1}{\gcd(q^2 - a_2 + a_1, a_1 q - a_2 + 1)}\right\} \ ,$$

where $M = \max\left\{|q^2 + a_1 q - 2a_2 + a_1 + 1|,\ |q^2 + a_1 - a_1 q - 1|\right\}$.

Once a 2-tuple or 4-tuple $(m_0, ..., m_{n-2})$ has been generated whose components are smaller than r, scalar multiplication can be performed by multi-exponentiation techniques. The problems arise when generating the curves for cryptographic purposes. In fact, we want r to be $O(q^g)$ with the implied constant as close to 1 as possible, and this requires testing more curves. If r is too small, less points on the curve can be generated than desired. In odd characteristic this makes generating good curves even more difficult, but scalar splitting is slower than generating the multi-scalars directly, so this can be the preferred way. In our context, however, scalar splitting is much faster and thus probably the preferred way: note, however, that it requires

the implementation of a simple multiprecision integer library besides the binary arithmetic.

4. Security

To prevent Weil Descent attacks [20,22] on the elliptic and hyperelliptic curves which we use to construct the trace zero varieties, the degree k of the definition field $\mathbb{F}_q$ over $\mathbb{F}_2$ will be chosen to be a prime.

According to the analysis in [15,31] and [47] the security of a trace zero variety is comparable to that of a hyperelliptic curve of low genus g' over $\mathbb{F}_{q'}$ with $q' \sim (n-1)\frac{g}{g'}$ for $|G| \sim 128$ bits. In other words, these trace zero varieties are suitable for low security applications.

For the cases $n = 3, g = 2$ and $n = 5, g = 1$, we can not a priori exclude that G may be contained in the Jacobian of a curve of genus six. Therefore some index calculus variations whose performance is better than the square root of the size of the whole group must be taken into account. Note that these algorithms compute in the whole Jacobian and the size of the latter determines the complexity. In case of high security applications $|G| \sim 2^{256}$ we can estimate that the security is reduced by at most six bits.

At this point it is important to observe that the situation is different than in the case of low genus hyperelliptic curves. In [6] in order to compare curves of different genera the group sizes for the Jacobians of hyperelliptic curves of genera three and four had to be increased in order to correctly compare to curves of genera one and two. In fact, for curves of genus three and four, the fastest known attack is a double large prime variation of the index calculus algorithm [19,38]. Ignoring logarithmic terms, this attack requires $O(q^{2-2/g})$ group operations for a genus g over of field of q elements. For a curve of genus three, resp. four over $\mathbb{F}_q$, this means $O(q^{4/3})$, resp. $O(q^{3/2})$ group operations. Ignoring also the constants, we see that to compare with an elliptic curve over a field of n bits, we need a field of $n/2$ bits for genus two, $3n/8$ bits for genus three and $n/3$ bits for genus four. At the current state of research, such losses do not seem to affect trace zero varieties.

5. Implementation

5.1. *Base fields*

Our implementation of binary fields is the one described in the paper [5] (see also [6]). For prime fields, the implementation is described in [1], which is used also in [4].

5.2. *Extension fields of small degree*

For binary fields the construction of small extensions follows a different approach with respect to the odd characteristic case, as described in [4].

If α is an algebraic element of degree n over $\mathbb{F}_2$, it is also an algebraic element of degree n over $\mathbb{F}_{2^k}$ for every k prime to n. This is our case, since n is 3 or 5 and k is prime and greater than n.

Hence we can construct $\mathbb{F}_{q^n}$ with an irreducible polynomial over $\mathbb{F}_2$, which is fixed and independent from the ground field $\mathbb{F}_q$. We used $f(X) = X^3 + X + 1$ for $n = 3$ and $f(X) = X^5 + X^2 + 1$ for $n = 5$. In other words, we construct $\mathbb{F}_{q^n}$ as the composite of the two fields $\mathbb{F}_q$ and $\mathbb{F}_{2^n}$ over $\mathbb{F}_2$.

Let $\mathbb{F}_{q^n} = \mathbb{F}_q(\alpha)$. The set of elements $1, \alpha, \ldots, \alpha^{n-1}$ is a base of $\mathbb{F}_{q^n}$ as a vector space over $\mathbb{F}_q$ and we represent a generic element of $\mathbb{F}_{q^n}$ as a polynomial in α of degree lower than n:

$$\mathbb{F}_{q^n} \ni a = \sum_{i=0}^{n-1} a_i \alpha^i = a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0 \ .$$

Sums in $\mathbb{F}_{q^n}$ are done componentwise, so they cost n additions in the base field.

For reductions modulo $f(X)$, since this polynomial is fixed and depends only on the degree n, we can precompute the expressions of the necessary powers of α larger than $n-1$. Because of the choice of α, these are polynomials in α of degree at most $n-1$ and with coefficients in $\mathbb{F}_2$. Therefore, reducing a polynomial in α to one of degree at most $n-1$ requires only to perform a few additions in $\mathbb{F}_q$, and no field multiplications.

In what follows we shall use the following symbols to denote field operations:

- `m`,`s`,`i`= multiplication, resp. squaring, inversion in $\mathbb{F}_q$;
- `M`,`S`,`I`= multiplication, resp. squaring, inversion in $\mathbb{F}_{q^n}$.

5.2.1. *Multiplication*

Multiplication of two degree 2 polynomials over $\mathbb{F}_q$ is done with the Karatsuba method, and cost 6 `m`, instead of the 9 required by naïve (schoolbook) multiplication:

$$\begin{aligned}(a_0 + a_1\alpha + a_2\alpha^2)(b_0 + b_1\alpha + b_2\alpha^2) = \\ = a_0b_0 + \big((a_0 + a_1)(b_0 + b_1) - a_0b_0 - a_1b_1\big)\alpha \\ + \big((a_0 + a_2)(b_0 + b_2) - a_0b_0 - a_2b_2 + a_1b_1\big)\alpha^2 \\ + \big((a_1 + a_2)(b_1 + b_2) - a_1b_1 - a_2b_2\big)\alpha^3 + (a_2b_2)\alpha^4 \ .\end{aligned}$$

Now observe that $\alpha^3 = \alpha+1$ and $\alpha^4 = \alpha^2+\alpha$ and we see that multiplications in $\mathbb{F}_{q^3}$ are performed by 6 $\mathtt{m}$ and a few additions.

For degree 5 extensions, put $\xi = \alpha^3$ and define the following quantities: $A_0 = \sum_{i=0}^{2} a_i\alpha^i$, $A_1 = a_3 + a_4\alpha$, $B_0 = \sum_{i=0}^{2} b_i\alpha^i$ and $B_1 = b_3 + b_4\alpha$. Using Karatsuba,

$$\begin{aligned} a \cdot b &= (A_0 + A_1\xi)(B_0 + B_1\xi) \\ &= A_0B_0 + \big((A_0 + A_1)(B_0 + B_1) - A_0B_0 - A_1B_1\big)\xi + A_1B_1\xi^2 \\ &= \sum_{i=0}^{8} c_i\alpha^i \end{aligned}$$

and the last expression is reduced as

$$(c_0 + c_5 + c_8) + (c_1 + c_6)\alpha + (c_2 + c_5 + c_7 + c_8)\alpha^2 + (c_3 + c_6 + c_8)\alpha^3 + (c_4 + c_7)\alpha^4 \ .$$

The product A_1B_1 is computed with the Karatsuba method in 3 $\mathtt{m}$. The products A_0B_0 and $(A_0 + A_1)(B_0 + B_1)$ are again computed in the same way as the product for the case $n = 3$, in 6 $\mathtt{m}$. Notice that the term in α^4 in A_0B_0 and $(A_0 + A_1)(B_0 + B_1)$ is the same, so save one $\mathtt{m}$ and the total cost is $3 + 2 \times 6 - 1 = 14$ $\mathtt{m}$.

5.2.2. *Squaring*

Computing one $\mathtt{S}$ needs only n $\mathtt{s}$. For instance, for $n = 3$ and $a = a_0 + a_1\alpha + a_2\alpha^2$ we have

$$a^2 = (a_1 + a_2)^2\alpha^2 + a_2^2\alpha + a_0^2 \ .$$

5.2.3. *Frobenius Operation*

The Frobenius automorphism of $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$ has the following properties:

(1) It is $\mathbb{F}_q$-linear: $\sigma(\lambda a + \mu b) = \lambda\sigma(a) + \mu\sigma(b)$;
(2) It commutes with powers: $\sigma(a^r) = (a^r)^{2^k} = (a^{2^k})^r = (\sigma(a))^r$.

Thus, its computation can be reduced to the computation of $\sigma\alpha^i$ for $i \leq n-1$ which is, once again, a polynomial in α of degree at most $n-1$ and coefficients in $\mathbb{F}_2$. We observe that, for $q = 2^k$, the results depend only on the remainder of k modulo n. Since powering by q in $\mathbb{F}_q$ is the identity operation, a Frobenius operation is thus implemented by only a few additions.

5.2.4. *Field Inversion*

The performance of inversion is one of the strongest points of the arithmetic of extension fields.

To obtain the inverse of an element $a \in \mathbb{F}_{q^n}$ we can use two methods.

The first method is due to Kobayashi et al. [25] and consists in viewing the multiplication as a $\mathbb{F}_q$-linear map in $\mathbb{F}_{q^n}$ and determine its inverse. For instance, for $n = 3$ let $b = b_0 + b_1\alpha + b_2\alpha^2$ $(b_0, b_1, b_2 \in \mathbb{F}_q)$ be a^{-1} where $a = a_0 + a_1\alpha + a_2\alpha^2 \in \mathbb{F}_{q^3}$.
Now $ab = 1$, w.r.t. basis $\{1, \alpha, \alpha^2\}$ can be written as

$$A \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{where} \quad A = \begin{bmatrix} a_0 & a_2 & a_1 \\ a_1 & a_0 + a_2 & a_1 + a_2 \\ a_2 & a_1 & a_0 + a_2 \end{bmatrix} .$$

Hence

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_0 & a_2 & a_1 \\ a_1 & a_0 + a_2 & a_1 + a_2 \\ a_2 & a_1 & a_0 + a_2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = (\det A)^{-1} \begin{bmatrix} a_0^2 + a_1^2 + a_2^2 + a_1a_2 \\ a_2^2 + a_0a_1 \\ a_1^2 + a_2^2 + a_0a_2 \end{bmatrix} ,$$

where $\det A = (a_0(a_0^2 + a_1^2 + a_2^2 + a_1a_2) + a_1^3 + a_2(a_1a_2 + a_2^2))$.

The second method is sometimes called the Itoh-Tsujii inversion [21,23]. Write

$$a^{-1} = \frac{a^q a^{q^2} \cdots a^{q^{n-1}}}{\mathrm{N}(a)} ,$$

where we may take advantage of the Frobenius automorphism to efficiently compute the numerator, and where $\mathrm{N}(a)$ is the norm from $\mathbb{F}_{q^n}$ to $\mathbb{F}_q$ of a. Observe that the norm is obtained just by multiplying the numerator of the expression for the inversion by a and taking into account the fact that the result is an element $\mathbb{F}_q$ to avoid unnecessary operations.

For $n = 3$ the two methods have the same cost, which is $3\,\mathtt{s} + 9\,\mathtt{m} + 1\,\mathtt{i}$; instead for $n = 5$ the number of terms using the first approach grows quickly, so we only implemented the latter.

For $n = 5$ we have: $a^{-1} = a^q a^{q^2} (a^q a^{q^2})^{q^2} / \mathrm{N}(a)$. Since the computation of a^q is for free, the computation of the numerator costs $2\,\mathtt{M}$, *i.e.* $28\,\mathtt{m}$. Computing $\mathrm{N}(a) \in \mathbb{F}_q$ costs $6\,\mathtt{m}$ (it is easy to verify that

$$\left(\sum_{i=0}^{4} a_i\alpha^i\right)\left(\sum_{i=0}^{4} b_i\alpha^i\right) = a_0b_0 + a_1b_4 + a_2b_3 + a_3b_2 + a_4b_1 + a_4b_4$$

if the result is known beforehand to lie in $\mathbb{F}_q$), and multiplying by $1/\mathrm{N}(a)$ costs $5\,\mathtt{m}$. The total cost of an inversion is then $39\,\mathtt{m} + 1\,\mathtt{i}$. Note, however, that the latter cost is in fact only an upper estimate: for instance, the multiplication by $1/\mathrm{N}(a)$ is one of the many places where the technique of *Sequential Multiplications* from [5] can be put to use: for the fields we work with, the 5 multiplications in $\mathbb{F}_q$ with a common operand in fact costs less than $3.5\,\mathtt{m}$. Similar simplifications can be made to the computations of products in $\mathbb{F}_{q^5}$ (for which in fact in some cases the schoolbook method with $25\,\mathtt{m}$ was faster than Karatsuba if the $25\,\mathtt{m}$ are collected in 5 groups of $5\,\mathtt{m}$ each with a common multiplicand) and to norm computation.

5.2.5. *Performance*

In Table 1 we point out the cost of an operation in the extension field in terms of operations in the base field, both in even and odd characteristic and for the two interesting extension degrees. The considered operations are squaring, multiplication, inversion, Frobenius operation.

Table 1. Costs of $\mathbb{F}_{q^n}$-operations in terms of $\mathbb{F}_q$-operations.

	$n=3$		$n=5$	
	$\mathrm{char}\,\mathbb{F}_q = 2$	$\mathrm{char}\,\mathbb{F}_q > 2$	$\mathrm{char}\,\mathbb{F}_q = 2$	$\mathrm{char}\,\mathbb{F}_q > 2$
S	$3\,\mathtt{s}$	$3\,\mathtt{s} + 3\,\mathtt{m}$	$5\,\mathtt{s}$	$5\,\mathtt{s} + 10\,\mathtt{m}$
M	$6\,\mathtt{m}$	$6\,\mathtt{m}$	$14\,\mathtt{m}$	$14\,\mathtt{m}$
I	$3\,\mathtt{s} + 9\,\mathtt{m} + 1\,\mathtt{i}$	$3\,\mathtt{s} + 9\,\mathtt{m} + 1\,\mathtt{i}$	$39\,\mathtt{m} + 1\,\mathtt{i}$	$50\,\mathtt{m} + 1\,\mathtt{i}$
σ	negligible	$2\,\mathtt{m}$	negligible	$4\,\mathtt{m}$

Table 2 shows the implementation results on a PowerPC G4 1.5 Ghz: timings are in microseconds. We compared small extensions and base fields performance; we grouped together fields suitable to define curves (ordinary or TZV) with the same security strength (medium or high). The notation $k \times n$ denotes the degree n field extension of the field $\mathbb{F}_{2^k}$.

In general we can see that multiplication is always slower in the small extension and the inversion is faster only for fields suitable for elliptic curves. It should also be noticed that the ratio inversion–multiplication (I/M) is quite low, and this motivate our choice to implement TZV only in affine coordinates.

6. Construction of Varieties and Generators

For cryptographic applications we need to be able to construct TZV whose order is prime or almost prime, and a generator of the subgroup G_0 of large

Table 2. Benchmarking of implementations of some finite fields. Timings in microseconds on a PowerPC G4 1.5 Ghz.

	For normal ($\approx$ 160-bit) security curves					For high ($\approx$ 190-bit) security curves				
Genus:	$g=1$			$g=2$		$g=1$			$g=2$	
Field:	83×3	41×5	163	41×3	83	89×3	47×5	191	47×3	89
σ	0.043	0.050	–	0.031	–	0.038	0.055	–	0.033	–
S	0.170	0.120	0.057	0.110	0.029	0.177	0.120	0.057	0.112	0.029
M	1.649	1.434	0.858	0.639	0.292	1.643	1.458	0.858	0.633	0.292
I	3.517	4.446	6.023	1.420	1.028	3.421	4.446	6.619	1.520	1.129
I/M	2.078	2.995	6.685	2.118	3.203	2.034	2.938	7.462	2.280	4.342

group order.

In order to find a good curve we generate random ones with random coefficients until we find one with the correct order. By the Hasse-Weil Theorem, we thus need only to fix the field size in order to get groups whose orders are in a relatively narrow interval. To compute the characteristic polynomial of an elliptic curve we use MAGMA, and to find good curves of genus two we modified a software package written by Frederic Vercauteren.

Once a curve has been found, we need only take a random divisor D on it and observe that $E = D + \sigma(D)$ lies in the TZV. To ensure that it has the right order we replace E by its multiple by the index of G_0 in G: if the result is the neutral element of the group, we try with another choice of D.

We report in Appendix A the examples of curves used in our experiments. Notice that, in all the examples, finding good groups required only a few minutes of computation on old workstations.

7. Experimental results

In this section we present the experimental results. In particular we show the performance of TZV with respect to ordinary curves and provide a comparison between even and odd characteristic.

The test have been done using the curves presented in the previous section, that provide security levels of approximately 160 and 190 bits. We benchmarked these curves on two platforms: an AMD Athlon K6-2 at 1GHz and Motorola PowerPC G4 at 1.5GHz. The first machine is the same used in [1]: even though more modern Intel(-compatible) CPUs are available and our finite field library supports them, we ran the tests on that CPU to provide a comparison with already published results, in particular to show the performance differences between groups in even and odd characteristic. The second platform is widely used in embedded systems, for which lightweight structures based on small(er) fields, such as TZV, are attractive.

7.1. *Scalar multiplication techniques*

We used several scalar multiplication techniques to perform our tests. The first group of scalar multiplication techniques consists in writing a scalar m in the form $m = \sum_{i=0}^{n} d_i 2^i$, where the digits d_i belong to a suitable digit set $\mathcal{D}$ and in the evaluation of $m \cdot D$ as the sum $\sum_{i=0}^{n} d_i 2^i D$ via a Horner scheme. The recodings of the scalar (which differ in the digit set and in the algorithms used to generate the expansion) are the binary representation of the scalar, the Non-Adjacent Form (NAF) [42], and signed windowing methods (w-NAF) [14].

We also split the scalar using the method described in Theorem 3.1, and performed the multiplication $m \cdot D$ by computing $\sum_{i=0}^{n-2} m_i s^i \bmod \ell$ via interleaved multiple scalar multiplication (described by Möller in [37], but in fact an older idea which has been rediscovered several times: see [33], a two base case appears in [44], and indeed the principles can be traced back to the work of Pippenger [41]; see also [7]). To recode the smaller scalars m_i we used the binary representation, the NAF, windowed methods with interleaved multiple scalar multiplication, and in the case of curves with $n = 3$ also Solinas' Joint Sparse Form (JSF) [46].

7.2. *Comparison between ordinary curves and TZV*

In Tables 3 and 4 we report results about scalar multiplication timing in different groups: `ec` and `hec` for plain elliptic and genus 2 hyperelliptic curves, `ec3`, resp. `ec5`, for TZV from elliptic curves over extension fields of degrees 3, resp. 5, and `hec3` for TZV constructed from genus 2 hyperelliptic curves over a degree 3 extension field.

The scalar multiplication has been computed with binary (mul) and NAF (mul NAF) representation for the scalar. For TZV we also report results for the multi–multiplication (m-mul) achieved using the scalar splitting. In this case we also perform different tests, using different representation for the scalar: binary, NAF and, in the case of degree 3 extensions, JSF.

The last lines show the timing for mul and m-mul based on the sliding window representation for scalars (w-NAF); between parentheses we report the optimal windows size.

From the results, we can do the following remarks:

(1) TZV constructed from elliptic curves have the most efficient scalar mul-

Table 3. Comparison between ordinary curves and TZV at 160 bits. Timings in microseconds.

AMD Athlon K6 1Ghz

	`ec`	`ec3`	`ec5`	`hec`	`hec3`
$k =$	163	83	41	83	41
mul	5831	4485	3723	3014	4324
mul NAF	5185	3884	3453	2628	3139
mul w-NAF	(3)4647	(3)3570	(4)3113	(3)2278	(4)2700
m-mul	–	2511	1964	–	2727
m-mul NAF	–	2269	1534	–	2260
m-mul JSF	–	2162	–	–	2099
m-mul w-NAF	–	(4)2054	(4)1157	–	(4)1857

PowerPC G4 1.5 Ghz

	`ec`	`ec3`	`ec5`	`hec`	`hec3`
$k =$	163	83	41	83	41
mul	1914	1820	1898	1023	2403
mul NAF	1666	1580	1758	865	1730
mul w-NAF	(3)1494	(4)1442	(3)1575	(4)704	(4)1472
m-mul	–	1032	1008	–	1520
m-mul NAF	–	910	784	–	1256
m-mul JSF	–	914	–	–	1167
m-mul w-NAF	–	(4)871	(4)595	–	(4)1027

tiplication;

(2) In the case of genus two hyperelliptic curves, TZV are less efficient. This is due to the higher number of field multiplications in the formulæ for genus two curves with respect to the elliptic curves. Since multiplications in extension fields are less efficient (cf. Table 2), their impact vanifies the advantages of the fast inversion;

(3) Using multiple scalar multiplication with scalar splitting, performance increases by a factor 2 over extensions of degree 3 and by a factor of 2.5-3 over extensions of degree 5 with respect to the naive scalar multiplication using only one full length scalar. Hence `ec3` and `ec5` perform much better than `ec`, while `hec3`, though still faster than `hec`, achieves a lower overall speed–up: in the case of AMD Athlon K6 1Ghz it is enough to achieve better performance than `hec;` in the case of PowerPC G4 1.5 Ghz it is not, but we can see a decreasing gap passing from 160 to 190 bits, due to the decreasing relative cost of the inversion.

In conclusion TZV, and in particular `ec5` from the performance point of view, come out as a good alternative to ordinary elliptic and hyperelliptic curves, since they shows interesting performance, both for the scalar multiplication and the construction of groups of cryptographic relevant size.

Table 4. Comparison between ordinary curves and TZV at 190 bits. Timings in microseconds.

AMD Athlon K6 1Ghz

	`ec`	`ec3`	`ec5`	`hec`	`hec3`
$k =$	191	89	47	89	47
mul	7168	5167	4691	6037	4988
mul NAF	6764	4476	4063	4772	4072
mul w-NAF	(4)6064	(5)4090	(5)3669	(5)4162	(4)3247
m-mul	–	2861	2099	–	3023
m-mul NAF	–	2529	1740	–	2583
m-mul JSF	–	2422	–	–	2502
m-mul w-NAF	–	(3)2350	(3)1390	–	(4)2135

PowerPC G4 1.5 Ghz

	`ec`	`ec3`	`ec5`	`hec`	`hec3`
$k =$	191	89	47	89	47
mul	2425	2051	2338	1473	2716
mul NAF	2290	1782	2017	1163	2188
mul w-NAF	(4)2064	(5)1636	(5)1801	(4)974	(4)1729
m-mul	–	1145	1039	–	1665
m-mul NAF	–	1000	854	–	1424
m-mul JSF	–	962	–	–	1379
m-mul w-NAF	–	(4)935	(4)681	–	(3)1177

7.3. *Comparison between even and odd characteristic*

We conclude with a comparison between TZV over binary fields and over fields of odd characteristic (as described in [4]). Data for ordinary curves in characteristic p is obtained from [1].

Tables 5 and 6 presents the scalar multiplication timings for the five curves analyzed, with binary, NAF and, when available, JSF representations for the scalar (or split scalars).

Table 5. Comparison between TZV at 160 bits for $\operatorname{char} K = 2$ or $\operatorname{char} K = p > 2$. Timings in microseconds.

AMD Athlon K6 1Ghz

		`ec`	`ec3`	`ec5`	`hec`	`hec3`
	bits	160	80	40	80	40
BIN	char $= 2$	5831	2511	1964	3014	2727
	char $= p$	3074	1320	1035	1899	2415
NAF	char $= 2$	5185	2269	1534	2628	2260
	char $= p$	2701	1245	1020	1706	2265
JSF	char $= 2$	–	2162	–	–	2099
	char $= p$	–	1125	–	–	–

PowerPC G4 1.5 Ghz

		`ec`	`ec3`	`ec5`	`hec`	`hec3`
	bits	160	80	40	80	40
BIN	char $= 2$	1914	1032	1008	1023	1520
	char $= p$	3852	1860	1000	2545	2660
NAF	char $= 2$	1666	910	784	865	1256
	char $= p$	3480	1580	920	2298	2320
JSF	char $= 2$	–	914	–	–	1167
	char $= p$	–	1540	–	–	–

Table 6. Comparison between TZV at 190 bits for $\operatorname{char} K = 2$ or $\operatorname{char} K = p > 2$. Timings in microseconds.

AMD Athlon K6 1Ghz

		ec	ec3	ec5	hec	hec3
	bits	190	90	48	90	48
BIN	char = 2	7168	2861	2099	6037	3023
	char = p	5385	1845	1545	2546	3060
NAF	char = 2	6764	2529	1740	4772	2583
	char = p	4809	1680	1470	2265	2805
JSF	char = 2	–	2422	–	–	2502
	char = p	–	1575	–	–	–

PowerPC G4 1.5 Ghz

		ec	ec3	ec5	hec	hec3
	bits	190	90	48	90	48
BIN	char = 2	2425	1145	1039	1473	1665
	char = p	6181	2420	1420	3270	3220
NAF	char = 2	2290	1000	854	1163	1424
	char = p	5520	2240	1380	2964	3000
JSF	char = 2	–	962	–	–	1379
	char = p	–	2060	–	–	–

From the data we can observe similar results for the curves, both over binary or prime fields. However odd characteristic shows better performance on AMD Athlon K6 1Ghz, while even characteristic is better on PowerPC G4 1.5 Ghz, in particular for genus two hyperelliptic curves. The main reason for this discrepancy is the inefficiency of the integer multiplication unit on the PowerPC, while on the other hand the AMD processor has a much faster multiplier that can be better pipelined.

8. Conclusions

In this paper we consider trace zero varieties (TZV) over binary fields, showing that they share some of the advantages of TZV over fields of odd characteristic, beside the many difference in the construction: base field, extension field and underlying curve arithmetic.

TZV deliver better scalar multiplication performance than elliptic curves; furthermore the varieties constructed from elliptic curves over degree 5 extension fields come out as one of the most efficient groups, suitable to build cryptographic systems based on the discrete logarithm problem: for the same group size they are almost 3 times faster than elliptic curves.

Trace zero varieties however require a larger number of bits to represent group elements for the same group size with respect to elliptic and hyperelliptic curves, assuming that affine coordinates are used for the latter: 50%, resp. 25% more bits if the extension degree is 3, resp. 5. This means that in order to get vastly superior performance one has to sacrifice some bandwidth.

TZV also simplify the construction of groups of cryptographic relevant size. The order of the group can be computed from the characteristic polynomial of the Frobenius endomorphism of the curve. At comparable level of the security provided (*i.e.* comparable group size), being the ground field

of (the curve under) a TZV smaller than the corresponding ground field of an ordinary curve, it is simpler to compute the characteristic polynomial (and hence the group size) for TZV.

Future research will include the use of point halving on trace zero varieties constructed from elliptic curves. Point halving, i.e. the inverse operation of point doubling, was introduced independently by Knudsen [24] and Schroeppel [43] (then revisited by Fong et al. [16], see also [2] for the impact of faster square root extraction on point halving) allowing for a faster arithmetic.

Trace Zero Varieties are therefore very interesting groups for the design of cryptographic systems around the discrete logarithm problem.

Acknowledgments

The authors would like to thank Gerhard Frey for initiating their research on this subject and his constant interest. We are also grateful to Ottavio Rizzo for his support, and to David Kohel, Tanja Lange, and Nicolas Thériault for interesting discussions and remarks.

References

1. R. M. Avanzi, *Aspects of Hyperelliptic Curves over Large Prime Fields in Software Implementations*. In: *CHES 2004*, LNCS volume 3156, pages 148–162. Springer, 2004.
2. R. Avanzi. *Another Look at Square Roots and Traces (and Quadratic Equations) in Fields of Even Characteristic*. In: *SAC 2007*, LNCS volume 4876, Springer 2007 (page numbers not currently available). See also: IACR ePrint 2007/103, `http://eprint.iacr.org/2007/103`.
3. R. Avanzi, H. Cohen, C. Doche, G. Frey, T. Lange, K. Nguyen, and F. Vercauteren. *The Handbook of Elliptic and Hyperelliptic Curve Cryptography*. CRC Press, 2005.
4. R. M. Avanzi and T. Lange, *Cryptographic Applications of Trace Zero Varieties*. Preprint.
5. R. M. Avanzi and N. Thériault, *Effects of Optimizations for Software Implementations of Small Binary Field Arithmetic*. In: *WAIFI 2007*. LNCS volume 4547, pages 69–84. Springer, 2007.
6. R. M. Avanzi, N. Thériault, and Z. Wang, *Rethinking Low Genus Hyperelliptic Jacobian Arithmetic over Binary Fields: Interplay of Field Arithmetic and Explicit Formulæ*. CACR Technical report CACR 2006-07. `http://www.cacr.math.uwaterloo.ca/ techreports/2006/ tech_reports2006.html`.
7. D.J. Bernstein, *Pippenger's exponentiation algorithm*. Preprint. Available from `http:// cr.yp.to`

8. G. Blady, *Die Weil-Restriktion elliptischer Kurven in der Kryptographie*. Master's thesis, University Essen, 2002.
9. Wieb Bosma, John Cannon, and Catherine Playoust, *The* Magma *Algebra System I: The User Language*, J. Symbolic Comput. **24**, pages 235–265 (1997).
10. B. Byramjee, and S. Duquesne, *Classification of genus* 2 *curves over* $\mathbb{F}_{2^n}$ *and optimization of their arithmetic*. Preprint, 2004.
11. E. Cesena. *Varietá a Traccia Zero su Campi Binari: Applicazioni Crittografiche. (Trace Zero Varieties over Binary Fields: Cryptographic Applications.)* Master's Thesis. Universitá degli Studi di Milano. 2005. In Italian.
12. D. G. Cantor, Computing in the Jacobian of hyperelliptic curves. *Mathematics of Computation*, 48 (177), pages 95–101 (1987).
13. Y. Choie and D. Yun, *Isomorphism classes of hyperelliptic curves of genus* 2 *over* $\mathbb{F}_q$. In: *ACISP 2002*, LNCS volume 2384, 190–202. Springer–Verlag, 2002.
14. H. Cohen, A. Miyaji and T. Ono, *Efficient elliptic curve exponentiation*. In: *ICICS '97*, LNCS 1334, pages 282–290. Springer–Verlag, 1997.
15. C. Diem, and J. Scholten, *Cover Attacks – A report for the AREHCC project*. see `http://www.arehcc.org`, 2003.
16. K. Fong, D. Hankerson, J. Lopez, and A. Menezes, *Field inversion and point halving revisited*. In *IEEE Transactions on Computers*, pages 1047–1059, 2004.
17. G. Frey, *How to disguise an elliptic curve*. Talk at Waterloo workshop on the ECDLP, 1998.
`http://cacr.math.uwaterloo.ca/conferences/1998/ecc98/slides.html`.
18. G. Frey, *Applications of arithmetical geometry to cryptographic constructions*. In *Finite fields and applications (Augsburg, 1999)*, pages 128–161. Springer–Verlag, 2001.
19. P. Gaudry, E. Thomé, N. Thériault and C. Diem, *A double large prime variation for small genus hyperelliptic index calculus*. Math. Comp. **76**, pages 475–492 (2007) See also: IACR ePrint 2004/153, available from `http://eprint.iacr.org/2004/153`
20. P. Gaudry, F. Hess, and N. P. Smart, *Constructive and destructive facets of Weil descent on elliptic curves*, J. Cryptology **15** (1), pages 19–46.
21. J. Guajardo, and C. Paar, *Itoh-Tsujii Inversion in Standard Basis and Its Application in Cryptography and Codes. Designs, Codes and Cryptography*, 25, pages 207–216 (2002).
22. F. Hess, and N. Smart, *Extending the GHS Weil-descent attack*, Advances in Cryptology – Eurocrypt 2002, LNCS 2332, pages 29–44. Springer-Verlag, 2002.
23. T. Itoh, and S. Tsujii, *A Fast Algorithm for Computing Multiplicative Inverses in* $GF(2^m)$ *using Normal Bases. Information and Computation*, 78, pages 171–177 (1988).
24. E. Knudsen, *Elliptic scalar multiplication using point halving*. In *Advances in cryptology – ASIACRYPT 1999*, LNCS 1716, pages 135–149. Springer-Verlag, 1999.

25. T. Kobayashi, H. Morita, K. Kobayashi, and F. Hoshino, *Fast elliptic curve algorithm combining frobenius map and table reference to adapt to higher characteristic.* In *Theory and Application of Cryptographic Techniques*, pages 176–189 (1999).
26. N. Koblitz, *Elliptic curve cryptosystems. Mathematics of Computation*, 48 (177), pages 203–209 (1987).
27. N. Koblitz, *Algebraic aspects of cryptography.* Springer-Verlag, 1998.
28. N. Koblitz, *Hyperelliptic cryptosystems. J. Cryptology*, 1, pages 139–150 (1989).
29. N. Koblitz, *CM-curves with good cryptographic properties.* In: *CRYPTO 1991*, LNCS 576, pages 279–287. Springer-Verlag, 1991.
30. T. Lange, *Efficient Arithmetic on Hyperelliptic Curves.* PhD. thesis, University Essen, 2001.
31. T. Lange, *Trace zero subvarieties of genus 2 curves for cryptosystems.* J. Ramanujan. Math. Soc. **19** (1), pages 15–33, 2004.
32. T. Lange and M. Stevens, *Efficient doubling for genus two curves over binary fields.* In *Selected Areas in Cryptography – SAC 2004*, LNCS 3357, pages 170–181. Springer-Verlag, 2005.
33. C.H. Lim, *Efficient multi-exponentiation and application to batch verification of digital signatures.* Unpublished manuscript. August 2000.
See: `http://dasan.sejong.ac.kr/~chlim/english_pub.html`
34. A. Menezes, Y.-H. Wu, and R. Zuccherato, *An Elementary Introduction to Hyperelliptic Curves.* In [27], pages 155–178.
35. J.-F. Mestre, *Applications de l'AGM au calcul du nombre de points d'une courbe de genre* 1 *ou* 2 *sur* $\mathbb{F}_{2^n}$. Talk given to the Séminaire de Cryptographie de l'Université de Rennes, March 2002. See
`http://www.maths.univ-rennes1.fr/crypto/2001-02/Mestre2203.html`
for a brief summary.
36. V. S. Miller, *Use of elliptic curves in cryptography.* In *Advances in Cryptology – Crypto 1985*, LNCS 218, pages 417–426. Springer-Verlag, 1986.
37. B. Möller, *Algorithms for Multi-exponentiation.* In *Selected Areas in Cryptography – SAC 2001*, LNCS 2259, pages 165–180. Springer-Verlag, 2001.
38. K. Nagao, *Improvement of Thériault Algorithm of Index Calculus for Jacobian of Hyperelliptic Curves of Small Genus.* IACR ePrint 2004/161.
Available from `http://eprint.iacr.org/2004/161`
39. National Institute of Standards and Technology. *Recommended Elliptic Curves for Federal Government Use.* NIST Special Publication, July 1999.
Available from: `http://csrc.nist.gov/csrc/fedstandards.html`
40. N. Naumann, *Weil-Restriktion abelscher Varietäten.* Master's thesis, University Essen, 1999.
41. N. Pippenger, *On the evaluation of powers and related problems (preliminary version).* 17th Annual Symp. on Foundations of Comp. Sci., IEEE Computer Society, 1976, pages 258–263.
42. G. W. Reitwiesner, *Binary arithmetic.* Advances in Computers **1**, pages 231–308 (1960).
43. R. Schroeppel, *Elliptic curve point halving wins big.* 2nd Midwest Arithmeti-

cal Geometry in Cryptographic Workshop, Urbana, Illinois, November 2000.
44. S.G. Sim, and P.J. Lee, *An efficient implementation of two-term exponentiation in elliptic curves* In *Japan–Korea Joint Workshop on Information Security and Cryptology (JW-ISC 2000)*, Jan. 25-26, 2000, Naha, Okinawa, Japan, pages 61–68.
45. J. A. Solinas, *Efficient Arithmetic on Koblitz Curves*. Designs, Codes and Cryptography, Vol. 19, No. 2/3, pages 125–179 (2000).
46. J.A. Solinas, *Low-Weight Binary Representations for Pairs of Integers*. Centre for Applied Cryptographic Research, University of Waterloo, Combinatorics and Optimization Research Report **CORR 2001-41**, 2001. Available from: `http://www.cacr.math.uwaterloo.ca/techreports/2001/corr2001-41.ps`
47. N. Thériault, *Index calculus attack for hyperelliptic curves of small genus*. In *Advances in Cryptology - ASIACRYPT 2003*, LNCS 2894, pages 75–92. Springer–Verlag, 2003
48. A. Weimerskirch, *The Application of the Mordell-Weil Group to Cryptographic Systems*. Master's thesis, Worchester polytechnic institute, 2001.

Appendix A. Examples

In this Appendix we list examples from the actual curves used in our experiments. The construction methodology is addressed in Section 6. Notice that, in all the examples, finding good groups required only a few minutes of computation on old workstations.

Elliptic, resp. hyperelliptic curves have equations of the form (2), resp. (3). Let s be the constant such that $s \cdot D = \sigma(D)$ on the TZV, r be the bound given by Theorem 3.2, ℓ the order of the group G_0 and c the cofactor (in the whole elliptic curve, in the divisor class group, or in trace-zero variety). P, resp. D denotes the random point, resp. divisor, picked on the curve, and Q, resp. E the generator of the cryptographically interesting group.

Field elements are given by the hexadecimal strings corresponding to the internal representation, the least significant digits being to the right. For instance, let $K = \mathbb{F}_{2^7} = \mathbb{F}_2(\beta)$ with $\beta^7 + \beta + 1 = 0$. The elements of K will be represented by hexadecimal strings, for instance the hexadecimal string `0x32` is (1100 0100) in binary notation, and represents the field element $1 + \beta + \beta^5 = \beta^{19}$.

Appendix A.1. *160-Bit Groups*

Appendix A.1.1. *Elliptic curve on* $\mathbb{F}_{2^{163}}$

The field is defined by the standard NIST [39] polynomial $X^{163} + X^7 + X^6 + X^3 + 1$.

$$
\begin{aligned}
a &= \texttt{0} \ , \ b = \texttt{0x1A76A76F7B346488C9A35692BA68599FCCDFF9253} \\
\ell &= 2923003274661805836407369488607882210206410807107 \ , \ c = 4 \\
P &= (\texttt{0xC4C83506090F50668DC9712FB341E40404991E015}, \\
&\quad \texttt{0x0F40F3EDF53B5F1AB2292AEE8F89491B58545E2DB}) \\
Q &= 4P \\
&= (\texttt{0xBA37855C3EDE2A548259D15103D22D93B2F8523A6}, \\
&\quad \texttt{0x6B52552D6CCEFDBEA0771E24EDAAA3F3FA1626141})
\end{aligned}
$$

Appendix A.1.2. *TZV over* $\mathbb{F}_{2^{83}}$ *with parameters* $g = 1$, $n = 3$

$\mathbb{F}_{2^{83}}$ is defined by the polynomial $X^{83} + X^7 + X^4 + X^2 + 1$.

$$
\begin{aligned}
a &= \texttt{0} \ , \ b = \texttt{0x1000000010000000042} \\
\ell &= 9353610478921145403801704743947372123090419683579 9 \\
s &= 52717605029603870685346358367337123607533507700722 \\
r &= 96714065569170333976494 07 \\
\tilde{P} &= ([\texttt{0xEDB6F92A2EA71F604D95 0xE384C42131F580D858611} \\
&\qquad \texttt{0x95B3DEDEC0414F10C4C83}], \\
&\quad [\texttt{0xAB38D1A19CCE1769FF067 0x06642ED8A050A85391AC5} \\
&\qquad \texttt{0x495194FD9ED1A2E7CCAA7}]) \\
\tilde{S} &= \sigma(\tilde{P}) = \\
&\quad ([\texttt{0xEDB6F92A2EA71F604D950 0x76371AFFF1B4CFC89CA92} \\
&\qquad \texttt{0xE384C42131F580D858611}], \\
&\quad [\texttt{0xAB38D1A19CCE1769FF067 0x4F35BA253E810AB45D062} \\
&\qquad \texttt{0x06642ED8A050A85391AC5}]) \\
\tilde{Q} &= \tilde{P} \ominus \sigma(\tilde{P}) = \\
&\quad ([\texttt{0x10DB4388B93301F4D4656 0xCE290BD4544646EBF51F6} \\
&\qquad \texttt{0x10FDAE7FF034D341BC7A3}],
\end{aligned}
$$

[0xDC53D4F26C72BCC645F52 0x05439588DE86A7328AA97
0xEA099914821C1AEB94604])

Appendix A.1.3. *TZV over $\mathbb{F}_{2^{41}}$ with parameters $g = 1$, $n = 5$*

$\mathbb{F}_{2^{41}}$ is defined by the polynomial $X^{41} + X^3 + 1$.

$a =$ 0 , $b =$ 0x0000400028

$\ell = 233840240720611172362151321949161174780039413297 01$

$s = 859689839731687837068346974751971996174147121558 2$

$r = 11206585$

$\tilde{P} =$ ([0xCDAFD8B0131 0x53A8AF48421 0x44F5A3920E1
0xF67993BA441 0x34A0F297D41],
[0x7E36B87B880 0x976D00397A0 0x9A21D55E041
0xD77D34B65F1 0x2E6F45751B0])

$\tilde{S} = \sigma(\tilde{P}) =$ ([0xF90F2A27C7 0xF67993BA441 0x67085DDF96
0xC2D9612D9 0x44F5A3920E1],
[0x5059FD0E93 0xD77D34B65F1 0xB902454C61
0xF91271C3441 0x9A21D55E041])

$\tilde{Q} = \tilde{P} \ominus \sigma(\tilde{P}) =$
([0x61A7B457D01 0x4B539AE2A31 0xA69138609A0
0x29AA7A28A90 0x9D910198BD1],
[0x5EF090213D1 0x691E2B69FA0 0x6576B231FC1
0x3E13973F631 0x233C039C141])

Appendix A.1.4. *Genus two hyperelliptic curve over $\mathbb{F}_{2^{83}}$*

$f_0 =$ 0x24132CDF7FC8FBDAA4DA , $f_2 =$ 1 ,

$f_3 =$ 0xFFE59AEF1B8E5FD3646C1

$\ell = 467680523946084570060257334802947903980347475558 11$, $c = 2$

$D = \langle x^2 +$ 0x89D6EC10E10487B5E3721 $x +$ 0xA5FF4CB5B115F3C2BD864,
0xB75D17A7D2A09015A9DC1 $x +$ 0x287BA5DE2133DDC3B9A64$\rangle$

$E = 2D = \langle x^2 +$
0x35DFF70E4C99F567C6BA6 $x +$ 0xC2275602FE2B755FAD714,

$\texttt{0xC4E6C2F02879D966B0602}\, x + \texttt{0xA3DDF3E70326B1AF04D47}\rangle$

Appendix A.1.5. *TZV over $\mathbb{F}_{2^{241}}$ with parameters $g = 2$, $n = 3$*

$f_0 = \texttt{0x540B020A91}$, $f_2 = 1$, $f_3 = \texttt{0x60905698DF1}$

$\ell = 233840392727907606896827009737486517737624992086 57$

$s = 289135746804572963956156494693406763288841787 5314$

$r = 4835703278454286981086417$

$\tilde{D} = \langle x + [\texttt{0x1895B3DEDE 0x414F10C4C81 0x06090F5066}],$
$[\texttt{0x53EB12A874 0x41741262CE 0x0493906BEF}]\rangle$

$\tilde{S} = \sigma(\tilde{D}) = \langle x + [\texttt{0x1895B3DEDE 0x47461F94AE1 0x414F10C4C81}],$
$[\texttt{0x53EB12A874 0x45E7820921 0x41741262CE}]\rangle$

$\tilde{E} = \tilde{D} \ominus \sigma(\tilde{D}) = \langle x^2 + [\texttt{0x0 0x06090F5066 0x47461F94AE1}]\, x +$
$+ [\texttt{0xF60614DB19 0xD83A289AF 0xB760A34826}],$
$[\texttt{0xCF8CB7DBEC1 0x712573AEA 0x7AF53377461}]\, x +$
$+ [\texttt{0x925BFFEB5 0x1C38024F751 0xCAA48998E1}]\rangle$

Appendix A.2. *190-Bit Groups*

Appendix A.2.1. *Elliptic curve on $\mathbb{F}_{2^{191}}$*

The field is defined by the NIST recommended [39] polynomial $X^{191} + X^9 + 1$.

$a = 1$, $b = \texttt{0xF03EDF4221A3657D20962F478BF6BDA2C0B960ECBEAA57A1}$

$\ell = 1569275433846670190958947355849253382243216021981899129049$,

$c = 2$

$P = (\texttt{0x1895B3DEDEC0414F10C4C83506090F50668DC9712FB341E4},$
$\texttt{0x1F96687BCC87070803C3C6372BD76BE153E0BF57FCE27E97})$

$Q = 2P = (\texttt{0xB1939FD4FB0D2CE94CD337481B0FAE274FA63409AF6DF675},$
$\texttt{0x1ED8914CB28E2A7A1986A460F4729FEF2A94C7DDA6B8A016})$

Appendix A.2.2. *TZV over $\mathbb{F}_{2^{97}}$ with parameters $g = 1$, $n = 3$*

$\mathbb{F}_{2^{97}}$ is defined by the polynomial $X^{97} + X^6 + 1$.

$a = 1$, $b =$ 0x9C4CF3F01C54F4E7A7FD7D781

$\ell = 25108406941546760434751930836434834976964168002830535418893$

$s = 10669767213033396095159100125257086651731213520935893527279$

$r = 158456325028528675187087900671$

$\tilde{P} = ($[0x04991E0152ED73795E539BC10 0xA5F59DEE5B30BE1F0C2AA3391
0x03566BA1DAC2D797FE636CD81],
[0x48C5B9E5B8B798F8FBEEE5A31 0xF1178A7BBCB1DACB04CC2C9A1
0xA535887A2267B3D51AC9736B1])

$\tilde{S} = \sigma(\tilde{P}) =$
([0x04991E0152ED73795E539BC10 0x03566BA1DAC2D797FE636CD81
0xA6A3F64F81F26988F249CFE1],
[0x48C5B9E5B8B798F8FBEEE5A31 0xA535887A2267B3D51AC9736B1
0x542202019ED6691E1E055FF1])

$\tilde{Q} = \tilde{P} \ominus \sigma(\tilde{P}) =$
([0xA1F68A3A248E955BCFC957FA1 0x58C53431D1D4C80C2E2E32A1
0x3D3102B2B31E1CE6CECC4F550],
[0x80EAAD5092B9A5C9A6ECB3510 0x364CB1D3144600AC81A894D1
0xF9C2A43BDA602825A063ED061])

Appendix A.2.3. *TZV over* $\mathbb{F}_{2^{47}}$ *with parameters* $g = 1$, $n = 5$

$\mathbb{F}_{2^{47}}$ is defined by the polynomial $X^{47} + X^5 + 1$.

$a = 0$, $b =$ 0x10004802

$\ell = 392318871423095243169862765890601877920346380738406349101$

$s = 183867226781893594594157220308313210869565837762023539 19$

$r = 35761759$

$\tilde{P} = ($[0x414F10C4C835 0x06090F506685 0xC9712FB341E4
0x0404991E0152 0xED73795E5393],
[0xB1D3D34DBF95 0xA1AABDAE73B6 0x09D8309E59D4
0x4B045454A757 0x6CDD3F7A7C31])

$\tilde{S} = \sigma(\tilde{P}) =$

$$\begin{aligned}
&([\texttt{0x654D4629DA42 0xE977E04052C1 0xCD75B6AD40B6}\\
&\qquad\qquad \texttt{0x2006CFF31325 0xEB7A760E3516}],\\
&\quad [\texttt{0xD4D6DCA99A70 0x27D96B2EDB66 0x42DC64CAFE83}\\
&\qquad\qquad \texttt{0x2E015BB082B2 0x6CDD3F7A7C31}])\\
\tilde{Q} = \tilde{P} \ominus \sigma(\tilde{P}) &=\\
&([\texttt{0x20AE8C7594F7 0x51FDA31983F4 0xE7237AA4FD56}\\
&\qquad\qquad \texttt{0xC86CB1516B21 0xAECA20300294}],\\
&\quad [\texttt{0x8805317028F5 0x3FC11C8D1E77 0xBFBBC6E2BDF2}\\
&\qquad\qquad \texttt{0xD05E8885C1C0 0x383CA9E92C96}])
\end{aligned}$$

Appendix A.2.4. *Genus two hyperelliptic curve over* $\mathbb{F}_{2^{97}}$

$$\begin{aligned}
f_0 &= \texttt{0x97411E31BEAEF44EA0D228E4}\ ,\ \ f_2 = 1\ ,\\
f_3 &= \texttt{0xD55942A3DD16DDA11AFE21A1}\ ,\ \ c = 4\\
\ell &= 627710173538666682931088974786241698430381421854821592237 9\ ,\\
D &= \langle x^2 +\\
&\quad \texttt{0xC1D9D62997C8DFAE7F470155}\ x + \texttt{0x8774458EDE8C4DF4B58AB3621},\\
&\quad \texttt{0xADE99329E2F1B6C430479E4D1}\ x + \texttt{0x0D37022B4D9CBBAE5B191A731}\rangle\\
E = 4D &= \langle x^2 +\\
&\quad \texttt{0x6EE871688004C82DF06D1076}\ x + \texttt{0xFB32FEB7C9AA3BDCDE385D41},\\
&\quad \texttt{0x91875AAF171D035D97D2EE3D}\ x + \texttt{0x900B248F3EB86AE7DD0850541}\rangle
\end{aligned}$$

Appendix A.2.5. *TZV over* $\mathbb{F}_{2^{47}}$ *with parameters* $g = 2$, $n = 3$

$$\begin{aligned}
f_0 &= \texttt{0x134F44BA1CF3}\ ,\ \ f_2 = 1\ ,\ \ f_3 = \texttt{0x1337CAD53025}\\
\ell &= 3923189389732390362618918952794526697136996158042653463 89\\
s &= 1433077383704080007532542666603325915220504817551595826 20\\
r &= 19807040628565647205297769285\\
\tilde{D} &= \langle x + [\texttt{0x4C3A52EDB6F1 0x2A2EA71F6045 0x958DE384C421}],\\
&\quad [\texttt{0xB1F6DE1C9327 0x3298791E4667 0xD078ABD1819}]\rangle\\
\tilde{S} = \sigma(\tilde{D}) &= \langle x + [\texttt{0x4C3A52EDB6F1 0xBFA3449BA464 0x2A2EA71F6045}],\\
&\quad [\texttt{0xB1F6DE1C9327 0xE2E0D2CFC7F7 0x3298791E4667}]\rangle
\end{aligned}$$

$$\tilde{E} = \tilde{D} \ominus \sigma(\tilde{D}) = \langle x^2 + [\texttt{0x0 0x958DE384C421 0xBFA3449BA464}]\ x + \\ + [\texttt{0xA7A1F454D931 0x2AF2C903858 0x1C3DF68C4EC}], \\ [\texttt{0x2B41D88C5393 0xE4C0516BD0C7 0x67AF7DD4F906}]\ x + \\ + [\texttt{0xBAD12BA8B824 0x64D6CA8CC5B1 0x42B5E64AB197}]\rangle$$

Group Law Algorithms For Jacobian Varieties Of Curves Over Finite Fields

Ran Cohen

Department of Mathematics,
Bar-Ilan University, Ramat Gan, Israel
E-mail : cohenran@yahoo.com

This paper reviews the group law algorithms of nonsingular C_{ab} curves, and provides implementation results comparing it to an algorithm for computing in the generalized Jacobian of C_A curves.

Keywords: generalized Jacobian, hyperelliptic curves, cryptography

1. Introduction

The success of Jacobian groups of (nonsingular) elliptic and hyperelliptic curves in cryptography has drawn a lot of interest over the recent years. One of the main areas of research is to find suitable Jacobian groups of other families of curves, and implement their group operation efficiently. The first part of this work is a survey of the group law algorithms of nonsingular C_{ab} curves. The second part is a review of the algorithm, proposed recently by Arita, Miura and Sekiguchi, for generalized Jacobian of a special family of singular curves, so called C_A curves. Implementation results provide a comparison of the running time between C_{ab} curves and C_A curves.

In this paper we follow the notations of [41]. Let $\Bbbk = \mathbb{F}_q$, where $q = p^t$ for some prime p, and denote by $\overline{\Bbbk}$ its algebraic closure. Let $C_0/\Bbbk$ be an absolutely irreducible projective curve. We fix a $\Bbbk$-rational point $P_\infty \in C_0$, and let C be the curve obtained by desingularizing only the point P_∞. We assume there is only one point lying above P_∞ in C which is also denoted P_∞. In addition, we assume that $C_a = C\backslash\{P_\infty\}$ is a nonsingular affine curve. Let $R = \Bbbk[C_a]$ be its coordinate ring, and $\Bbbk(C)$ its field of rational functions. Let g be the genus of the curve C.

The Jacobian variety $\mathrm{Jac}(C)$ is isomorphic as a group to the degree zero subgroup of the Picard group $\mathrm{Pic}^0(C) := \mathrm{Div}^0(C)/\mathrm{PDiv}(C)$. We focus on

$\mathrm{Pic}^0_{\Bbbk}(C)$, the invariant subgroup w.r.t. $\mathrm{Gal}(\overline{\Bbbk}/\Bbbk)$. Recall that a divisor of the form $E-nP_\infty$, where $E \geq 0$ is affine and of degree n, is called a *semi-reduced divisor* with *weight* n. Using Riemann-Roch Theorem one can prove that every divisor class $[D] \in \mathrm{Pic}^0_{\Bbbk}(C)$ has a unique representative $D \sim E-mP_\infty$, with minimal weight m. The divisor $E - mP_\infty$ is called a *reduced divisor*.

The coordinate ring R is obviously Noetherian and of Krull dimension 1. For curves the notions of normality and nonsingularity are equivalent, therefore C_a is a normal curve. It follows that R is a Dedekind domain, and the ideal class group $\mathrm{IdCl}(R)$ is well defined. In this case we have a natural isomorphism between $\mathrm{Pic}_{\Bbbk}(C_a)$ and $\mathrm{IdCl}(R)$. In addition, P_∞ is the single point at infinity, and so we have an isomorphism between $\mathrm{Pic}^0_{\Bbbk}(C)$ and $\mathrm{Pic}_{\Bbbk}(C_a)$, by $\sum n_P P - (\sum n_P)P_\infty \leftrightarrow \sum n_P P$. For conclusion, we get the following sequence of isomorphisms of groups:

$$\mathrm{Jac}_{\Bbbk}(C) \simeq \mathrm{Pic}^0_{\Bbbk}(C) \simeq \mathrm{Pic}_{\Bbbk}(C_a) \simeq \mathrm{IdCl}(R).$$

2. C_{ab} Curves

Let $C/\Bbbk$ be a plane irreducible projective curve, and let $P \in C$. We define

$$\mathcal{L}(\infty P) := \bigcup_{n \geq 0} \mathcal{L}(nP).$$

Define $M(P) = \big\{ -\mathrm{ord}_P(f) : f \in \mathcal{L}(\infty P) \backslash \{0\} \big\}$. If $a, b \in M(P)$ there exist functions $f, h \in \mathcal{L}(\infty P)$ such that $-\mathrm{ord}_P(f) = a$ and $-\mathrm{ord}_P(h) = b$. Clearly $fh \in \mathcal{L}(\infty P)$, and $-\mathrm{ord}_P(fh) = a + b$, so $M(P)$ is a unitary semigroup w.r.t. addition.

Definition 2.1 (C_{ab} curve, [2]). *If the semigroup $M(P)$ is generated by two relatively prime positive integers a and b, then the pair (C, P) is called a C_{ab}* curve.

Let (C, P) be a C_{ab} curve. Then by definition there are functions $x, y \in \mathcal{L}(\infty P)$ with poles of order a and b respectively. Using these two functions we obtain the affine model of the C_{ab} curve $F(x, y) = \sum_{ai+bj \leq ab} c_{ij} x^i y^j = 0$, where $0 \leq i \leq b$, $0 \leq j \leq a$, $c_{ij} \in \Bbbk$, $c_{b0} \neq 0$ and $c_{0a} \neq 0$. This affine model of C is called the *Miura canonical form*. We assume that C_{ab} curves satisfy the conditions in the introduction, i.e. nonsingular in the affine plane, and P is the only point at infinity, denoted P_∞. The genus of a C_{ab} curve is $g = \frac{1}{2}(a-1)(b-1)$.

Definition 2.2 (C_{ab} order, [2]). *Let $\alpha = (\alpha_1, \alpha_2), \beta = (\beta_1, \beta_2) \in \mathbb{N}^2$. We say that $\alpha >_{ab} \beta$, if one of the following holds:*

(1) $a\alpha_1 + b\alpha_2 > a\beta_1 + b\beta_2$.
(2) $a\alpha_1 + b\alpha_2 = a\beta_1 + b\beta_2$, *and* $\alpha_1 < \beta_1$.

We can order the monomials $x^{\alpha_1}y^{\alpha_2} \in \Bbbk(C)$ by their pole order at P_∞, by setting $-\mathrm{ord}_{P_\infty}(x^{\alpha_1}y^{\alpha_2}) = a\alpha_1 + b\alpha_2$, and when two monomials have the same pole order at infinity, the monomial with the larger degree of x is smaller.

2.1. *Classification of algorithms*

Various algorithms were proposed for Jacobians of C_{ab} curves, with special care for curves appealing for cryptography such as C_{25} and C_{34} curves. The algorithms consist of two stages: composition and reduction, and can be classified to the following types:

The first type is based on *hyperplane intersection*. The underlying idea is to construct a hyperplane interpolating the affine points in the support of two reduced divisors. The intersection of this hyperplane with the curve forms a divisor in the opposite class. Inverting this divisor yields the reduced divisor.

The second type is based on Cantor's algorithm. This approach is inspired from Gauß' algorithm for quadratic forms, which was utilized by E. Artin to the case of hyperelliptic fields, quadratic extensions of $\Bbbk(x)$. Cantor used this approach for computing in Jacobians of hyperelliptic curves.

The third type is based on Gröbner basis manipulation and includes two subfamilies: algorithms based on lexicographic order and algorithms based on C_{ab} order. Algorithms in the first subfamily use an LLL-like algorithm in order to reduce a divisor. This algorithm finds a reduced basis for a lattice in a function field, allowing us to compute the minimal element in the lattice based on a specific metric. Algorithms in the second subfamily compute the minimal element in the reduced Gröbner basis (with respect to C_{ab} order) of the ideal corresponding to the divisor. The divisor is then reduced using this minimal element.

In the remaining part of this section we review special families of C_{ab} curves, and the algorithms proposed for each family based on chronological order.

2.2. *Elliptic Curves*

The first example of C_{ab} curves are C_{23} curves, elliptic curves. By definition, elliptic curves are nonsingular curves of genus 1 along with a fixed

base point. In this case the Miura canonical form is mostly known as the Weierstraß form, $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$, where the fixed point P_∞ corresponds to $(0, 1, 0)$.

The set of rational points on an elliptic curve has a natural composition, the *chord and tangent* law. This algorithm obviously belongs to the first type, hyperplane intersection.

Based on the chord and tangent law, one can derive explicit formulae valid over any field. These formulae can be simplified based on the characteristic of the field. The efficiency of the computation can be further improved based on the coordinates system: affine coordinates, projective coordinates, Jacobian coordinates, Chudnovski-Jacobian coordinates, modified Jacobian coordinates, or mixed coordinates (see [12]).

2.3. *Hyperelliptic Curves*

A *hyperelliptic curve* $C/\Bbbk$ is a projective curve of genus $g \geq 1$ that admits a nonsingular affine model of the form $y^2+H(x)y = F(x)$, where $H, F \in \Bbbk[x]$, F is monic of degree $2g + 1$, and $\deg(H) \leq g$. Clearly, hyperelliptic curves are C_{2b} curves. If $\mathrm{chr}(\Bbbk) = 2$ then we have $H(x) \neq 0$, and if $\mathrm{chr}(\Bbbk) \neq 2$, C can be represented in the form $y^2 = F(x)$. The order of the Jacobian of a hyperelliptic curve is approximately q^g, so we can obtain a similar group as of an elliptic curve while working over a field with $\sqrt[g]{q}$ elements.

In ([32]) Mumford showed that every semi-reduced divisor on a hyperelliptic curve can be represented as the gcd of two principal divisors. Using Mumford's representation one can represent a divisor over $\Bbbk$ even if its support is contained in some extension of $\Bbbk$. Let $D = \sum n_P P - \big(\sum n_P\big)P_\infty$ be a semi-reduced divisor, and set $U(x) = \prod(x - x_P)^{n_P} \in \Bbbk[x]$, then there is a unique polynomial $V(x) \in \Bbbk[x]$ satisfying: $\deg(V) < \deg(U)$, $V(x_P) = y_P$ for all P for which $n_P \neq 0$, and $U|(V^2+VH-F)$. It follows that $D = \gcd\big(\mathrm{div}(U), \mathrm{div}(y - V)\big)$. We simplify this notation to $\mathrm{div}(U, y - V)$.

In [18] Gaudry used a variant of index calculus for the Jacobian group of hyperelliptic curves of genus g defined over $\mathbb{F}_q$, and managed to compute the DLP with complexity $O(q^2)$. Note that the group size is approximately q^g, and so Pollard's Rho attack computes the DLP with complexity $O(q^{g/2})$. Thus, Gaudry's attack is faster than Pollard's Rho when the genus is greater than 4.

2.3.1. *Cantor's algorithm*

In 1987 Cantor utilized Mumford's representation to propose an algorithm for addition in the Jacobian of hyperelliptic curves over fields of odd characteristic (see [10]). In 1989 Koblitz generalized Cantor's algorithm to arbitrary characteristic (see [24]). This algorithm consists of two steps: composition and reduction. The reduction algorithm is based on the classical method due to Gauß. Nevertheless, there are other reduction algorithms that are faster asymptotically, i.e. as the genus grows larger.

In 2000 Nagao improved Cantor's algorithm over odd characteristic, based on the following ideas: (see [33])

- Dividing polynomials without inversion in the base field.
- Computing gcd of polynomials using only one inversion in the base field.
- Ignoring superfluous operations during the algorithm.
- Representing the Jacobian's elements differently.

2.3.2. *Harley's algorithm*

In 2000 Harley proposed a generalization of the chord and tangent law to the case of hyperelliptic curves of genus 2 over odd characteristic ([22,23]). This algorithm is based on hyperplane intersection. Given two reduced divisors D_1 and D_2, after the composition step we have a semi-reduced divisor R of weight at most $2g$. If $\text{weight}(R) \leq g$ then R is already reduced, otherwise, for genus 2, $\text{weight}(R)$ can be either 3 or 4.

If $\text{weight}(R) = 3$, then $R = P_1 + P_2 + P_3 - 3P_\infty$. Denote $y = A(x)$ the hyperplane (parabola or straight line) passing through the three points. The roots of $F - A^2$ are the x-coordinates of the intersection points between the hyperplane and C. $F - A^2$ is a polynomial of degree 5, hence there are 5 intersection points, denote Q_1, Q_2 the remaining two, and define $S = Q_1 + Q_2 - 2P_\infty$, then the result is $-S$.

If $\text{weight}(R) = 4$, then $R = P_1 + P_2 + P_3 + P_4 - 4P_\infty$. Denote $y = A(x)$ the hyperplane interpolating the four points, this is a polynomial of degree at most 3. $F - A^2$ is a polynomial of degree 5 or 6, and we know 4 of the roots. Construct S as in the previous case, and the result is $-S$.

Harley implemented the algorithm using relatively fast techniques such as CRT, Newton's iteration, and Karatsuba method for polynomials multiplication. In this way the polynomial arithmetic was reduced to arithmetic over the base field.

2.3.3. *Improvements for genus two*

Since Harley's algorithm many improvements were suggested for genus 2. The main ideas of generalizations and improvements were in the following issues:

Generalization to arbitrary characteristic [26,27,42,43].

Reduce the number of operations [30,31,44].

Use of different sets of coordinates ([28,29,31]).

Focus on special families of curves [9,36,37].

The comparison between various algorithms for hyperelliptic curves of genus 2 can be found in Appendix A.1.

2.3.4. *Improvements for genus three*

Genus 3 hyperelliptic curves were considered to be attractive because they are resilient against Gaudry's attack on one hand, and allow working over smaller fields than genus 2 curves do. However, a recent variant of Gaudry's attack ([19]) appears to be faster than Pollard Rho attack in this case, so one should be careful about using these curves in standard cryptosystems. Genus 3 hyperelliptic curves are C_{27} curves, i.e. of the form

$$y^2+(h_3x^3+h_2x^2+h_1x+h_0)y = x^7+f_6x^6+f_5x^5+f_4x^4+f_3x^3+f_2x^2+f_1x+f_0.$$

In [25] Kuroki et al. generalized Harley's algorithm to genus 3 over odd characteristic. In [35] Pelzl et al. generalized it to arbitrary characteristic. In [20] the authors managed to save some multiplications by using Toom's polynomial multiplication instead of Karatsuba's multiplication. In [13] Fan and Wang implemented the algorithm over odd characteristic using projective coordinates. In [14] Fan et al. focused on the doubling operation over binary fields. In [5] Avanzi et al. focused on the arithmetic over binary fields.

The comparison between various algorithms for hyperelliptic curves of genus 3 can be found in Appendix A.2.

2.3.5. *Improved Algorithms for Genus four*

Genus 4 hyperelliptic curves allow working over smaller fields than genus 2 and 3 curves. However, these curves are less attractive for standard cryptosystems due to recent attacks ([19]). Genus 4 hyperelliptic curves are C_{29} curves, i.e. of the form

$$y^2 + (h_4x^4 + h_3x^3 + h_2x^2 + h_1x + h_0)y = x^9 + f_8x^8 + f_7x^7 + \ldots + f_1x + f_0.$$

In [38] Pelzl et al. introduced the first explicit formulae for genus 4 hyperelliptic curves. In [5] Avanzi et al. used the same methods both for genus 3 and for genus 4. The authors focused on curves over binary fields, and implemented the arithmetic "tricks" that were used for smaller genus as well as more efficient field arithmetic.

The comparison between various algorithms for hyperelliptic curves of genus 4 can be found in Appendix A.3.

2.4. *Superelliptic Curves*

A *superelliptic curve* C is a curve that admits an affine model of the form

$$y^a = F(x) = f_b x^b + f_{b-1} x^{b-1} + \ldots + f_1 x + f_0.$$

We can see that superelliptic curves are C_{ab} curves where $c_{ij} = 0$ for $0 \leq i \leq b$ and $0 < j \leq a-1$, $c_{ia} = 0$ for $0 < i \leq b$, and $c_{0a} = 1$.

In order for the superelliptic curve to be nonsingular in the affine plane we assume that $\gcd\big(F(x), F'(x)\big) = 1$, and that $\mathrm{chr}(\Bbbk) \nmid a$. To ensure one and only one point at infinity we assume that $\gcd(a, b) = 1$. The field extension $\Bbbk(C)/\Bbbk(x)$ is a Galois extension, and the Galois group is $\mathrm{Gal}\big(\Bbbk(C)/\Bbbk(x)\big) = \langle\sigma\rangle$, where σ is of the form $\sigma(x, y) \mapsto (x, \zeta y)$ and ζ is a primitive a-th root of unity.

2.4.1. *GPS Algorithm*

In [17] Galbraith et al. proposed an algorithm for computing in the Jacobian of superelliptic curves by adopting an LLL-like algorithm for lattice reduction to provide the reduction method. Their approach is analogous to the strategy of computing with ideals in number fields ([11] Section 6.5). This algorithm belongs to the third type, Gröbner basis manipulation.

In [34] Paulus modified the LLL algorithm to compute a reduced basis of a lattice in a function field. First we embed $\Bbbk[C]$ into $\Bbbk[x]^a$ in the following way:

$$\varphi : \sum_{i=1}^{a} h_i(x) y^i \mapsto \big(h_1(x), \ldots, h_a(x)\big).$$

Denote $A = \big(h_1(x), \ldots, h_a(x)\big) \in \Bbbk[x]^a$, we can define a metric on A by setting $|A|_i := \deg_x(h_i(x)) + \frac{b}{a} i$, and then $|A| := \max_i\{|A|_i\}$.

Consider an ideal $\mathfrak{a} \subseteq \Bbbk[C]$, and let $[\alpha_1, \ldots, \alpha_a]$ be a $\Bbbk[x]$-basis of $\mathfrak{a}$, then the image of $\mathfrak{a}$ under φ is a lattice over $\Bbbk[x]$ generated by $\{\varphi(\alpha_i)\}$.

The authors represented divisors classes using the unique HNF (Hermite normal form) of the reduced ideal. Given reduced ideals $\mathfrak{a}_1, \mathfrak{a}_2$, the algorithm consists of four steps

(1) $\mathfrak{b} \leftarrow \mathfrak{a}_1\mathfrak{a}_2$.
(2) $\mathfrak{c} \leftarrow$ the semi-reduced ideal equivalent to $\mathfrak{b}^{-1}$.
(3) $b \leftarrow$ a minimal nonzero element in $\mathfrak{c}$.
(4) $\mathfrak{a}_3 \leftarrow$ the HNF of $b\mathfrak{c}^{-1}$.

The complexity of the GPS algorithm is $O\big(a^6b^2g^2\big)$ operations in $\Bbbk$.

2.4.2. *Cantor's Algorithm for Superelliptic Cubics*

In [6] Basiri et al. focused on superelliptic cubics, i.e. superelliptic curves with $a = 3$. In this case we have that $\mathrm{Gal}\big(\Bbbk(C)/\Bbbk(x)\big) = \{\mathrm{Id}, \sigma, \sigma^2\}$, where σ is of the form $\sigma(x, y) \mapsto (x, \zeta y)$ and ζ is a primitive third root of unity.

Let D be a $\Bbbk$-rational divisor. The *conjugates* of D are D^σ and D^{σ^2}. The authors noticed that Mumford's representation is suitable for a special class of divisors, namely divisors that do not have any two conjugate points in their support. A divisor D is called *typical* if it is of the form $D = \mathrm{div}(U, y - V)$ for some $U, V \in \Bbbk[x]$ such that $\deg(V) < \deg(U) \leq g$ and $U \mid (V^3 - F)$. For a fixed g, the probability that a reduced divisor is not typical is $O(\frac{1}{q})$.

Given a superelliptic cubic of genus 3 or 4, a typical divisor $\mathrm{div}(U, y-V)$ is reduced whenever $\deg(U) < g$, or $\deg(U) = g$ and $\deg(V) = g - 1$.

Based on the similarity between Mumford's representation of reduced divisors on a hyperelliptic curve and the representation of typical divisors on a superelliptic cubic, Basiri et al. generalized Cantor's algorithm to the case of superelliptic cubics. The classic approach of Gauß' reduction fails for superelliptic cubics, so Lagrange reduction is used in this case.

2.4.3. *Bauer's Algorithm for Superelliptic Cubics*

In [8] Bauer implemented GPS algorithm for superelliptic curves in the specific case of superelliptic cubics. The author noticed that because GPS algorithm is very general it contains certain inefficiencies, and by exploiting the underlying structure of superelliptic cubics the arithmetic can be improved.

2.4.4. *Flon-Oyono Algorithm for Picard Curves*

A *Picard curve* is a superelliptic curve of genus 3: $y^3 = F(x) = x^4 + f_3x^3 + f_2x^2 + f_1x + f_0$.

In [15] Flon and Oyono took the approach of hyperplane intersection, and obtained explicit formulae for computing in the Jacobian of Picard curves. Following [6], the authors concentrated on addition in the most frequent cases, i.e. of typical divisors.

Given two reduced divisors $D_1 = P_1 + P_2 + P_3 - 3P_\infty$ and $D_2 = Q_1 + Q_2 + Q_3 - 3P_\infty$, we want to find the reduced divisor equivalent to $P_1 + P_2 + P_3 + Q_1 + Q_2 + Q_3 - 6P_\infty$. For this we look at the divisor

$$D = -\big(P_1 + P_2 + P_3 + Q_1 + Q_2 + Q_3 - 9P_\infty\big).$$

D is a $\Bbbk$-rational divisor of degree 3, so by Riemann-Roch there exists a function $W \in \Bbbk(C)^*$ such that $\mathrm{div}(W) \geq -D$. The only pole of W is P_∞, hence $W \in \Bbbk[C]$. In addition, $\mathrm{ord}_{P_\infty}(W) \geq -9$, so W is an element of the $\Bbbk$-vector space spanned by $1, x, x^2, xy, y^2, x^3$. Take W to be the unique such element with max. order at P_∞.

If W is a conic, then $\mathrm{Supp}(D_1 + D_2)$ consists of 6 points aside from P_∞ that lie on W. This conic intersects C in exactly two more points R_1 and R_2. Taking a hyperplane through these points gives us two new points K_1, K_2. Thus, the reduction of $D_1 + D_2$ is $K_1 + K_2 - 2P_\infty$.

If W is a cubic, then by Bézout theorem W intersects C in exactly three more points R_1, R_2, R_3. We get that

$$(P_1+P_2+P_3-3P_\infty)+(Q_1+Q_2+Q_3-3P_\infty) = -(R_1+R_2+R_3-3P_\infty)+\mathrm{div}(W).$$

Using Riemann-Roch again we get that there exists a unique conic W_2 passing through R_1, R_2, R_3 and through three more points K_1, K_2, K_3. Thus, $(K_1 + K_2 + K_3 - 3P_\infty)$ is the reduced representative of $D_1 + D_2$.

2.5. C_{ab} Curves

In this section we review some generic algorithms for C_{ab} curves, and focus on the more cryptographically interesting C_{34} curves.

2.5.1. *Arita's Algorithm for C_{ab} Curves*

In [2] Arita proposed an addition algorithm in the Jacobian of C_{ab} curves. Reduced ideals are represented by their reduced Gröbner basis w.r.t. C_{ab} order.

Let C be a C_{ab} curve, and let $\mathfrak{a} \subset \Bbbk[C]$ be an ideal. We denote by $f_{\mathfrak{a}}$ the nonzero 'monic' polynomial with smallest leading monomial (w.r.t. C_{ab} order) in $\mathfrak{a}$, and $\mathfrak{a}^* := \big(\langle f_{\mathfrak{a}}\rangle :_{\Bbbk[C]} \mathfrak{a}\big) = \big\{g \in \Bbbk[C] : g\mathfrak{a} \subseteq \langle f_{\mathfrak{a}}\rangle\big\}$. Notice that $\mathfrak{a}\mathfrak{a}^* = \langle f_{\mathfrak{a}}\rangle$, thus $\mathfrak{a}^* = \mathfrak{a}^{-1}$. An ideal $\mathfrak{a}$ is reduced iff $\mathfrak{a} = \mathfrak{a}^{**}$.

Given two reduced ideals $\mathfrak{a}_1, \mathfrak{a}_2 \subseteq \Bbbk[C]$, in the composition step we compute the ideal $\mathfrak{b} = \mathfrak{a}_1\mathfrak{a}_2$. In the reduction step we first compute the reduced inverse $\mathfrak{b}^*$ and later the reduced ideal $\mathfrak{a}_3 = \mathfrak{b}^{**}$. We need to find the polynomial $g = f_{\mathfrak{b}}$ such that $\langle g\rangle = \mathfrak{b}\mathfrak{b}^*$ and the polynomial $h = f_{\mathfrak{b}^*}$ such that $\langle h\rangle = \mathfrak{b}^*\mathfrak{b}^{**}$. Combining these relations we get that $\mathfrak{b}\langle h\rangle = \mathfrak{b}\mathfrak{b}^*\mathfrak{b}^{**} = \langle g\rangle\mathfrak{b}^{**}$, hence $\mathfrak{b}^{**} = \frac{h}{g}\mathfrak{b}$.

(1) $\mathfrak{b} \leftarrow \mathfrak{a}_1\mathfrak{a}_2$.
(2) $g \leftarrow$ the minimal nonzero element in $\mathfrak{b}$ w.r.t. C_{ab} order.
(3) $h \leftarrow$ the minimal nonzero element w.r.t. C_{ab} order, satisfying $h\mathfrak{b} \subseteq \langle g\rangle$.
(4) $\mathfrak{a}_3 \leftarrow \frac{h}{g}\mathfrak{b}$.

The minimal elements in steps 2 and 3 are found by computing the reduced Gröbner bases of the ideals $\mathfrak{b}$ and $\mathfrak{b}^*$. Arita used Buchberger's method for this computation. The complexity of this algorithm is $O(g^3 \log^2 q)$.

2.5.2. *GPS Algorithm for C_{ab} Curves*

In [21] Harasawa and Suzuki generalized the GPS algorithm of superelliptic curves to C_{ab} curves. In order to generalize GPS algorithm, the authors had to address two issues:

(1) Given an ideal $\mathfrak{a}$, how to compute the inverse $\mathfrak{a}^{-1}$.
(2) How to compute the minimal element over an ideal w.r.t. C_{ab} order.

The first issue is treated with methods of computing inverse ideals in the integral closure of a number field (see [11, Prop. 4.8.19]). In the case of C_{ab} curves, $\Bbbk[C]$ is the integral closure of $\Bbbk[x]$ in $\Bbbk(C)$, and the set $\{1, y, \ldots, y^{a-1}\}$ is a $\Bbbk[x]$-basis of $\Bbbk[C]$.

Regarding the second issue, given $h = \sum_{i=1}^{a} h_i(x)y^i \in \Bbbk[C]$, by the definition of the metric $|\cdot|$, we have

$$-\mathrm{ord}_{P_\infty}(h) = \max_{1\le i\le a}\Big\{a\deg_x(h_i)+bi\Big\} = a\max_{1\le i\le a}\Big\{\deg_x(h_i)+\frac{b}{a}i\Big\} = a|\varphi(h)|.$$

Therefore, for an ideal $\mathfrak{a} \subseteq \Bbbk[C]$, finding the minimal element over $\mathfrak{a}$ w.r.t. the C_{ab} order is the same as finding the minimal element over $\varphi(\mathfrak{a})$ w.r.t. the metric $|\cdot|$. Thus we can apply Paulus' method for lattice reduction in order to find the minimal element w.r.t. the C_{ab} order.

The overall complexity Harasawa and Suzuki obtained for the addition in the Jacobian of a C_{ab} curve is $O(a^8 g^2 \log^2 q)$.

2.5.3. *Arita's Algorithm for C_{34} Curves*

In [3] Arita managed to simplify his algorithm in the case of C_{34} curves. This was accomplished by classifying the Gröbner bases of the ideals, and so computing their Gröbner bases without the use of Buchberger algorithm. The author carried out the computation symbolically and managed to obtained explicit formulae.

The genus of a C_{34} curve is 3, therefore, after the composition step the order of the ideals is at most 6. The minimal six polynomials w.r.t. C_{34}-order are $M = \{1, x, y, x^2, xy, y^2\}$. Arita classified the ideals of degree at most 6 based on the different possibilities of linear independence of polynomials in M.

2.5.4. *BEFG Algorithm for C_{34} Curves*

In [6] and [7] the authors managed to obtain explicit formulae for adding and doubling typical reduced ideals in C_{34} curves. The underlying method for the reduction step was the FGLM algorithm for switching between Gröbner bases of different orderings, and so compute the Gröbner basis in a C_{34} order from the Gröbner basis in lexicographic order.

Let C be a C_{34} curve of the form $y^3 + H(x)y = F(x)$ where $\deg(F) = 4$ and $\deg(H) \leq 2$. In the composition step, given two typical ideals $\mathfrak{a}_i = \langle U_i, y - V_i \rangle$ where $\deg(U_i) = 3$ and $\deg(V_i) = 2$, we get the product $\mathfrak{b} = \mathfrak{a}_1\mathfrak{a}_2 = \langle U, y - V \rangle$ where $U = U_1U_2$ of degree 6 and $\deg(V) = 5$. In the case of addition, where $U_1 \neq U_2$ and $\gcd(U_1, U_2) = 1$, we can find V using the CRT as follows:

$$S_1 \equiv U_1^{-1} \pmod{U_2}, \quad T \equiv S_1(V_2 - V_1) \pmod{U_2}, \quad V = V_1 + TU_1.$$

In the reduction step we have the ideal $\mathfrak{b} = \langle U, y - V \rangle$. Let E be the minimal element w.r.t. C_{34} order in the ideal $\mathfrak{b}^{-1} = \langle U, y^2 + Vy + V^2 + H \rangle$. The reduced ideal is $\mathfrak{a}_3 = \frac{E}{U}\mathfrak{b} = \langle U_3, y - V_3 \rangle$.

2.5.5. *FOR Algorithm for C_{34} Curves*

In [16] Flon et al. focused on Jacobians of non-hyperelliptic curves of genus 3. Such a curve can be represented as a smooth projective plane quartic C. We assume there exists a rational line ℓ^∞ which crosses C in four $\mathbb{k}$-rational points P_1^∞, P_2^∞, P_3^∞ and P_4^∞. There are 5 possibilities:

(1) The four points are distinct.
(2) $P_1^\infty = P_2^\infty$, then ℓ^∞ is tangent to C at P_i^∞.
(3) $P_1^\infty = P_2^\infty = P_3^\infty$, then the point P_1^∞ is called a *flex*.
(4) $P_1^\infty = P_2^\infty$ and $P_3^\infty = P_4^\infty$, then ℓ^∞ is call *bitangent*.
(5) $P_1^\infty = P_2^\infty = P_3^\infty = P_4^\infty$, then the point P_1^∞ is called a *hyperflex*.

Denote $D^\infty = P_1^\infty + P_2^\infty + P_3^\infty$. For every $D \in \mathrm{Div}^0_{\Bbbk}(C)$ let D^+ be an effective divisor such that $D^+ - D^\infty \sim D$. The algorithm is based on the following theorem.

Theorem 2.1 ([16] p. 4). *Let $D_1, D_2 \in \mathrm{Div}^0_{\Bbbk}(C)$. Then $D_1 + D_2$ is equivalent to a divisor $D = D^+ - D^\infty$, where the points in* $\mathrm{Supp}(D^+)$ *are given by the following:*

(1) Take the unique cubic E which goes (with multiplicity) through the support of D_1^+, D_2^+ and P_1^∞, P_2^∞ and P_4^∞. This cubic also crosses C in the residual effective divisor D_3.
(2) Take the unique conic Q which goes through the support of D_3 and P_1^∞, P_2^∞. This conic also crosses C in the residual effective divisor D^+.

The authors implemented this algorithm for the flex case, with cost of 2I + 163M for addition and 2I + 185M for doubling. In the subcase of hyperflex we get a C_{34} curve, and the cost is 2I + 145M for addition and 2I + 167M for doubling.

The comparison of the algorithms for C_{34} curves can be found in Appendix A.4. *

3. C_A Curves

In the previous section we described the group law operation in the Jacobian group of certain nonsingular curves. In this section we discuss a generalization of this situation, and explain the group law in the generalized Jacobian of (possibly singular) C_A curves. C_{ab} curves are a special case of C_A curves. The concept of generalized Jacobian was first introduced by Rosenlicht in [39] More details can be found in [40].

3.1. *Generalized Jacobian Varieties*

Let $C/\Bbbk$ be a complete irreducible nonsingular projective curve, let $\mathfrak{m} = \sum m_P P$ be an effective $\Bbbk$-rational divisor, and let $S = \mathrm{Supp}(\mathfrak{m})$. We call

*After the release of this paper another algorithm was published in [1], using an approach which reduces the computation to linear algebra.

$\mathfrak{m}$ a *modulus*. Given a function $f \in \Bbbk(C)^*$ we denote $f \equiv 1 \pmod{\mathfrak{m}}$ if for every $P \in C$ we have $\mathrm{ord}_P(1-f) \geq m_P$.

Definition 3.1 ($\mathfrak{m}$-equivalence). *Let D_1 and D_2 be two divisors over C prime to S. We say that D_1 and D_2 are* $\mathfrak{m}$-equivalent*, and write $D_1 \sim_{\mathfrak{m}} D_2$, if there is a function $f \in \Bbbk(C)^*$ such that $div(f) = D_1 - D_2$ and $f \equiv 1$ (mod $\mathfrak{m}$).*

Given an irreducible nonsingular curve C, a finite subset of $S \subseteq C$ and an equivalence relation $\sim$ on S, one can construct a singular curve $C' = (C \backslash S) \cup (S/\sim)$ with C its normalization. Given a modulus $\mathfrak{m}$ with $\deg(\mathfrak{m}) \geq 2$, $\sim_{\mathfrak{m}}$ is an equivalence relation and we denote the singular curve C' by $C_{\mathfrak{m}}$.

Just like linear equivalence gives rise to the Jacobian variety, given a modulus $\mathfrak{m}$ the equivalence relation $\sim_{\mathfrak{m}}$ gives rise to a generalized Jacobian variety, $J_{\mathfrak{m}}$.

The dimension of $J_{\mathfrak{m}}$ is the arithmetic genus of the curve $C_{\mathfrak{m}}$, that is

$$\pi = \begin{cases} g, & \mathfrak{m} = 0 \\ g + \deg(\mathfrak{m}) - 1, & \mathfrak{m} \neq 0 \end{cases}$$

Let $\mathfrak{m}$ be a modulus on C and $S = \mathrm{Supp}(\mathfrak{m})$, we define $\mathrm{Div}_{\mathfrak{m}}(C)$ to be the subgroup of $\mathrm{Div}(C)$ formed by divisors prime to S and $\mathrm{Div}^0_{\mathfrak{m}}(C)$ its subgroup of degree zero. We denote by $\mathrm{Pic}_{\mathfrak{m}}(C)$ (and $\mathrm{Pic}^0_{\mathfrak{m}}(C)$) the quotient group of $\mathrm{Div}_{\mathfrak{m}}(C)$ (resp. $\mathrm{Div}^0_{\mathfrak{m}}(C)$) of $\mathfrak{m}$-equivalence classes. $J_{\mathfrak{m}}$ is isomorphic (as a group) to $\mathrm{Pic}^0_{\mathfrak{m}}(C)$.

In order to understand the structure of $J_{\mathfrak{m}}$ first note that there are isomorphisms of groups $\varphi : \mathrm{Pic}^0(C) \to J$ and $\psi : \mathrm{Pic}^0_{\mathfrak{m}}(C) \to J_{\mathfrak{m}}$. In addition, every pair of $\mathfrak{m}$-equivalent divisors is obviously also linearly equivalent, so we have an epimorphism $\sigma : \mathrm{Pic}^0_{\mathfrak{m}}(C) \to \mathrm{Pic}^0(C)$. Combining the three maps together, we get an epimorphism $\tau : J_{\mathfrak{m}} \to J$ defined by $\tau = \varphi \circ \sigma \circ \psi^{-1}$. Denote by $L_{\mathfrak{m}}$ the kernel of τ, and we get the following short exact sequence of groups

$$0 \longrightarrow L_{\mathfrak{m}} \hookrightarrow J_{\mathfrak{m}} \xrightarrow{\tau} J \longrightarrow 0.$$

Thus, the generalized Jacobian $J_{\mathfrak{m}}$ is an extension of the Jacobian J by $L_{\mathfrak{m}}$. The algebraic group $L_{\mathfrak{m}}$ is biregular isomorphic to $R_{\mathfrak{m}}/\mathbb{G}_m$, i.e.

$$L_{\mathfrak{m}} \simeq \prod_{i=1}^{\#S-1} \mathbb{G}_m \times \prod_{P \in S} V_{(m_P)},$$

where $V_{(m_P)}$ is a unipotent group isomorphic to a group of matrices.

3.2. *The Ideal Class Group of Singular Curves*

Let $C_0/\Bbbk$ be a plane irreducible projective curve. We fix a $\Bbbk$-rational point $P_\infty \in C_0$. Let C_1 be the curve obtained by desingularizing only the point P_∞. We assume there is only one point lying above P_∞ in C_1 which is also denoted P_∞. Let S be the set of singular points of C_1, and let C be the normalization of C_1 (and of C_0) with the map $d : C \to C_1 \to C_0$. Denote by $\mathfrak{m}$ the modulus defined by $d^{-1}(S)$. Let g be the genus of the curve C.

We assume $C_1^a = C_1 \backslash \{P_\infty\}$ is an affine curve, and let $R = \Bbbk[C_1^a]$ be its coordinate ring. Note that if C_1^a has singular points R is not a Dedekind domain. However, we can still denote by $\mathrm{Id}(R)$ the group of invertible fractional ideals of R, and its subgroup $\mathrm{PId}(R) = \{\langle f\rangle = fR : f \in \Bbbk(C)^*\}$. Thus, we can define with this notation the ideal class group $\mathrm{IdCl}(R) = \mathrm{Id}(R)/\mathrm{PId}(R)$.

The local ring of a nonsingular point P is regular, i.e. the corresponding maximal ideal $\mathfrak{m}_P$ is principal. However, the maximal ideal corresponding to a singular point is generated by more than one element. Therefore, the group of invertible fractional ideals of R is the free abelian group generated by maximal ideals corresponding to nonsingular points. We generalize the situation of the previous section by assigning divisors prime to S, i.e. divisors generated by nonsingular points, to invertible fractional ideals generated by the corresponding regular maximal ideals, and vice-versa. For conclusion, we get the following sequence:

$$J_{\mathfrak{m}}(C) \simeq J_{\mathfrak{m}}(C_1) \simeq \mathrm{Pic}^0_{\mathfrak{m}}(C_1) \simeq \mathrm{Pic}_{\mathfrak{m}}(C_1^a) \simeq \mathrm{IdCl}(R).$$

3.3. *C_A Curves*

We denote by $\mathbb{N}$ the set of non-negative integers. Let $M \subset \mathbb{N}$ be a finitely generated semigroup, i.e.

$$M = \langle a_1, \ldots, a_t\rangle = \mathbb{N}a_1 + \ldots + \mathbb{N}a_t$$

where $t \leq a_1$, and $a_i \neq 0$. The complement of M in $\mathbb{N}$ is finite iff $\gcd(a_1, \ldots, a_t) = 1$. If we denote $b_i = \min\{a \in M : a \equiv i \pmod{a_1}\}$, then

$$\#(\mathbb{N}\backslash M) = \sum_{i=1}^{a_1-1} \left[\frac{b_i}{a_1}\right].$$

In this case M is called a *numerical semigroup*.

Let M be a numerical semigroup with a system of generators $A = \{a_1, \ldots, a_t\}$, where $t \leq a_1$. We now define a surjective map $\Psi : \mathbb{N}^t \longrightarrow M$ by

$$\Psi(n_1, n_2, \ldots, n_t) = \sum_{i=1}^{t} n_i a_i.$$

Using this map, we can define a monomial order on $\mathbb{N}^t$ as follows.

Definition 3.2 (C_A order). *Given $\alpha = (\alpha_1, \ldots, \alpha_t), \beta = (\beta_1, \ldots, \beta_t) \in \mathbb{N}^t$ we say that $\alpha <_A \beta$, if one of the following holds:*

(1) $\Psi(\alpha) < \Psi(\beta)$;
(2) $\Psi(\alpha) = \Psi(\beta)$, and $\alpha_1 = \beta_1, \ldots, \alpha_i = \beta_i, \alpha_{i+1} > \beta_{i+1}$.

Definition 3.3. For $a \in M$ we define $\mathbf{m}(a) = \min\{n \in \mathbb{N}^t : n \in \Psi^{-1}(a)\}$, and as before $b_i = \min\{a \in M : a \equiv i \pmod{a_1}\}$. In addition we set
$B(A) = \{\mathbf{m}(a) : a \in M\} \subset \mathbb{N}^t$.
$T(A) = \{\mathbf{m}(b_i) \in B(A) : i = 0, \ldots, a_1 - 1\}$.
$V(A) = \{\ell \in \mathbb{N}^t \backslash B(A) :$ if $\ell = m + n$ with $m \in \mathbb{N}^t \backslash B(A)$ and $n \in \mathbb{N}^t$, then $n = 0\}$.

$V(A)$ is a finite set and $B(A) = T(A) + \mathbb{N} \times \{0\}^{t-1}$.

Under these notations we consider polynomials $F_m \in \Bbbk[x_1, \ldots, x_t]$, for $m \in V(A)$, satisfying the following conditions:

(D1) For each $m \in V(A)$ set $\ell = \mathbf{m}\big(\Psi(m)\big)$, then

$$F_m = X^m + a_\ell X^\ell + \sum_n a_n X^n,$$

where $a_\ell \neq 0$, and the sum runs over $n \in B(A)$ with $n < \ell$.

(D2) $\big\{\sum_{n \in B(A)} \Bbbk X^n\big\} \bigcap \big\{F_m : m \in V(A)\big\} = \big\{0\big\}$.

The notation X^m means $x_1^{m_1} \cdots x_t^{m_t}$, where $m = (m_1, \ldots, m_t)$.

Let $C/\Bbbk$ be an irreducible nonsingular curve. Let $P \in C$ be a $\Bbbk$-rational point, like in C_{ab} curves we define

$$M(P) = \big\{ - \operatorname{ord}_P(f) : f \in \mathcal{L}(\infty P) \backslash \{0\}\big\} \subset \mathbb{N},$$

then $M(P)$ is a numerical semigroup.

Let R be a subalgebra of $\mathcal{L}(\infty P)$ such that $\Bbbk \subset R \subset \mathcal{L}(\infty P)$, and define a semigroup by

$$M(R) = \big\{ - \operatorname{ord}_P(f) : f \in R \backslash \{0\}\big\} \subset \mathbb{N}.$$

Lemma 3.1 ([4], Lemma 3.1). *The field of fractions of R coincides with $\Bbbk(C)$ iff $M(R)$ is a numerical semigroup.*

Hereafter we assume that $M(R)$ is a numerical semigroup. For every $i = 1, \ldots, t$ we choose a function $f_i \in R$ such that $\mathrm{ord}_P(f_i) = -a_i$, and consider the surjection $\Theta : \Bbbk[x_1, \ldots, x_t] \longrightarrow R$ defined by $\Theta(F) = F(f_1, \ldots, f_t)$ for $F \in \Bbbk[x_1, \ldots, x_t]$. The kernel of this map, denoted $I(R) = \ker \Theta$, is generated by polynomials satisfying conditions (D1) and (D2). Miura showed the converse is also true.

Theorem 3.1 ([4], Theorem 3.2). *Let M be a numerical semigroup with a system of generators $A = \{a_1, \ldots, a_t\}$. Then an ideal $I \subset \Bbbk[x_1, \ldots, x_t]$ is generated by polynomials satisfying conditions (D1) and (D2) iff there exist a function field $\mathbb{K}$ of one variable over $\Bbbk$, a $\Bbbk$-rational point P of the nonsingular model of $\mathbb{K}$, and a subalgebra $\Bbbk \subset R \subset \mathcal{L}(\infty P)$ such that the field of fractions of R is $\mathbb{K}$, and $I = I(R)$.*

In this case the polynomials satisfying (D1) and (D2) form a Gröbner basis of I, and the projective model $C_0(A)$ of $\mathrm{Spec}(\Bbbk[x_1, \ldots, x_t]/I) \subset \mathbb{A}^t_\Bbbk$ in $\mathbb{P}^t_\Bbbk$ has only one infinite point P, which is at most a cuspidal singularity. Moreover, the affine model $\mathrm{Spec}(\Bbbk[x_1, \ldots, x_t]/I)$ of $\mathbb{K}$ is nonsingular iff $R = \mathcal{L}(\infty P)$.

Definition 3.4 (C_A curve, [4]). *Under the notation of the theorem, when the ideal $I = I(R) \subset \Bbbk[x_1, \ldots, x_t]$ corresponds to the numerical semigroup M with a system of generators $A = \{a_1, \ldots, a_t\}$, we call $\mathrm{Spec}(\Bbbk[x_1, \ldots, x_t]/I)$, or $C_0(A)$, a C_A* curve. *For a numerical semigroup M generated by A, if there exists an affine nonsingular C_A curve, M is called a Weierstraß numerical semigroup.*

Using the Riemann-Roch theorem for singular curves we obtain the following formula for the genus of a C_A curve.

Proposition 3.1 ([4], Prop. 3.3). *Let C_1 be the curve given by desingularizing only the point at infinity of a C_A curve $C_0(A)$. The arithmetic genus of C_1 is*

$$p_a(C_1) = \#(\mathbb{N} \backslash M) = \sum_{i=1}^{a_1 - 1} \left[\frac{b_i}{a_1} \right].$$

If a C_A curve is generated by two coprime elements, $A = \{a, b\}$, we obtain a C_{ab} curve. A beautiful example of a C_{357} curve can be found in [4].

3.4. *Arita's Algorithm for C_A Curves*

In [4] the authors generalized Arita's algorithm for C_{ab} curves to the case of C_A curves. Let C be a C_A curve, and following the notations of Theorem 3.1 let $R = \Bbbk[C] = \Bbbk[x_1, \ldots, x_t]/I$ be its coordinate ring, and consider the canonical map $\Theta : \Bbbk[x_1, \ldots, x_t] \to R$ whose kernel is I. Note that C_A order is a monomial order on $\Bbbk[x_1, \ldots, x_t]$. For every ideal $\mathfrak{a} \subset R$ we denote $\mathfrak{A} = \Theta^{-1}(\mathfrak{a})$.

Let $\mathfrak{a} \subset R$ be an invertible ideal, denote $h \in \left(1 :_{\Bbbk(C)} \mathfrak{a}\right) = \mathfrak{a}^{-1}$ the nonzero element of $\mathfrak{a}^{-1}$ with smallest minus order $-\mathrm{ord}_{P_\infty}(h)$, and define $\mathfrak{a}^* = h\mathfrak{a}$.

Lemma 3.2 ([4]). *For an invertible ideal $\mathfrak{a}^*$ of R, a nonzero element h of $\mathfrak{a}$ with smallest minus order $-ord_{P_\infty}(h)$ is unique up to constant multiplication.*

It follows that given an ideal class $[\mathfrak{a}]$ the ideal $\mathfrak{a}^*$ is unique, this is the reduced representative of the ideal class.

In order to find the element $h \in \mathfrak{a}^{-1}$ we take an element $f \in \mathfrak{a}$ and find $g \in \left(\langle f\rangle :_{\Bbbk(C)} \mathfrak{a}\right) = \left(\langle f\rangle :_R \mathfrak{a}\right) = f\mathfrak{a}^{-1}$ with smallest minus order $-\mathrm{ord}_{P_\infty}(g)$, then $h = \frac{g}{f}$. The element g is found by computing the Gröbner basis w.r.t. C_A order of the ideal $\Theta^{-1}(f\mathfrak{a}^{-1}) = \Big((f\Bbbk[x_1, \ldots, x_t] + I) :_{\Bbbk[x_1,\ldots,x_t]} \mathfrak{A}\Big)$.

Given two reduced invertible ideals $\mathfrak{a}_1, \mathfrak{a}_2 \subseteq R$, represented with the Gröbner bases $\mathfrak{a}_1 = \langle f_1, \ldots, f_l\rangle, \mathfrak{a}_2 = \langle g_1, \ldots, g_m\rangle$:

(1) $\mathfrak{A}_1 \leftarrow \langle f_1, \ldots, f_l\rangle + I, \quad \mathfrak{A}_2 \leftarrow \langle g_1, \ldots, g_m\rangle + I.$
(2) $\mathfrak{B} \leftarrow \mathfrak{A}_1\mathfrak{A}_2.$
(3) Take a nonzero element $f \in \mathfrak{B}\backslash I.$
(4) Take $g \in \Big((f\Bbbk[x_1, \ldots, x_t] + I) :_{\Bbbk[x_1,\ldots,x_t]} \mathfrak{B}\Big)$ with smallest C_A order.
(5) Compute Gröbner basis of $\mathfrak{B}^* = \frac{g}{f}\mathfrak{B}$.

4. Conclusions

The Arita-Miura-Sekiguchi algorithm for C_A curves does not attempt to replace the faster algorithms for C_{ab} curves, but provides a way to work with more general curves than C_{ab} curves. Nevertheless, it is interesting to see the difference in performance between the AMS algorithm and the fast C_{ab} curves algorithms. AMS algorithm relies on Buchberger's algorithm to find the reduced Gröbner Basis, and this algorithm is hard for analysis,

therefore, it is difficult to perform a theoretical comparison. In this work we implemented some of the C_{ab} curves algorithms and compared them to AMS C_A curves algorithm. The purpose of this implementation is to provide a feeling of the running time.

All computations were performed using Intel Pentium 4 CPU 3.00GHz, 1GB RAM, using Magma V2.11-14.

Examples comparing C_{34} curves can be found in Appendix A.5. The results show that operations that take 0.016 seconds using the FOR and BEFG algorithms take 0.125 seconds using AMS algorithm. This means that AMS algorithms is approximately 8 times slower than the FOR and BEFG algorithms for C_{34} curves.

Acknowledgment

The results of this paper were obtained during my Master studies at Bar-Ilan University. I would like to express deep gratitude to my supervisor Professor Boris Kunyavskii whose guidance and support were crucial for the successful completion of this project.

References

1. F. Abu Salem, K. Makdisi, *Fast Jacobian group operations for C_{34} curves over a large finite field*, LMS J. Comput. Math., **Vol. 10** (2007), 307-328.
2. S. Arita, *Algorithms for computations in Jacobian group of C_{ab} curve and their application to discrete-log-based public key cryptosystems*, The Institute of Electronics, Information and Communication Engineers - IEICE Trans. Fundamentals, **Vol. J82-A, No. 8** (1999), 1291-1299.
3. S. Arita, *An addition algorithm in Jacobian of C_{34} curve*, In Information Security and Privacy, ACISP 2003, Springer-Verlag, LNCS 2727, 2003, pp. 93-105.
4. S. Arita, S. Miura, T. Sekiguchi, *An addition algorithm on the Jacobian varieties of curves*, J. Ramanujan Math. Soc., **Vol. 19, No. 4** (2004), 1-17.
5. R. Avanzi, N. Theriault, Y. Wang, *Rethinking low genus hyperelliptic Jacobian arithmetic over binary fields: interplay of field arithmetic and explicit formulae*,
   http://www.cacr.math.uwaterloo.ca/techreports/2006/cacr2006-07.pdf, 2006.
6. A. Basiri, A. Enge, J.C. Faugère, N. Gürell, *The arithmetic of Jacobian groups of superelliptic cubics*, Math. of Comp., **Vol. 74, No. 249** (2004), 389-410.
7. A. Basiri, A. Enge, J.C. Faugère, N. Gürell, *Implementing the arithmetic of C_{34} curves*, ANTS-VI, Springer-Verlag, LNCS 3076, 2004, pp. 87-101.
8. M. Bauer, *The arithmetic of certain cubic function fields*, Math. of Comp., **Vol. 73, No. 245** (2003), 387-413.

9. B. Byramjee, S. Duqesne, *Classification of genus two curves over $\mathbb{F}_{2^n}$ and optimization of their arithmetic*, Cryptology ePrint Archive, Report 2004/107, 2004, http://eprint.iacr.org.
10. D. Cantor, *Computing in the Jacobian of a hyperelliptic curve*, Math. of Comp., **Vol. 48** (1987), 95-101.
11. H. Cohen, *A course in computational algebraic number theory*, Graduate texts in Math. 138, Springer-Verlag, Berlin, 1993.
12. H. Cohen, A. Miyaji, T. Ono, *Efficient elliptic curve exponentiation using mixed coordinates*, In Proceedings of ASIACRYPT 1998, LNCS 1514, 1998, pp. 51-65.
13. X. Fan, Y. Wang, *Inversion free arithmetic on genus 3 hyperelliptic curves*, Cryptology ePrint Archive, Report 2004/223, 2004, http://eprint.iacr.org.
14. X. Fan, T. Wollinger, Y. Wang, *Efficient doubling on genus 3 curves over binary fields*, Topics in Cryptology - CT-RSA 2006, Springer-Verlag, LNCS 3860 (2006), pp. 64-81.
15. S. Flon, R. Oyono, *Fast arithmetic on Jacobians of Picard curves*, Public Key Cryptography - PKC 2004, Springer-Verlag, LNCS 2947 (2004), pp. 55-68.
16. S. Flon, R. Oyono, C. Ritzenthaler *Fast addition on non-hyperelliptic genus 3 curves*, Cryptology ePrint Archive, Report 2004/118, 2004, http://eprint.iacr.org.
17. S. Galbraith, S. Paulus, N. Smart, *Arithmetic on superelliptic curves*, Math. of Comp., **Vol. 71, No. 237** (2002), 393-405.
18. P. Gaudry, *An algorithm for solving the discrete log problem on hyperelliptic curves*, EUROCRYPT 2000, Springer-Verlag, LNCS 1807, 2000, pp. 19-34.
19. P. Gaudry, N. Thériault, E. Thomé, C. Diem, *A double large prime variation for small genus hyperelliptic index calculus*, , Math. of Comp., **Vol. 76** (2007), 475-492.
20. M. Gonda, K. Matsuo, K. Aoki, J. Chao, S. Tsujii, *Improvements of addition algorithm on genus 3 hyperelliptic curves and their implementation*, In Proc. of SCIS 2004, Japan, 2004.
21. R. Harasawa, J. Suzuki, *Fast Jacobian group arithmetic on C_{ab} curves*, ANTS-IV, Springer-Verlag, LNCS 1838, 2000, pp. 359-376.
22. R. Harley, *Fast addition on genus two hyperelliptic curves*, http://cristal.inria.fr/∼harley/hyper/adding.text, 2000.
23. R. Harley, *Fast doubling on genus two hyperelliptic curves*, http://cristal.inria.fr/∼harley/hyper/doubling.c, 2000.
24. N. Koblitz, *Hyperelliptic cryptosystems*, J. of Cryptology, 1 (1989), 139-150.
25. J. Kuroki, M. Gonda, K. Matsuo, J. Chao, S. Tsujii, *Fast genus three hyperelliptic curve cryptosystems*, In Proc. of SCIS2002, IEICE Japan, pp. 503-507, 2002.
26. T. Lange, *Efficient arithmetic on hyperelliptic curves*, PhD thesis, University Essen, 2001.
27. T. Lange, *Efficient arithmetic on genus 2 hyperelliptic curves over finite fields via explicit formulae*, Cryptology ePrint Archive, Report 2002/121, 2002, http://eprint.iacr.org.

28. T. Lange, *Inversion free arithmetic on genus 2 hyperelliptic curves*, Cryptology ePrint Archive, Report 2002/147, 2002, http://eprint.iacr.org.
29. T. Lange, *Weighted coordinates on genus 2 hyperelliptic curves*, Cryptology ePrint Archive, Report 2002/153, 2002, http://eprint.iacr.org.
30. K. Matsuo, J. Chao, S. Tsujii, *Fast genus two hyperelliptic curve cryptosystems*, Technical report of IEICE, ISEC2001-31, 2001, pp. 89-96.
31. Y. Miamoto, H. Doi, K. Matsao, J. Chao, S. Tsujii, *A fast addition algorithm of genus two hyperelliptic curve*, The 2002 Symp. on Cryptography and Info. Security, Japan, SCIS, pp. 497-502, 2002.
32. D. Mumford, *Tata lectures on theta II - Jacobian theta functions and differential equations*, In Prog. Math. Volume 43, Birkhäuser, 1984.
33. K. Nagao, *Improving group law algorithm for Jacobians of hyperelliptic curves*, W. Bosma ed., ANTS-IV, Springer-Verlag, LNCS 1838, 2000, pp. 439-448.
34. S. Paulus, *Lattice basis reduction in function fields*, In J. Buhler ed., Proceedings of ANTS-III, Springer-Verlag, LNCS 1423 (1998), pp. 567-575.
35. J. Pelzl, T. Wollinger, J. Guajardo, C. Paar, *Hyperelliptic curve cryptosystems: closing the performance gap to elliptic curves*, C. Walter, Ç. Koç, and C. Paar, ed., CHES, Springer-Verlag, LNCS 2779 (2003), pp. 349-365.
36. J. Pelzl, T. Wollinger, C. Paar, *High performance arithmetic for hyperelliptic curve cryptosystems of genus two*, Cryptology ePrint Archive, Report 2003/212, 2003, http://eprint.iacr.org.
37. J. Pelzl, T. Wollinger, C. Paar, *High performance arithmetic for special hyperelliptic curve cryptosystems of genus two*, International Conference on Information Technology: Coding and Computing - ITCC 2004, **Vol. 2**, 2004, pp. 513-557.
38. J. Pelzl, T. Wollinger, C. Paar, *Low cost security: explicit formulae for genus 4 hyperelliptic curves*, Selected Areas in Cryptography SAC 2003, Springer-Verlag, LNCS 3006 (2004), pp. 1-16.
39. M. Rosenlicht, *Generalized Jacobian varieties*, The Annals of Math., 2nd Ser., **Vol. 59, No. 3** (1954), 505-530.
40. J P. Serre, *Algebraic groups and class fields*, Springer-Verlag, 1988.
41. J.H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, New York, 1986.
42. H. Sugizaki, K. Matsuo, J. Chao, S. Tsujii, *An extension of Harley addition algorithm for hyperelliptic curves over finite fields of characteristic two*, IEICE Technical report, ISEC2002-9, pp. 49-56, 2002.
43. H. Sugizaki, K. Matsuo, J. Chao, S. Tsujii, *A generalized Harley algorithm for genus two hyperelliptic curves*, In Proc. of SCIS2003, IEICE Japan, pp. 917-921, 2003.
44. M. Takahashi, *Improving Harley algorithms for Jacobians of genus 2 hyperelliptic curves*, In Proc. of SCIS2002, IEICE Japan, pp. 155-160, 2002.

Appendix A.

Appendix A.1. *Genus 2 hyperelliptic curves – summary*

The following table compares various algorithms for genus 2 hyperelliptic curves. The hyperelliptic curve is of the form

$$y^2 + (h_2x^2 + h_1x + h_0)y = x^5 + f_4x^4 + f_3x^3 + f_2x^2 + f_1x + f_0.$$

Table A1. Genus 2 hyperelliptic curve arithmetic – summary

Algorithm	chr(𝕜)	Properties	Addition	Doubling
1987 Cantor	general		3I + 70M/S	3I + 76M/S
2000 Nagao	odd	regular representation	2I + 52M/S	2I + 59M/S
	odd	alternative representation	I + 56M/S	I + 66M/S
2000 Harley	odd		2I + 24M + 3S	2I + 30M/S
2001 Lange	general		2I + 24M + 3S	2I + 26M + 6S
2001 MCT	odd		2I + 22M + S	2I + 23M + 2S
2002 MDMCT	odd	affine, $f_4 = 0$	I + 24M + 2S	I + 23M + 4S
	odd	projective, $f_4 = 0$	51M + 3S	47M + 6S
2002 Takahashi	odd		I + 23M + 2S	I + 21M + 8S
2002 Lange a	general	$h_2, h_1, f_4 \in \{0, 1\}$	I + 22M + 3S	I + 22M + 5S
	even	$h_2, h_1, f_4 \in \{0, 1\}$	I + 22M + 2S	I + 20M + 5S
2002 Lange b	general	projective	47M + 4S	40M + 6S
	general	mixed	40M + 3S	
2002 Lange c	odd	weighted, $f_4 = 0$	47M + 7S	34M + 7S
	odd	mixed, $f_4 = 0$	36M + 5S	
	even	weighted, $f_4 = 0$, $h_2 \neq 0$	46M + 4S	35M + 6S
	even	mixed, $f_4 = 0$, $h_2 \neq 0$	35M + 5S	
	even	weighted, $f_4 = 0$, $h_2 = 0$	44M + 6S	29M + 6S
	even	mixed, $f_4 = 0$, $h_2 = 0$	34M + 6S	
2002 SMCT	even	$f_4 = f_2 = 0$, $h_2 = 1$	I + 23M + 2S	I + 26M + S
2003 SMCT	general		I + 28M + S	I + 38M
2003 PWP	even	$y^2 + xy = x^5 + f_1x + f_0$		I + 9M +6S
2004 BD	even	affine, general	I + 25M	I +27M
	even	affine, $\deg(H) = 2$	I + 25M	I +26M
	even	affine, $\deg(H) = 1$	I + 24M	I +18M
	even	projective, general	45M	45M
	even	projective, $\deg(H) = 2$	45M	44M
	even	projective, $\deg(H) = 1$	42M	31M
	even	modified, general	45M	43M
	even	modified, $\deg(H) = 2$	45M	42M
	even	modified, $\deg(H) = 1$	42M	31M
	even	weighted, general	42M	46M
	even	weighted, $\deg(H) = 2$	42M	45M
	even	weighted, $\deg(H) = 1$	40M	27M

Appendix A.2. *Genus 3 hyperelliptic curves – summary*

The following table compares various algorithms for genus 3 hyperelliptic curves. The hyperelliptic curve is of the form

$$y^2+(h_3x^3+h_2x^2+h_1x+h_0)y = x^7+f_6x^6+f_5x^5+f_4x^4+f_3x^3+f_2x^2+f_1x+f_0.$$

Table A2. Genus 3 hyperelliptic curve arithmetic – summary

Algorithm	chr(𝕜)	Properties	Addition	Doubling
1987 Cantor	general		4I + 200M/S	4I + 207M/S
2000 Nagao	odd	regular representation	2I + 144M/S	2I + 153M/S
	odd	alternative representation	2I + 157M/S	2I + 170M/S
2002 KGMCT	odd	$f_6 = 0$	I + 81M/S	I + 74M/S
2003 PWGP	general	$h_i \in \{0,1\}$, $f_6 = 0$	I + 70M + 6S	I + 61M + 10S
	even	$h_i \in \{0,1\}$, $f_6 = 0$	I + 65M + 6S	I + 53M + 10S
	even	$H(x) = 1$, $f_6 = 0$	I + 65M + 6S	I + 14M + 11S
2004 GMACT	odd	$f_6 = 0$	I + 67M + 3S	I + 61M + 8S
2004 FW	odd	projective, $f_6 = 0$	132M + 8S	120M + 12S
	odd	mixed, $f_6 = 0$	101M + 7S	
2005 FWW a	odd	projective, $f_6 = 0$	122M + 9S	110M + 11S
	odd	mixed, $f_6 = 0$	105M + 8S	
	even	projective, $H(x) = 1$, $f_6 = 0$	119M + 9S	42M + 15S
	even	mixed, $H(x) = 1$, $f_6 = 0$	102M + 8S	
2005 FWW b	even	$H(x) = 1$		I + 11M + 11S
	even	$H(x) = x$		I + 13M + 13S
	even	$H(x) = x^2$		I + 20M + 12S
	even	$H(x) = x^3$		I + 26M + 11S
2006 ATW	even	classical, $H(x) = 1$, $f_6 = 0$	I + 57M + 6S	I + 11M + 11S
	even	effective, $H(x) = 1$, $f_6 = 0$	I +47.7M + 6S	I + 9.3M + 11S

Appendix A.3. *Genus 4 hyperelliptic curves – summary*

The following table compares various algorithms for genus 4 hyperelliptic curves. The hyperelliptic curve is of the form

$$y^2 + (h_4x^4 + h_3x^3 + h_2x^2 + h_1x + h_0)y = x^9 + f_8x^8 + f_7x^7 +$$
$$f_6x^6 + f_5x^5 + f_4x^4 + f_3x^3 + f_2x^2 + f_1x + f_0.$$

Table A3. Genus 4 hyperelliptic curve arithmetic – summary

Algorithm	chr(𝕜)	Properties	Addition	Doubling
1987 Cantor	general		6I + 386M/S	6I + 395M/S
2000 Nagao	odd	regular representation	3I + 286M/S	3I + 296M/S
	odd	alternative representation	2I + 292M/S	2I + 307M/S
2004 PWP	general	$h_i \in \{0,1\}$, $f_8 = 0$	2I + 160M + 4S	2I + 193M + 16S
	even	$H(x) = x$, $f_8 = 0$	2I + 148M + 6S	2I + 75M + 14S
2006 ATW	even	classical, $H(x) = 1$,	I + 119M +10S	I + 28M + 16S
	even	effective, $H(x) = 1$,	I +98.1M +10S	I +23.7M + 16S

Appendix A.4. *Genus 3 C_{ab} Curves – Summary*

The following table compares various algorithms C_{34} curves.

Table A4. Genus 3 C_{ab} Curves Arithmetic – Summary

Algorithm	Curve	Algorithm Type	Addition	Doubling
1998 GPS	Superelliptic	LLL	20I + 600M	
1999 Arita	C_{ab}	Gröbner basis (Buchberger)		
2000 HS	C_{ab}	LLL		
2001 Bauer	$y^3 = F(x)$	LLL	10I + 547M	
2001 Arita	C_{34}	Explicit formulae	5I + 204M	5I + 284M
2002 BEFG	$y^3 = F(x)$	Cantor based	10I + 200M	
2003 FO	Picard	Explicit formulae	2I + 156M	2I + 174M
2003 BEFG	Picard	Explicit formulae	2I + 140M	2I + 164M
2003 BEFG	C_{34}	Explicit formulae	2I + 150M	2I + 174M
2004 FOR	Picard	Explicit formulae	2I + 130M	2I + 152M
2004 FOR	C_{34}	Explicit formulae	2I + 145M	2I + 167M
2007 ASM	C_{34}	Linear Algebra	2I + 117M	2I + 129M

Appendix A.5. *Implementation Results*

Example 4.1. The first example is a C_{23} (elliptic) curve over a prime field of size of the 250 bits prime number

$$78350495509812662593956461946222915591170600900120350269718238148567069661 3.$$

The elliptic curve is defined as $y^2 = x^3 + ax + b$, where

$$a=67932267643457968166209555940584451919343347299360812115802973037931427095 7,$$

$$b=77528910601603326472479335451945678736513769256954927837743210082920534314 7.$$

The base point P is set to be (x_0, y_0) where

$$x_0=66317733687373309421808144531088250138064976604969512753834018164547535747 0,$$

$$y_0=68975808239733787064730883358065000353293664399791187512477883059733111390 9.$$

The order of P is

$$19587623877453165648489115486555728898111241884175537829271429386661210467 5.$$

The example computes the scalar product mP where m is set to be

$$19587623877453165648489115486555728898111241884175537829271429386661210467 0.$$

Using Magma's built-in elliptic curves arithmetic the computation took 0.016 seconds, and gave the result (x_1, y_1) where

$$x_1=63019862982573127760531339108981032362362387592017401492469040142089174785 6,$$

$$y_1=12917565692266723499615524166679158798653555968872041531702467902052164851 8.$$

Using AMS algorithm the computation took 1.219 seconds, and returned the following Gröbner basis which corresponds to the ideal $\langle x - x_1, y - y_1 \rangle$.

$$\left\{ \begin{array}{l} y+65432929817545939094340937779543756792517044931248308738015770246514904809 5, \\ x+15330632527239534833425122837241883228808213308102948777249198006477894875 7 \end{array} \right\}$$

Example 4.2.

The following example is a C_{34} curve over the prime field of size 25033. The curve is defined to be

$$6567x^3y+y^3+25032x^4+18877x^2y+162xy+4738x^2+14333y+7218x+21234.$$

We represent the divisor $\mathrm{div}(x^3 + u_2x^2 + u_1x + u_0, y - (v_2x^2 + v_1x + v_0))$ by the sextuple $[u_2, u_1, u_0, v_2, v_1, v_0]$.

We select the divisor $D = [3904, 5539, 5752, 19670, 14925, 12954]$ and the scalar $m = 1341$ and compute mP.

Using FOR algorithm the computation took 0.016 seconds and returned the divisor $D_1 = [11095, 5932, 17083, 12380, 15154, 10043]$.

Using AMS algorithm the computation took 0.125 seconds and returned the following Gröbner basis

$$\left\{ \begin{array}{c} y^2 + 17380y + 4174x + 17473, \\ xy + 4646y + 21534x + 13556, \\ x^2 + 840y + 12437x + 14528 \end{array} \right\}$$

which corresponds to the ideal of D_1.

Note that BEFG algorithm is not applicable for this curve, because it is only effective for C_{34} curves where $h_3 = 0$.

Example 4.3.

The following example is a C_{34} curve over the prime field of size 2003. The curve is defined to be

$$y^3 + 2002x^4 + 1550x^2y + 1224xy + 1679x^2 + 856y + 1882x + 1302.$$

We select the divisor $D = [721, 1735, 1698, 1360, 1449, 1465]$ and the scalar $m = 10000$ and compute mP.

Using FOR algorithm the computation took 0.016 seconds and returned the divisor $D_1 = [1164, 1124, 904, 1260, 79, 581]$.

Using BEFG algorithm the computation took 0.016 seconds and returned the divisor $D_1 = [1164, 1124, 904, 1260, 79, 581]$.

Using AMS algorithm the computation took 0.125 seconds and returned the following Gröbner basis

$$\left\{ \begin{array}{c} y^2 + 258y + 921x + 279, \\ xy + 1683y + 50x + 1870, \\ x^2 + 1379y + 1224x + 1064 \end{array} \right\}$$

which corresponds to the ideal of D_1.

Example 4.4.

The following example is a C_{357} curve over the prime field of size 83. C is defined by

$$\left\{ \begin{array}{c} 64+4x+30x^2+30x^3+76y+75xy+y^2+52z+4xz, \\ 10+27x+16x^2+6x^3+44x^4+69y+16xy+27x^2y+31z+xz+yz, \\ 22+32x+77x^2+30x^3+11x^4+17y+25xy+3x^2y+72x^3y+45z+32xz+76x^2z+z^2 \end{array} \right\}$$

We select a point $P = (2, 33, 62)$. According to [4] the order the ideal $I = \langle x - 2, y - 33, z - 62 \rangle$, corresponding to P, is $m = 1848$. We verified this statement. The computation of mI took 0.313 seconds and returned the Gröbner basis $\{1\}$.

Discrete Logarithms, Duality, and Arithmetic in Brauer Groups

Gerhard Frey

IEM
Universiy of Duisburg-Essen
dedicated to Gilles Lachaud on the occasion of his 60th birthday

Duality theorems are in the heart of class field theory both for number fields and geometric objects like curves and abelian varieties. In particular, class groups of rings of integers and group schemes attached to Jacobians of curves are involved in this game. Since these groups are the most popular for producing crypto primitives based on discrete logarithms (which use a priori only the $\mathbb{Z}$-linear structure) these systems carry unavoidably a bilinear structure which offers possibilities of attacks as well as new applications.
This bilinear structure is made explicit by the Lichtenbaum-Tate pairing which links Brauer groups of local fields to torsion points of (generalized) Jacobian varieties over finite fields.
By local class field theory elements in the Brauer group of local fields can be presented by cyclic algebras which are determined by their invariants. We explain how this is related to the classical discrete logarithm in finite fields.
By the duality theorem for number fields and in particular because of the Hasse-Brauer-Noether-Sequence for Brauer groups we can try to globalize and find an Index-Calculus-Algorithm to determine local invariants. We shall apply this both to the discrete logarithm in finite fields and to the computation of the Euler totient function. Moreover it becomes clear that there are "reciprocity laws" for discrete logarithms but it is till now a challenging task to exploit them.*

1. Data Security Meets Class Field Theory

Duality is one of the basic principles for the study of mathematical objects. We give two examples which, a priori, seem to be not related. The purpose of this article is to convince the reader that this is not so, and that their interaction is fruitful for both sides.

*This report is based on a lecture given at the conference **The first Symposium on Algebraic Geometry and its Applications (SAGA 2007)**, May 7 to 11, 2007, at the University of French Polynesia in Papeete, Tahiti.
The author would like to thank the organizers for their warm hospitality in a most beautiful environment which made the conference an unforgettable experience.

1.1. *Class Field Theory*

One of the most fascinating objects in number theory is class field theory ruling over the extensions of number fields which are Galois with abelian Galois groups. The essentials of this theory are formulated in a very elegant way by a fundamental duality theorem involving étale cohomology. For a nice introduction we refer to B. Mazur's paper [16].
Let O be a ring of integers of a number field and $X = \mathrm{Spec}(O)$, let F be a constructible abelian sheaf (i.e. there are finitely many points $\{x_1, \ldots, x_n\}$ such that the pullback of F to $X \setminus \{x_1, \ldots, x_n\}$ and to $\{x_1, \ldots, x_n\}$ is locally constant.
With G_m we denote the group scheme attached to the multiplicative group.

Theorem 1.1. *For $0 \leq i \leq 3$ we have a perfect pairing*

$$H^i_{et}(X, F) \times Ext^{3-i}_X(F, G_m) \to H^3_{et}(X, G_m) \cong \mathbb{Q}/\mathbb{Z}$$

of finite groups.

From this pairing we get duality theorems both for local fields (e.g. finite algebraic extensions of $p-$adic fields) and for global fields (here we consider finite algebraic extensions of $\mathbb{Q}$) which will be explained later on.

1.2. *Discrete Logarithm Systems with Bilinear Structure*

1.2.1. *DL-systems*

Let ℓ be a prime number. A group $(A, \circ)$ of order ℓ is called a DL-system if

i) the elements in A are presented in a compact way, for instance by $O(log(\ell))$ bits,
ii) it is easy to implement the group composition $\circ$ such that it is very fast, for instance has complexity $O(log(\ell))$, but
iii) to compute, for randomly chosen elements $g_1, g_2 \in A$, a number $k \in \mathbb{Z}$ such that $g_2^k = g_1$ (the discrete logarithm problem (DL-problem)) is hard.

In the ideal case this complexity were $exp(O(\log(\ell)))$. This is obtained in black box groups, and, as we hope, in certain groups related to elliptic curves and abelian varieties.
If the complexity is smaller, e.g. subexponential, one may still get usable systems by taking the parameters larger. For instance, one knows that there are algorithms for computing discrete logarithms in the multiplicative group

of finite fields $\mathbb{F}_q$ with q elements which have subexponential complexity in q. One is forced to take q as a number with at least 2000 bits.

1.2.2. *Bilinear Structures*

Let $(A, \circ)$ be a DL-system.

Definition 1.1. Assume that there are $\mathbb{Z}$-modules B and C and a bilinear map

$$Q : A \times B \to C$$

with

i) the group composition laws in A, B and C as well as the map Q are fast computable (e.g. in polynomial time).
ii) Q is non-degenerate in the first variable. Hence, for random $b \in B$ we have $Q(a_1, b) = Q(a_2, b)$ iff $a_1 = a_2$.

Then we call (A, Q) a *DL-system with bilinear structure.*

Remark 1.1. One is used to describe bilinear maps on free modules by matrices whose entries consist of the values of pairs of elements in a fixed basis. This is not enough for our purposes.
For instance assume that $A = B$ is a cyclic group with n elements with generator P_0 and take $C = \mathbb{Z}/n$.
Choose $m \in \mathbb{Z}$ prime to n. Let

$$Q_m : A \times A \to \mathbb{Z}/n$$

be the pairing determined by $Q_m(P_0, P_0) := m + n\mathbb{Z}$.
Without further information the computation of $Q_m(P, Q)$ is equivalent with the Discrete Logarithm in A. So, though from the algebraic point of view pairings are "everywhere" it is much harder to find DL-systems with bilinear structure.

1.2.3. *Some Applications of Bilinear Structures*

There are destructive aspects which may weaken DL-systems if they carry a bilinear structure.
Here is one.

The DL-system $(A, \circ)$ is at most as secure as the discrete logarithm in $(C, \circ)$ [1].

Example 1.1. As we shall see one can transfer discrete logarithms of groups of order ℓ in abelian varieties over $\mathbb{F}_q$ to the multiplicative group μ_ℓ of roots of unity of order ℓ in $\mathbb{F}_{q'} := \mathbb{F}_q(\mu_\ell)$. We have to make sure that q' is large enough so that the subexponential algorithm in $\mathbb{F}_{q'}$ is not feasible.

And there are constructive aspects, for instance

Tripartite Key Exchange [13],
Identity Based Protocols [4], and
Short Signatures [5].

For more information the interested reader is advised to visit Paulo Barretos Pairing Based Crypto Lounge.

2. Geometric Construction of DL-Systems

Let O be a commutative noetherian ring with unit element 1 without zero divisors.
We generalize the notion of ideals of O slightly. Let F be the quotient field of O and $M \subset F$. The set M is an O-ideal if M is an O-module and if there is an element $f \in F^*$ with $fM \subset O$. We multiply ideals as usual.
The ideal M is called invertible if there is an ideal M' with $M \cdot M' = O$.
The set $I(O)$ of invertible ideals is a commutative group called the ideal group of O.
$I(O)$ modulo the subgroup $\mathrm{Princ}(O)$ of invertible principal ideals is the Picard group $\mathrm{Pic}(O)$.
The aim is to find suitable rings O such that for a large prime number ℓ the group $\mathbb{Z}/\ell$ can be embedded into $\mathrm{Pic}(O)$, that the elements in $\mathrm{Pic}(O)$ can be described in a compact way and that the composition in the ideal class group has complexity $O(\log(\ell))$.

There are only two types of rings O used today:

(1) O is an order or a localization of an order in a number field, or
(2) O is the ring of holomorphic functions
of a curve defined over a finite field $\mathbb{F}_q$ with q elements.

We shall restrict ourselves to the second case but we remark that most of the considerations done in the following have analogues in the number field case.

2.1. *Ideal Classes of Function Rings*

Though we are interested in rings of holomorphic functions on curves over finite fields we switch now from $\mathbb{F}_q$ to an arbitrary ground field K.
Let C_O be an absolutely irreducible curve defined over K with function field F and let O the ring of holomorphic functions on C_O. We assume that $Quot(O) = F$ and so C_O is an affine curve. We allow that C_O has (at most) one singular point. Let $\widetilde{C_O}$ be its desingularisation with ring of holomorphic functions $\tilde{O}$. We assume that the conductor ideal of the singular point is, as ideal of $\tilde{O}$, a product of prime ideals occurring with multiplicities at most 1[†]. The ring O is contained in $\tilde{O}$ and $\tilde{O}$ is a Dedekind domain.
There is a unique projective irreducible regular curve C with function field F containing $\widetilde{C_O}$ as affine part.

We use the opportunity to introduce the natural Galois structure coming with the definition of O and then relate the Picard groups of these curves. We extend scalars and interpret C_O as well as $\widetilde{C}$ and C as curves defined over the separable closure K_s of K with corresponding ring of holomorphic functions. [‡] We denote by T_∞ the set $\overline{C}(K_s) \setminus \overline{\widetilde{C}}(K_s)$ and we assume that there is an K-rational point P_∞ in T_∞.
We denote by S the set of points on $\overline{C}(K_s)$ corresponding to the singular point on C_O.

By the theory of Generalized Jacobians and using the Approximation Theorem we relate the ideal class groups of the rings of functions to the points of the Jacobian variety J_C of C by the following exact sequences of Galois modules[§].

Theorem 2.1. *We have the exact sequences of Galois modules*

$$1 \to \mathrm{Princ}(\overline{O}) \to I(\overline{O}) \to \mathrm{Pic}(\overline{O}) \to 0$$

$$1 \to \mathcal{T}_S(K_s) \to \mathrm{Pic}(\overline{O}) \to \mathrm{Pic}(\overline{\widetilde{O}}) \to 0$$

[†]This type of singularity is interesting for cryptography for it enables us to treat tori [21]
[‡]In the following $\overline{*}$ always indicates that we are extending scalars to K_s.
[§]For the definition of Galois modules see Definition 3.1

and

$$0 \to \mathcal{C}_{T_\infty} \to J_C(K_s) \to \mathrm{Pic}(\overline{\overline{O}}) \to 0.$$

where $\mathcal{T}_S$ is a torus of dimension $\mid S \mid -1$ determined by the singularity and $\mathcal{C}_T$ is the ideal class group with support in $T_\infty \setminus P_\infty$.

For $K = \mathbb{F}_q$ all these modules are torsion modules, and we are interested in elements of order dividing n in $\mathrm{Pic}(O)$, which is, since this is a finite group, isomorphic to $\mathrm{Pic}(O)/n \cdot \mathrm{Pic}(O)$ which fits better into our frame. We want to apply the duality theorem 1.1 from Subsection 1.1. For this, it is convenient to work over local fields instead over finite fields. Moreover we shall see that the geometric situation can be made simpler.

2.2. *p-adic Lifting*

Assume that O is the ring of holomorphic functions of an affine curve C_O defined over $\mathbb{F}_q$ with corresponding projective curve C of genus g_0. We assume that C_O has only one singular point with square free conductor. Let S be the set of points on C corresponding to the singular point.
Let K be a local field with residue field $\mathbb{F}_q$.
Let n be a natural number prime to q. We can lift C_O to an affine *non-singular* curve C_O^l defined over K embedded in the projective curve C^l which is a lift of C such that all relevant data are preserved.
In particular,

$$\mathrm{Pic}(O^l)/[n]\,\mathrm{Pic}(O^l)$$

is canonically isomorphic to $\mathrm{Pic}(O)/[n]\,\mathrm{Pic}(O)$, the genus of C^l is

$$g_0 + \mid S \mid -1$$

and there exists a torus T_S/K of dimension $\mid S \mid -1$ and an exact sequence

$$1 \to T_S^l(U_K)/(T_S^l(U_K))^n \to \mathrm{Pic}(O^l)/[n]\,\mathrm{Pic}(O^l) \to \mathrm{Pic}(\tilde{O})/[n]\,\mathrm{Pic}(\tilde{O}) \to 0$$

with U_K the units of K.

Instead of a proof I give an example.

Example 2.1. Take

$$C_O : Y^2 + XY = X^3/\mathbb{F}_q.$$

Then $T_\infty = \{(0,1,0)\}$, the singular point $(0,0)$ corresponds to *two* points on the desingularization and $\mathrm{Pic}(O) \cong \mathbb{F}_q^*$.
Take $K = \mathcal{W}(\mathbb{F}_q)$ as field of Witt vectors over $\mathbb{F}_q$ and choose $\pi \in K$ with

$w_{\mathfrak{p}}(\pi) = 1$.
Then

$$C^l := E : Y^2 + XY = X^3 + \pi$$

is a Tate elliptic curve with

$$E(K) \cong K^*/ < Q_E > \cong U_K.$$

Our aim is to use duality over local fields to get bilinear structures. To do this we first study duality in more detail in a general frame.

3. Duality in Arithmetic Geometry

3.1. *Evaluation Pairings*

Bilinear maps are often obtained by evaluation of functions: Let S be a (non-empty) set and C an abelian group.

$$F(S, C) := \{f : S \to C\}$$

becomes, in a natural way, an abelian group, and the evaluation map

$$S \times F(S, C) \to C$$

is non-degenerate and $\mathbb{Z}$-linear in the second argument.¶
Let $\mathbb{Z}^{(S)}$ be the free abelian group generated by S, its elements are given by functions

$$g : S \to \mathbb{Z}$$

which have value different from 0 only for finitely many elements in S.
Hence we can define the degree $\deg(g) := \sum_{s \in S} g(s)$.
The elements of degree 0 are denoted by $\mathbb{Z}^{(S)^0}$ and form a group.
We can embed S into $\mathbb{Z}^{(S)}$ by sending $s \in S$ to g_s with $g_s(s) = 1$ and $g_s(s') = 0$ for $s \neq s'$.

A function f from S to C can be extended "linearly", i.e. it becomes a homomorphism from $\mathbb{Z}^{(S)}$ to C and is denoted again by f via

$$f : g \mapsto \sum_{s \in S} g(s) \cdot f(s).$$

¶In many contexts both the groups S and C are endowed with a topology. In this case we tacitly assume that all functions are continuous.

In this way, $F(S,C)$ is identified with $\mathrm{Hom}(\mathbb{Z}^{(S)},C)$. We get the non-degenerate evaluation pairing

$$E : \mathbb{Z}^{(S)} \times F(S,C) \to C$$

by

$$E(g,f) \mapsto \sum_{s\in S} g(s)\cdot f(s).$$

Now assume that S is a group.
We restrict the evaluation map from $F(S,C)$ to $\mathrm{Hom}(S,C)$, the group of homomorphisms from S to C and get

$$E_0 : S \times \mathrm{Hom}(S,C) \to C.$$

E_0 is linear and non-degenerate in the second argument.
As function of the first argument, it is a group homomorphism.
Since C is assumed to be abelian every homomorphism vanishes on the commutator subgroup S' of S, and hence D gives rise to a pairing

$$D : S/S' \times \mathrm{Hom}(S,C) \to C.$$

Example 3.1. Take $C = \mathbb{R}/\mathbb{Z}$ with the discrete topology and S a topological group.
The (topological) group $\mathrm{Hom}(S,\mathbb{R}/\mathbb{Z})$ of continuous homomorphisms is called the Pontryagin dual S^* of S.
Since $\mathbb{R}/\mathbb{Z}$ is an injective $\mathbb{Z}$-module the pairing

$$D : S/S' \times S^* \to \mathbb{R}/\mathbb{Z}$$

is non-degenerate in both variables.

3.2. *Arithmetical Duality*

We add more structure.
Let K be a field of characteristic $p \geq 0$.
Let K_s be the separable closure of K and $G_K = \mathrm{Aut}_K(K_s)$ the absolute Galois group of K.
This is a topological group with profinite topology and hence it is compact.
First we introduce the notion of Galois modules.

Definition 3.1. A Galois module (G_K-module) M is a discrete $\mathbb{Z}$-module with continuous G_K-action. In particular, this implies that

$$M = \bigcup_U M^U$$

where $U < G$ of finite index.

Attached to M is the functor

$$\mathcal{M} : \{ \text{ fields between } K \text{ and } K_s \} \mapsto \{ \text{ Abelian groups } \}$$

sending L to M^{G_L}.

Example 3.2. Assume that $\mathcal{A}$ is an étale commutative group scheme defined over K.
Then $A = \mathcal{A}(K_s)$ is a G_K-module and the attached functor $\mathcal{M}$ is given by

$$\mathcal{M}(L) = \mathcal{A}(L)$$

for L separable over K.

A basic example for this is the multiplicative group G_m which is represented by an affine plane curve given by the coordinate ring $K[X,Y]/(XY-1)$. For commutative algebras R over K we have

$$G_m(R) = R^*, \text{ the group of invertible elements in } (R, \cdot)$$

and so $G_m(L) = L^* = K_s^{*G_L}$.

Remark 3.1. A finite Galois module is always represented by an (affine) étale commutative group scheme, and conversely, the K_s-rational points of a finite étale commutative group scheme are a finite Galois module.

Definition 3.2. Let A, B, C be Galois modules, and

$$Q : A \times B \to C$$

a pairing.
Q is a Galois pairing iff for all $(a,b) \in A \times B$ and $\sigma \in G_K$ we have

$$Q(\sigma \circ a, \sigma \circ b) = \sigma \circ Q(a,b).$$

Let S be a set endowed with the discrete topology with continuous G_K-action, and let C be a Galois module. Then $F(S,C)$ becomes a Galois module by the action

$$f^\sigma := \sigma \circ f \circ \sigma^{-1}.$$

We apply this definition to $\mathbb{Z}^{(S)}$ ($\mathbb{Z}$ as Galois module with trivial action) and check that the evaluation pairing

$$E : \mathbb{Z}^{(S)} \times F(S,C) \to C$$

is a Galois pairing.

Let A, C be G_K-modules. Restricting to homomorphisms we endow

$$\mathrm{Hom}(A, C)$$

with a G_K-module in a natural way and the pairing

$$D : A \times \mathrm{Hom}(A, C) \to C$$

is a Galois pairing.

Let A be a finite Galois module.
We could apply the above definitions to the Pontryagin dual A^* by taking $C = \mathbb{R}/\mathbb{Z}$ as trivial G_K-module. More interesting from the arithmetical point of view is to take $C = K_s^*$.

Definition 3.3. Let A be a finite G_K-module.
The Cartier dual of A is $\widehat{A} := \mathrm{Hom}(A, K_s^*)$.

Example 3.3. Let n be a number prime to $\mathrm{char}(K)$ and $A = \mathbb{Z}/n$ with trivial Galois action.
Then the Pontryagin dual of A is again $\mathbb{Z}/n$ but its Cartier dual is μ_n, the group of roots of unity of order dividing n with natural Galois action.
So as $\mathbb{Z}$-module we get the same groups but the Galois action differs as soon as K does not contain all n-th roots of unity.

Theorem 3.1. *The evaluation pairing $D : A \times \widehat{A} \to K_s^*$ is a non-degenerate Galois pairing. If $\mathcal{A}$ is a finite étale group scheme with order prime to* $\mathrm{char}(K)$ *then $\widehat{\mathcal{A}} := \mathrm{Hom}(\mathcal{A}, G_m)$ is an étale group scheme (the Cartier dual of $\mathcal{A}$) and $\widehat{\mathcal{A}(K_s)} = \widehat{\mathcal{A}}(K_s)$.*

In particular, the Cartier dual of $\widehat{A}$ is A. So $\widehat{\mu_n} = \mathbb{Z}/n$.
Here is another key example. Let $\mathcal{A}$ be be an abelian variety defined over K. Take

$$\mathcal{A}[n] := ker(n \circ id_{\mathcal{A}}).$$

We have a Kummer sequence of étale group schemes

$$0 \to \mathcal{A}[n] \to \mathcal{A} \to \mathcal{A} \to 0$$

yielding the exact sequence

$$0 \to \mathcal{A}(K_s)[n] \to A(K_s) \to A(K_s) \to 0$$

of Galois modules.
There is an abelian variety $\widehat{\mathcal{A}}$ dual to $\mathcal{A}$ such that, in a canonical way, $\widehat{(\mathcal{A}[n])}$ is isomorphic to $\widehat{\mathcal{A}}[n]$.

In particular, we get a non-degenerate Galois pairing between the points of order dividing n of $\mathcal{A}(K_s)$ and $\widehat{\mathcal{A}}(K_s)$.
An important special case is that $\mathcal{A}$ is a principally polarized abelian variety and then it is canonically isomorphic to $\widehat{\mathcal{A}}$ (e.g., $\mathcal{A}$ is the Jacobian variety of a curve). In this case $\mathcal{A}[n]$ is self-dual.

3.2.1. *Computational Aspects*

- In general it is not clear how to compute the evaluation pairing fast.
- In special cases (e.g. if $\mathcal{A}$ is a subscheme of an abelian variety) there is an explicit and fast evaluation function based on the Weil pairing.
  But even in this case one has to deal with objects in large extension fields L of K in general (e.g.,$L = K(\mathcal{A}[n](K_s))$) even though one is only interested in the group of K-rational points. In general it is not true that the restriction of the pairing to $\mathcal{A}(K) \times \mathcal{A}(K)$ is non-degenerate.
- Caution: Assume that the exponent of A is n and that K contains the n-th roots of unity hence μ_n is isomorphic to $\mathbb{Z}/n$. Assume that we can compute the the evaluation pairing fast. Then we can transfer the arithmetic from A to μ_n but not necessarily to $\mathbb{Z}/n$ since there will be in general no isomorphism between μ_n and $\mathbb{Z}/n$ which can be computed fast. This leads to the question of discrete logarithms in μ_n.

3.3. *Galois Cohomology and Induced Pairings*

In this section we take G as profinite group. Of course $G = G_K$ is the motivating example. We shall use the elementary properties of Galois cohomology.[‖]

3.3.1. *Etale Cohomology Around the Corner*

We recall that the duality theorem in Section 1.1 makes a statement about étale cohomology. Galois cohomology is a special case of this cohomology. Take $X = Spec(K)$. Etale (connected) covers of X are separable extension fields L of K with the induced map

$$Spec(L) \to Spec(K).$$

They define the "open sets" of the étale topology of $Spec(K)$. Galois modules A define sheaves via the section functor

$$\Gamma(Spec(L), A) := A^{G_L}.$$

[‖] See [22], or [7]

The functor Γ is left-exact and there are enough flask sheaves (injective modules) and so we get a sheaf cohomology $H^n_{et}(X, A)$ resp. $H^n_{et}(X, \mathcal{A})$ which is nothing but the Galois cohomology of $A(= \mathcal{A}(K_s))$.
Hence we can embed Galois cohomology into the much wider and flexible frame of étale cohomology, and we can use results about étale cohomology groups to get results about Galois cohomology.

3.3.2. *Pairings in Cohomology*

Let A and B be two $G-$modules.
The tensor product (over $\mathbb{Z}$)

$$A \otimes B$$

becomes, in a natural way, a G-module.
We have a natural (and functorial) homomorphism $\cup^{0,0}$ from $A^G \otimes B^G$ to $(A \otimes B)^G$.
$\cup^{0,0}$ induces a unique family of homomorphisms

$$\cup^{p,q} : H^p(G, A) \times H^q(G, B) \to H^{p+q}(G, A \otimes B)$$

with functorial properties with respect to cohomology functors.
$\cup^{p,q}$ is called the cup product (for details see [21]).

Now assume that there is a G-pairing

$$Q : A \times B \to C.$$

Q defines a G-homomorphism ϕ_Q from $A \otimes B$ to C by sending $a \otimes b$ to $Q(a, b)$, and so we get induced homomorphisms $\phi_Q^{(n)}$ on cohomology groups.
Define a bilinear pairing from $H^p(G, A) \times H^q(G, B)$ to $H^{(p+q)}(G, C)$ by

$$Q^{p,q} = \phi_Q^{(n)} \circ \cup^{p,q}.$$

Example 3.4. Let p, q be non-negative integers with $p + q = 2$.

i) The evaluation pairing induces a pairing

$$D^{p,q} : H^p(G_K, A) \times H^q(G_K, \widehat{A}) \to H^2(G_K, K_s^*).$$

ii) Let S be a G_K-set. The evaluation pairing induces a pairing

$$E^{p,q} : H^p(G_K, \mathbb{Z}^{(S)}) \times H^q(G_K, F(S, K_s^*)) \to H^2(G_K, K_s^*).$$

The cohomology group $H^2(G_K, K_s^*)$ is important for the arithmetic of K, it is called the Brauer group of K and denoted by $\mathrm{Br}(K)$. We shall discuss it in Section 4.3.1 in more detail. One feature of Theorem 1.1 is that for number fields and local fields the cohomological derivates of the evaluation pairing relate Brauer groups to Galois modules in a natural way.

3.3.3. *The Tate Pairing*

We assume for sake of simplicity that J is a principally polarized abelian variety, for instance the Jacobian variety of a projective absolutely irreducible non-singular curve C, defined over K. Let, as always, $n \in \mathbb{N}$ be prime to $\mathrm{char}(K)$. As part of the long exact cohomology sequence attached to the Kummer sequence we have the exact sequence

$$0 \to J(K)/n \cdot J(K) \xrightarrow{\delta^0} H^1(G_K, J[n](K_s)) \to H^1(G_K, J(K_s))[n] \to 0.$$

The map δ^0 is given explicitly by the following recipe: to $P \in J(K)$ choose $Q \in J(K_s)$ with $n \cdot Q = P$.
Then $\delta^0(P + n \cdot J(K))$ is the cohomology class of the cocycle

$$\sigma \mapsto (\sigma(Q) - Q) \text{ for } \sigma \in G_K.$$

As seen above we have

$$E^{1,1} : H^1(G_K, J[n](K_s)) \times H^1(G_K, J[n](K_s)) \to \mathrm{Br}(K)[n].$$

One checks that $\delta^0(J(K)/n \cdot J(K))$ is isotrop w.r.t $E^{1,1}$ and so $E^{1,1}$ induces the *Tate-pairing*

$$T_n : J(K)/n \cdot J(K) \times H^1(G_K, J(K_s))[n] \to \mathrm{Br}(K)[n].$$

3.4. *The Lichtenbaum Pairing*

In this section we describe the work of Lichtenbaum presented in [15].
Let C be an absolutely irreducible non-singular projective curve defined over K with function field F with a $K-$rational point P_∞ (for simplicity). By $\overline{C}$ we denote the curve obtained from C by extending scalars to K_s. The function field of $\overline{C}$ is $\overline{F} = F \cdot K_s$.
G_K is acting in a natural way on $\overline{F}$ and $\overline{C}(K_s)$ with fixed sets F and $C(K)$. We apply the discussions in 3.2 to subsets $T \subset \overline{C}(K_s)$ which are assumed to be G_K-invariant, and interpret $\overline{F}_T \subset \overline{F}$, defined as the group of functions on $\overline{C}$ without zeroes and poles in T, as Galois invariant subset in $F(T, K_s^*)$,

the set of all maps from T to K_s^*.
The evaluation pairing

$$E_T : \mathbb{Z}^{(T)} \times \overline{F}_T \to K_s$$

is a Galois pairing inducing (for $p+q=2$) a pairing

$$E_T^{p,q} : H^p(G_K, \mathbb{Z}^{(T)}) \times H^q(G_K, \overline{F}_T) \to \mathrm{Br}(K).$$

Since C is assumed to be regular we can identify $\mathbb{Z}^{(T)}$ with the group $\overline{D}_T$ of divisors on $\overline{C}$ with support in T. For $T = \overline{C}(K_s)$ denote $\overline{D}_T$ by $\overline{D}$, the divisor group of $\overline{C}$. We recall that for elements in $\mathbb{Z}^{(T)}$ we have defined their degree and remark that this definition is the usual definition of the degree of a divisor as sum of the "multiplicities" of points occurring in the divisor. The group of divisors of degree 0 is denoted by $\overline{D}^0$.
The principal divisor of $f \in \overline{F}^*$ is denoted by (f) and is defined by

$$(f)(P) = w_P(f)$$

where $w_P(f)$ is the order of of vanishing of f in $P \in \overline{C}(K_s)$ resp. the negative of the order of the pole of f in P.
Principal divisors have degree 0 and form a subgroup $\mathrm{Princ}(\overline{C})$ of $\overline{D}^0$.
It is easy to see that $H^1(G_K, \overline{D}_T) = 0$ and so $p = q = 1$ is not interesting. In this paper we shall be interested in

$$E_T := E_T^{0,2} : D_T \times H^2(G_K, \overline{F}_T) \to \mathrm{Br}(K)$$

where $D_T = H^0(G_K, \mathbb{Z}^{(T)})$ can be identified with the group of K-rational divisors on C with support in T.

3.4.1. *Key Example*

We assume that L/K is a cyclic extension of order n with Galois group $G =< \tau >$.
Because of Hilbert's Theorem 90 the inflation from $H^2(G, (F \cdot L)_T)$ to $H^2(G_K, \overline{F}_T)$ is injective.
We have the following explicit description of $H^2(G, (F \cdot L)_T)$: every cohomology class c contains a cocycle given by

$$\zeta_f(\tau^i, \tau^j) := 1 \text{ if } i + j < n$$

$$\zeta(\tau^i, \tau^j) := f \text{ if } i + j \geq n$$

with $f \in F_T$. ζ_f lies in the same class as ζ_g iff

$$f \cdot g^{-1} \in \mathrm{Norm}_{F \cdot L/F}(F \cdot L).$$

Hence the restriction of E_T to $D_T \times H^2(G_K, (F \cdot L)_T)$ is given by

$$(D, c) \mapsto [f(D)]$$

where $[f(D)]$ is the inflation of the class of the cocycle $\zeta(\tau^i, \tau^j)$ with

$$\zeta(\tau^i, \tau^j) = 1 \text{ if } i + j < n$$

and

$$\zeta(\tau^i, \tau^j) = f(D) \text{ if } i + j \geq n.$$

This cocycle is a factor system for a cyclic algebra split by L (Subsection 4.3.1).

3.4.2. *The Brauer Group of C*

We would like to extend the pairing E_T to a pairing with $\overline{F}^*$ as domain for the second argument. The idea is to use that for a given finite set S of points on the projective non-singular $\overline{C}$ and a given K-rational divisor D on C we can always find a function $h \in F$ with principal divisor (h) such that $D + (h)$ is prime to S.
Beginning with $c \in H^2(G_K, \overline{F}^*)$ we represent c by a cocycle ζ determined by finitely many functions $f(\sigma, \tau)$ as values. This is possible since because of continuity there is a finite Galois extension L/K such that c is the inflation of an element $c^0 \in H^2(G(L/K), F \cdot L)$. Let S be the finite set of points on $\overline{C}$ which occur as zeroes of these functions, and take $T = \overline{C} \setminus S$.
Next, for given K-rational divisor D on C we choose a function $h \in F$ such that $D_h := D + (h)$ is, as divisor on C, prime to S. So $D_h \in D_T$ and $E_T(D_h, c)$ is an element in $\mathrm{Br}(K)$.
But this element may depend on the choice of h!
The question is: Let $h \in F$ be a function such that the principal divisor (h) is in D_T. Is the class of the cocycle ζ_0 given by

$$\zeta_0(\sigma, \tau) := f(\sigma, \tau)((h)); \sigma, \tau \in G(L)$$

trivial?
We use Weil's reciprocity law and get

$$f(\sigma, \tau)((h)) = h((f(\sigma, \tau))$$

and, since h is invariant under G_K, the class of ζ_0 is trivial if the class of

$$\zeta_1 : G(L/K) \times G(L/K) \to \overline{D}$$

given by

$$\zeta_1(\sigma,\tau) = (f(\sigma,\tau))$$

in $H^2(G_K, \overline{D})$ is trivial.

Definition 3.4. The Brauer group $\mathrm{Br}(C)$ of C is the kernel of the map

$$\alpha : H^2(G_K, \overline{F}^*) \to H^2(G_K, \overline{D})$$

induced by sending a function f on C to its principal divisor (f).

By the discussion above we see that we can define a pairing from $D \times \mathrm{Br}(C)$ in $\mathrm{Br}(K)$ by using appropriate pairings E_T and changing elements in D by principal divisors. By definition the resulting pairing depends only on the divisor class of the K-rational divisors on C and so get

Proposition 3.1. *Let C be a non-singular absolutely irreducible curve over K with divisor class group* $\mathrm{Pic}(C)$.
Then the evaluation map induces a pairing

$$E : \mathrm{Pic}(C) \times \mathrm{Br}(C) \to \mathrm{Br}(K).$$

In many cases one is interested in $\mathrm{Pic}^0(C)$, the group of K-rational divisor classes of degree 0. This group consists of the classes of divisors on C of degree 0 modulo the group of principal divisors. We observe that the evaluation of a function f at a divisor of degree 0 depends only on (f), and so E induces a pairing, also denoted by E, from $\mathrm{Pic}^0(C) \times \overline{\mathrm{Br}}(C)$ where $\overline{Br}(C)$ is the image of $\mathrm{Br}(C)$ in $H^2(G_K, \mathrm{Princ}(\overline{C}))$ induced by the map $f \mapsto (f)$.

Corollary 3.1. *The evaluation pairing induces a pairing*

$$E : \mathrm{Pic}^0(C) \times \overline{\mathrm{Br}}(C) \to \mathrm{Br}(K).$$

It remains to describe elements in $\overline{\mathrm{Br}}(C)$.
We use the exact G_K-module sequence

$$0 \to \mathrm{Princ}(\overline{C}) \to \overline{D}^0 \to \mathrm{Pic}^0(\overline{C}) \to 0$$

and get (since $H^1(G_K, \overline{D}^0) = 0$)

$$0 \to H^1(G_K, \mathrm{Pic}^0(\overline{C})) \stackrel{\delta^1}{\to} H^2(G_K, \mathrm{Princ}(\overline{C})) \to H^2(G_K, \overline{D}^0))$$

where δ^1 is the connecting homomorphism from $H^1(G_K, \mathrm{Pic}^0(\overline{C}))$ to $H^2(G_K, \mathrm{Princ}(\overline{C}))$ resulting from cohomology.
It follows that $\overline{\mathrm{Br}}(C) = \delta^1(H^1(G_K, \mathrm{Pic}^0(\overline{C})))$.

Proposition 3.2. *Let C be a non-singular absolutely irreducible projective curve.*
The evaluation pairing between points and functions on C induces a pairing

$$T_L : \mathrm{Pic}^0(C) \times H^1(G_K, \mathrm{Pic}^0(\overline{C})) \to \mathrm{Br}(K).$$

This pairing is called the *Lichtenbaum pairing.*

Explicit Pairing Since for computational purposes it is important to describe the pairing explicit we do this here.
Take $c \in H^1(G_K, \mathrm{Pic}^0(\overline{C}))$, represent it by a cocycle

$$\zeta : G_K \to \mathrm{Pic}^0(\overline{C}) \text{ with } \zeta(\sigma) = \bar{D}(\sigma)$$

and choose

$$D(\sigma) \in \bar{D}(\sigma).$$

The divisor

$$A(\sigma_1, \sigma_2) = \sigma_1(D(\sigma_2)) + D(\sigma_1) - (D(\sigma_1 \cdot \sigma_2))$$

is a principal divisor $(f(\sigma_1, \sigma_2))$ and $\delta^1(c)$ is the cohomology class of the 2-cocycle

$$\gamma : (\sigma_1, \sigma_2) \mapsto (f(\sigma_1, \sigma_2)).$$

For $c \in H^1(G_K, \mathrm{Pic}^0(\overline{C}(K_s))$ choose $D_0 := \sum_{P \in \overline{C}(K_s)} z_P \cdot P \in \bar{D}_0 \in \mathrm{Pic}^0(C)$ such that $\delta^1(c)$ is presented by a cocycle $(f(\sigma_1, \sigma_2))$ prime to D_0. Then $T_L(\bar{D}_0, c)$ is the cohomology class of the cocycle

$$\zeta(\sigma_1, \sigma_2) = \sum_{P \in \overline{C}(K_s)} f(\sigma_1, \sigma_2)(P)^{z_P}$$

in $H^2(G_K, K_s^*)$.

3.4.3. *Example*

We give the analogue of Example 3.4.1 for the Lichtenbaum pairing.

Example 3.5. Let L/K be cyclic of degree n and $G(L/K) =< \tau >$.
By $\mathrm{Pic}^0(C_L)$ we denote the divisor class group of degree 0 of $C \times L$.
Let ζ be a 1-cocycle from $< \tau >$ into $\mathrm{Pic}^0(C_L)$ representing the cohomology class $c \in H^1(G(L/K), \mathrm{Pic}^0(C_L))$.

ζ is determined by the value $\zeta(\tau) =: z$ since the cocycle condition implies that $\zeta(\tau^j) = \sum_{i=0}^{j-1} \tau^i z$ for $1 \leq j \leq n$. In particular,we get

$$Trace_{L/K}(z) = 0.$$

Choose a divisor $D \in z$ and $D(\tau^j) := \sum_{i=0}^{j-1} \tau^i D$. Then

$$Trace_{L/K}(D) = (f_D) \text{ with } f_D \in F.$$

Hence $\delta^1(c)$ is presented by the cocycle

$$f(\tau^i, \tau^j) = 1 \text{ for } i + j < n$$

and

$$f(\tau^i, \tau^j) = (f_D) \text{ for } i + j \geq n.$$

Next choose in the divisor class $\bar{D}_0 \in \mathrm{Pic}^0(C)$ a divisor D_0 with $D_0 = \sum_{P \in \overline{C}(K_s)} z_P \cdot P$ prime to the set of zeroes and poles of f_D.
Then $T_L(\bar{D}_0, c) \in H^2(G(L/K), L^*)$ is presented by the cocycle

$$\eta(\tau^i, \tau^j) = 1 \text{ for } i + j < n$$

and

$$\eta(\tau^i, \tau^j) = \prod_{P \in C(K_s)} f_D(P)^{z_P} \in K^* \text{ for } i + j \geq n.$$

This is a cocycle defining a cyclic algebra with center K and splitting field L.

3.4.4. *Comparison Theorem*

We have defined two pairings attached to Jacobian varieties, namely the Tate pairing T_n which uses crucially the Weil pairing on torsion points of the Jacobian of order n, and the Lichtenbaum pairing T_L which uses evaluation of functions on the curve. A priori, no number n appears in the latter pairing but we can look at it modulo n and get for all natural numbers prime to $\mathrm{char}(K)$

$$T_{L,n} : \mathrm{Pic}^0(C)/n \cdot Pic^0(C) \times H^1(G_K, \mathrm{Pic}^0(\overline{C}))[n] \to \mathrm{Br}(K)[n].$$

Lichtenbaum proves in [15]

Theorem 3.2. *Up to a sign, the pairing $T_{L,n}$ is equal to T_n.*

We shall call $T_{L,n}$ the Lichtenbaum-Tate pairing. For most purposes its interpretation by evaluation of functions on C is used.

3.4.5. *Pairings for Non-complete Curves*

We return to the more general situation that O is the ring of holomorphic functions of an affine curve C_O over K which may have singularities of a restricted type as defined in Subsection 2.1. Instead of divisor and divisor classes we work with invertible ideals and their classes and use the Galois module sequences in Theorem 2.1. By overcoming some technical difficulties (for details see [7]) we can generalize the definition of the Lichtenbaum-Tate pairing to this situation and get

Theorem 3.3. *For all n prime to* $\mathrm{char}(K)$ *there is the Lichtenbaum-Tate pairing*

$$T_{L,n} : \mathrm{Pic}(O)/n\,\mathrm{Pic}(O) \times H^1(G_K, \mathrm{Pic}(\overline{O}))[n] \to \mathrm{Br}(K)[n].$$

4. The Pairing over Local Fields

4.1. *Duality*

We apply Theorem 1.1 to finite Galois modules A over local fields K with residue field $\mathbb{F}_q$. We assume that the order of A is prime to q. In the language of Subsection 1.1 we interpret O_K, the ring of integers of K, as localization of the ring of integers of a number field, and look at abelian sheaves F trivial outside of $Spec(O_K)$.
$Spec(O_K)$ is a one-dimensional scheme with a closed point corresponding to the maximal ideal $\mathfrak{p}$, i.e. to $Spec(\mathbb{F}_q)$, and a generic point corresponding to $Spec(K)$ as open subscheme of X.
Galois modules over X consist of a generic fiber, i.e. a Galois module over G_K, a special fiber, which is a Galois module over the residue field, and a reduction map.
Etale neighborhoods are given by unramified extensions of O_K, by Galois extensions of K and by Galois extensions of k.
The duality theorem takes care of these data. Restricting to the generic fiber we come home to Galois cohomology.
We get the *Duality Theorem of Tate*:

Theorem 4.1.

(1) $H^3_{et}(X, G_m)$ *is isomorphic (in a natural way) to the Brauer group* $H^2(G_K, K_s^*) = \mathrm{Br}(K)$ *and hence this group is isomorphic to* $\mathbb{Q}/\mathbb{Z}$.
(2) Let A be a finite G_K-module with Cartier dual $\widehat{A}$.
Then for $0 \le i \le 2$ *the cohomology groups* $H^i(G_K, A)$ *are finite and*

the evaluation pairing induces non-degenerate pairings

$$H^i(G_K, A) \times H^{2-i}(G_K, \widehat{A}) \to \mathrm{Br}(K).$$

As a consequence of this duality theorem Tate proves in [24]

Theorem 4.2. *Let J be an abelian variety (for simplicity principally polarized). The Tate pairing*

$$T_n : J(K)/nJ(K) \times H^1(G_K, J(K_s))[n] \to \mathrm{Br}(K)$$

is a non-degenerate pairing.

Using Theorem 3.2 and results from [7] we get

Corollary 4.1. ***(Lichtenbaum-Tate)*** *Let K be a local field, C_O an affine regular curve over K with ring of holomorphic functions O.*
For every natural number n the Lichtenbaum-Tate pairing

$$T_{L,n} : \mathrm{Pic}(O)/n\,\mathrm{Pic}(O) \times H^1(G_K, \mathrm{Pic}(\overline{O}))[n] \to \mathrm{Br}(K)[n]$$

is non-degenerate.

This result encourages to investigate the modules $H^1(G_K, \mathrm{Pic}(\overline{O}))[n]$ and $\mathrm{Br}(K)[n]$ (known to be isomorphic to $\mathbb{Z}/n$) as well as the pairing from a computational point of view.

4.2. $H^1(G_K, \mathrm{Pic}(\overline{O}))[n]$

We assume (and this is so for all cryptographically interesting cases cf.[7]) that we can replace $\mathrm{Pic}(\overline{O})$ by $\mathcal{A}(K_s)$ where $\mathcal{A}$ is an abelian variety defined over K and isogenous to J_C.
$\mathcal{A}$ extends to a group scheme over O_K, its Néron model.
As always we assume that n is prime to q and to simplify the situation we assume in the whole section in addition that n is prime to the number of connected components of the special fiber of $\mathcal{A}^{**}$.
Let K_{nr} be the maximal unramified extension of K and denote by ϕ_q the Frobenius automorphism.
Then

$$H^1(G(K_{nr}/K), \mathrm{Pic}(O_{K_{nr}}))[n] = 0.$$

Via restriction we embed $H^1(G_K, \mathrm{Pic}(\overline{O}))$ into $H^1(G_{K_{nr}}, \mathrm{Pic}(\overline{O}))$, and the image is equal in the subgroup of elements which are ϕ_q-invariant (ϕ_q acts

**It can be interesting to study what happens if the last condition is not satisfied.

by conjugation on $G_{K_{nr}}$). Let L_n be the unique extension of K_{nr} of degree n which is totally ramified. It is equal to $K_{nr}(\pi^{1/n})$ where π is a uniformizing element of K.
So $G(L_n/K_{nr}) =< \tau_n >$ with $\tau_n(\pi^{1/n}) = \zeta_n \cdot \pi^{1/n}$ for an n-th root of unity ζ_n. The Frobenius automorphism ϕ_q acts by conjugation on τ sending τ to τ^q since q is the value of the cyclotomic character applied to ϕ_q.
These datas are sufficient to describe $H^1(G_K, Pic(\overline{O}))[n]$ in all concrete cases.
Here are two examples. First assume that J_C and hence $\mathcal{A}$ have good reduction. In this case $\mathcal{A}[n] := \mathcal{A}(K_s)[n] = \mathcal{A}(L_n)[n] = \mathcal{A}(K_{nr})[n]$ and hence

$$H^1(G_K, Pic(\overline{O}))[n] = Hom_{<\phi_q>}(< \tau >, \mathcal{A}[n]).$$

Definition 4.1. Let $\mathrm{Pic}(O)[n]^{(q)}$ be the subgroup in $\mathrm{Pic}(\overline{O})[n]$ consisting of elements c with $\phi_q(c) = q \cdot c$.

Proposition 4.1. *Assume that O is the ring of holomorphic functions of a regular affine curve with good reduction. Let n be prime to q.*
Then $H^1(G_K, Pic(\overline{O}))[n]$ is isomorphic to $\mathrm{Hom}(< \tau >, \mathrm{Pic}(O)[n]^{(q)})$*, and so, non-canonically since depending on the choice of τ, to* $\mathrm{Pic}(O)[n]^{(q)}$*.*

Corollary 4.2.

i) *If $\zeta_n \in K$ then $H^1(G_K, Pic(\overline{O}))[n]$ is isomorphic to* $\mathrm{Pic}(O)[n]$*.*
ii) *Let L be any extension field of K totally ramified of degree n.*
Then $H^1(G_K, \mathrm{Pic}(\overline{O}))[n]$ is equal to the kernel of the restriction map from G_K to G_L.

For general curves C_O it is more complicated to describe the result. One complication is that the torus part of the special fiber of J_C is in general not split. A complete treatment is possible in principle but not in the frame of this survey. So we restrict ourselves to an important example and take as ring O the holomorphic functions on Tate elliptic curves given by affine equations

$$E_Q : Y^2 + XY = X^3 + Q.$$

with $w_{\mathfrak{p}}(Q) = m \in \mathbb{N}$.
Since only one point is missing we get that $\mathrm{Pic}(\overline{O})$ is Galois isomorphic to $E_Q(K_s)$.
We assume that n is prime to m. Then $E_Q(K)$ contains elements of order n iff $\zeta_n \in K$, and hence by duality

$$H^1(G_K, E_Q(K_s))[n] \neq 0 \text{ iff } \zeta_n \in K,$$

and in this case it is cyclic of order n.
So we assume that $\zeta_n \in K$.
We take a special cyclic extension of degree n, namely $L_Q := K(Q^{1/n})$. By Tate's theory this field is equal to $K(E_Q[n])$.

Proposition 4.2. *Let τ be a generator of $G(L_Q/K)$, let $P \in E_Q[n]$ be any point of order n which is not K-rational, and let ζ be the cocycle from $< \tau >$ to $E_Q[n]$ determined by $\zeta(\tau) = P$.*
Then $H^1(G_K, E_Q(K_s))[n]$ is cyclic of order n and generated by the class of ζ.

4.3. *The Local Brauer Group*

4.3.1. *Cyclic Algebras*

For the moment let K be any field of characteristic prime to n.
The elements in the Brauer group $\mathrm{Br}(K)$ of K can be identified with classes of central simple algebras with center K. The group composition is the tensor product, and the trivial class consists of all algebras which are isomorphic to full matrix algebras over K.

Because of Hilbert's theorem 90 one sees that for Galois extensions L/K the inflation map from $H^2(G(L/K), L^*)$ to $\mathrm{Br}(K)$ is injective.
For any L/K the restriction map from $\mathrm{Br}(K)$ to $\mathrm{Br}(L)$ corresponds to base field extension applied to algebras, and its kernel consists of the classes of algebras which become, after tensoring with L, isomorphic to full matrix algebras. In this case L is called a splitting field of K. We have been confronted at different places with the special case that L/K is cyclic of degree n, for instance in Example 3.5 as result of the Lichtenbaum-Tate pairing.
In this case $H^2(G(L/K), L^*)$ consists of classes of *cyclic* algebras with 2-cocycles given in the following way:
Let σ be a generator of $G(L/K)$ and take a in K^*.
The map $f_{\sigma,a} : G \times G \to L^*$, given by

$$f_{\sigma,a}(\sigma^i, \sigma^j) = \begin{cases} 1 & \text{for} \quad i+j < n \\ a & \text{for} \quad i+j \geq n \end{cases}$$

is a cocycle.
The cocycles $f_{\sigma,a}$ and $f_{\sigma,a'}$ are in the same cohomology class if and only if $a \cdot a'^{-1} \in N_{L/K}L^*$. We denote the corresponding class of cyclic algebras by $(L, \sigma, a \cdot N_{L/K}L^*)$.

4.3.2. *Invariants*

Now we return to the case that K is a local field with residue field $\mathbb{F}_q$ and that n is prime to q.
Because of the local duality theorem we know already that $\mathrm{Br}(K)[n] \cong \mathbb{Z}/n$.

The unramified case: the invariant Let L_u be the unique unramified extension of K of degree n. It is cyclic.
$G(L_u/K)$ has as canonical generator the Frobenius automorphism ϕ_q.
We represent elements $c \in H^2(G(L_u/K), L_u^*)$ by a triple

$$(L_u, \phi_q, a \cdot N_{L_u/K}(L_u^*)).$$

Since

$$K^*/N_{L_u/K}(L_u^*) \cong < \pi > / < \pi^n >$$

the class of c is uniquely determined by $w_{\mathfrak{p}}(a) \bmod n$.

Definition 4.2. $w_{\mathfrak{p}}(a) \in \mathbb{Z}/n\mathbb{Z}$ is the *invariant* $\mathrm{inv}_K(c)$ of c.

The general case: the invariant From the paragraph above it follows that

$$\mathrm{Br}(K)[n] = inf_{L_u/K_s}(H^2(G(L_u/K), L_u^*)).$$

Definition 4.3. The map

$$\mathrm{inv}_K : Br(K)[n] \to \mathbb{Z}/n$$

is defined as follows:
For $c \in \mathrm{Br}(K)[n]$ take $c_0 \in H^2(G(L_u/K), L_u^*)$ with $\inf_{L_u/K_s}(c_0) = c$ and represent c_0 by a triple $(L_u, \phi_q, a \cdot N_{L_u/K}(L_u^*))$.
Then $\mathrm{inv}_K(c) := w_{\mathfrak{p}}(a) \mod n$ is well defined and determines c uniquely.

Though the invariant is defined in a seemingly very explicit way for cyclic algebras split by unramified extensions it may be difficult to compute it even in this case. To see this assume that τ is another generator of $G(L_u/K)$ and the cyclic algebra representing c is given by the triple $(L_u, \tau, a \cdot N_{L_u/K}(L_u^*))$.
We know that there exists $k \in \mathbb{Z}$ with $\tau^k = \phi_q$. Then $\mathrm{inv}_K(c) = k \cdot w_{\mathfrak{p}}(a) \bmod n$.
So we have to determine k, and this is a discrete logarithm problem.

4.3.3. *The Tamely Ramified Case*

The relation to Discrete Logarithms in finite fields becomes even more evident in the ramified case.
Let L_n a totally ramified Galois extension of degree n of K. It follows that L_n/K is cyclic and that $\mu_n \subset K$. Let τ be a fixed generator of $G(L_n/K)$. Since

$$K^*/N_{L_n/K}(L^*) \cong \mathbb{F}_q^*/\mathbb{F}_q^{*n}$$

the element $c \in H^2(G(L_n/K), L_n^*)$ is determined by the triple

$$(L_n, \tau, a \in \mathbb{F}_q^*/\mathbb{F}_q^{*n}).$$

Proposition 4.3. *For* $a_1, a_2 \in \mathbb{F}_q$

$$a_1^k \equiv a_2 \mod \mathbb{F}_q^{*n}$$

iff

$$k \cdot (\mathrm{inv}_K((L_n, \tau, a_1 \cdot \mathbb{F}_q^{*n}))) \equiv \mathrm{inv}((L_n, \tau, a_2 \cdot \mathbb{F}_q^{*n})) \text{ modulo } n.$$

Hence the computation of Discrete Logarithms in $\mathbb{F}_q^*$ *is equivalent with the computation of invariants of cyclic algebras.*

4.3.4. *The Frobenius Case*

The most important case for applications is that $c \in \mathrm{Br}(K)$ is represented as algebra split by an extension L of K which is totally ramified of degree n and which becomes Galois only after adjoining the n-th roots of unity. This is exactly the situations which occurs when one applies the Lichtenbaum-Tate pairing.
A description useful for algorithmic purposes is, at the moment, only available if one restricts c to $K(\zeta_n)$ and then uses the results obtained for cyclic ramified extensions over $K(\zeta_n)$ instead of K. Hence one has to pass to a field which will be, in general, much larger than K!
It is a challenge to do better.

5. The Bilinear Structure

Having information about the groups involved in the Lichtenbaum-Tate pairing over local fields we give it now in an explicit way, first over local fields, and then as application to cryptography, over finite fields.

5.1. *Algorithmic Description of the Lichtenbaum-Tate Pairing*

5.1.1. *The Pairing over Local Fields*

We continue to assume that K is a local field with residue field $\mathbb{F}_q$.
To make the situation not too complicated we discuss as example the case that the curve C has good reduction (hence is the lift of a nonsingular curve C_0 over $\mathbb{F}_q$) and that only one point "at infinity" is missing on C_O. So we have a non-degenerate pairing

$$T_{L,n} : J_C(K)/nJ_C(K) \times H^1(G_K, J_C(K_s))[n] \to \mathrm{Br}(K)[n].$$

Let k be the smallest number with $q^k \equiv 1 \mod n$. k is called the "embedding degree".
Define $K(\zeta_n) := K_n$. It is a local field with residue field $\mathbb{F}_{q^k}$.
We choose a uniformizing element $\pi \in K$, define $L := K_n(\pi^{1/n})$ and take τ as generator of $G(L/K_n)$.
As seen in Subsection 4.2 we can identify $H^1(G_K, J_C(K_n))[n]$ with group of homomorphisms

$$\{\varphi \in \mathrm{Hom}(G_K, J_C(K_n)[n]) \text{ with } \varphi(\tau) = P \text{ and } \phi_q(P) = q \circ P\}.$$

We use the explicit description of the Lichtenbaum pairing given in Example 3.5.
Take $c \in H^1(G_K, J_C(K_s))$ corresponding to φ with $\varphi(\tau) = P$ and $n \cdot P = (f_P)$.
Take $Q \in \overline{Q} \in J_C(K)$ such that $f_P(Q)$ is defined.
Then $T_{L,n}(\overline{Q}, c)$ is the class of cyclic algebra $(L, \tau, f_P(Q) \cdot N_{L/K_n}(L^*))$.
Hence we get a non-degenerate pairing

$$T_{n,0} : J_C(K)/n \cdot J_C(K) \times J_C(K_s)[n]^{(q)} \to \mathbb{F}_{q^k}^*/(\mathbb{F}_{q^k}^*)^n$$

by the evaluation *modulo* $\mathfrak{p}$ of the functions f_P with $(f_P) = n \cdot P$ and $P \in J_C(K_s)[n]^{(q)}$ on $J_C(K)$.

5.1.2. *The Pairing over Finite Fields*

Now begin with a curve C_O defined over $\mathbb{F}_q$. Since we evaluate functions on a p-adic lift C^l of C_O *modulo the maximal ideal* $\mathfrak{p} \subset O_K$ we get an explicit description of the Lichtenbaum-Tate pairing in the case of good reduction which only uses objects attached to the curve C_O. Using Corollary 4.1 we can generalize (see [7]) and get

Theorem 5.1. *Assume that C_O is an affine curve with one singular point over $\mathbb{F}_q$ whose conductor is square free. Let O be the ring of holomorphic functions O. Take n prime to q (and satisfying some "innocent" [††] extra conditions).*
Then we get a non-degenerate pairing

$$T_n : \mathrm{Pic}(O)/n \cdot \mathrm{Pic}(O) \times J_{C^l}(K_s)[n]^{(q)} \to \mathbb{F}^*_{q^k}/(\mathbb{F}^*_{q^k})^n$$

which is given by the evaluation modulo $\mathfrak{p}$ of a function f_P with $(f_P) = n \cdot P$ and $P \in J_{C^l}(K_s)[n]^{(q)}$ on $\mathrm{Pic}(O)$. If C_O is regular P and f_P can be replaced by their reduction modulo $\mathfrak{p}$, i.e by points and functions over $\mathbb{F}_q$, and we get a pairing

$$T_{n,0} : \mathrm{Pic}(O)/n \cdot \mathrm{Pic}(O) \times J_C(\mathbb{F}_{q_s})[n]^{(q)} \to \mathbb{F}^*_{q^k}/(\mathbb{F}^*_{q^k})^n.$$

5.1.3. *Evaluation*

To get a bilinear structure on class groups of rings of holomorphic functions on curves over $\mathbb{F}_q$ there is a last step to be done. One has to show that the computation of the pairing is fast.

As we have seen in Subsection 5.1.2 one has to evaluate a function f_P contained in the function field $F \cdot \mathbb{F}_{q^k}$ at a divisor D on C defined over $\mathbb{F}_q$ to compute T_n.
A naive approach is, because of the high degrees needed in practice, not possible. The way to reduce the problem to a square-and-multiply algorithm in a group was found by *V. Miller* for elliptic curves (applied to the Weil pairing). The general method uses as background the theory of Mumford's Theta groups which describe extensions of (finite subgroups of) abelian varieties by linear groups (see [11]).
For details and further accelerations we refer to [1].

5.2. *Conclusion*

Let O be the ring of holomorphic functions of an affine curve over a finite field $\mathbb{F}_q$.
Let n be a number prime to q, and let k be the smallest number such that $n \mid q^k - 1$.

Theorem 5.2. *The Lichtenbaum-Tate pairing induces a bilinear structure on* $\mathrm{Pic}(O)$ *of complexity* $O(k \cdot log(q))$ *with value group* $\mathrm{Br}(K)$.

[††]For a rigorous formulation see [7]

Corollary 5.1. *The discrete logarithm in* $\mathrm{Pic}(O)$ *is reduced, with costs* $O(k \cdot log(q))$*, to the discrete logarithm in* $\mathrm{Br}(K)$ *and hence to the computation of invariants in* $\mathrm{Br}(K)$*, or alternatively, to the computation of discrete logarithms in* $\mathbb{F}^*_{q^k}/\mathbb{F}^{*n}_{q^k}$.

To apply these results (and to have a bilinear structure in the strong sense) it is necessary that k is not too big.
In particular, for the constructive applications it is necessary to have an embedding degree $\sim 12 \cdot g$. It is a very nice problem in computational number theory to find such k. For elliptic curves we have rather satisfying results [3].
But for $g > 1$ nearly nothing is known if J_C is not supersingular.
A successful approach to this problem could be interesting since one can speed up the computation of T_n by a factor g in interesting protocols [9].

6. Global Situation

Let us formulate the quintessence of Section 4.3: If we could compute the invariants of algebras over local fields fast we could solve the Discrete Logarithm problem in systems attached to curves over finite fields. But this leads to the hard problem to compute discrete logarithms in large finite fields (at least in general). A possible way to overcome this difficulty is to lift again, now from local fields to global fields.

6.1. *Duality over Global Fields*

Let K be a global field, i.e. either a finite extension of $\mathbb{Q}$ or a function field of one variable over a finite field. To simplify we shall assume that K is a number field but the function field case can be treated analogously and is very interesting (key word: function field sieve).
An important method in Number Theory is to study global objects by passing to local ones. We shall go the converse way: to study local invariants we shall try to globalize.
Let Σ_K be the set of all places of K (including archimedean places). For $\mathfrak{p} \in \Sigma_K$ we denote by $K_\mathfrak{p}$ the completion of K at $\mathfrak{p}$. We choose an extension $\tilde{\mathfrak{p}}$ of $\mathfrak{p}$ to K_s and identify the decomposition group of $\tilde{\mathfrak{p}}$ with $G_{K_\mathfrak{p}}$.
Let A be a G_K-module.
We have restriction maps

$$\rho_\mathfrak{p} : H^n(G_K, A) \to H^n(G_{K_\mathfrak{p}}, A).$$

Define

$$f_n(A) : H^n(G_K, A) \stackrel{\prod \rho_{\mathfrak{p}}}{\to} \prod_{\mathfrak{p} \in \Sigma_K} H^n(G_{K_{\mathfrak{p}}}, A).$$

The key questions are: Describe the kernel and the cokernel of f_n! Here are consequences of Theorem 1.1 [19]. Assume that A is finite.

- The kernel of $f_1(A)$ is compact and dual to the kernel of $f_2(\widehat{A})$. In particular, the kernel of $f_2(A)$ is finite.
- We have an exact 9-term sequence, the Duality Theorem of Tate-Poitou

$$0 \to A^{G_K} \to \prod H^0(K_{\mathfrak{p}}, A) \to H^2(K, \widehat{A})^* \to H^1(K, A) \to \prod{}' H^1(K_{\mathfrak{p}}, A)$$

$$\to H^1(K, \widehat{A})^* \to H^2(K, A) \to \sum H^2(K_{\mathfrak{p}}, A) \to H^0(K, \widehat{A})^* \to 0.$$

  (Here G_K is replaced by K, and $\prod'$ is the restricted product with respect to the unramified cohomology.)
- The Hasse-Brauer-Noether Sequence:
  For all natural numbers n the sequence

$$0 \to \mathrm{Br}(K)[n] \stackrel{\oplus_{\mathfrak{p} \in \Sigma_K} \rho_{\mathfrak{p}}}{\longrightarrow} \bigoplus_{\mathfrak{p} \in \Sigma_K} \mathrm{Br}(K_{\mathfrak{p}})[n] \stackrel{\Sigma_{\mathfrak{p} \in \Sigma_K} \mathrm{inv}_{\mathfrak{p}}}{\longrightarrow} \mathbb{Z}/n \to 0$$

  is exact.

6.2. *Reciprocity Laws*

We use the sequence of Hasse-Brauer-Noether.

Proposition 6.1. *Assume that we have a curve C_O defined over K with properties as above.*
Take $c \in \mathrm{Pic}(O)$ and $\varphi \in H^1(G_K, \mathrm{Pic}(\overline{O}))[n]$ with localizations $c_{\mathfrak{p}}$ respectively $\varphi_{\mathfrak{p}}$.
Then

$$\sum_{\mathfrak{p} \in \Sigma_K} \rho_{\mathfrak{p}}(inv_{\mathfrak{p}}(T_{L,n}(c, \varphi))) = \sum_{\mathfrak{p} \in \Sigma_K} inv_{\mathfrak{p}}(T_{L,n}(c_{\mathfrak{p}}, \varphi_{\mathfrak{p}})) = 0.$$

Hence we have relations between local discrete logarithms modulo different places both on abelian varieties, e.g. elliptic curves, and in the multiplicative group.

The hope is that by these reciprocity laws we can compute discrete logarithms in geometrically defined groups A_q defined over $\mathbb{F}_q$ by first lifting them to groups $A_{\mathfrak{p}}$ over a local field $K_{\mathfrak{p}}$ with residue field $\mathbb{F}_q$, then lifting further to a global field K and finally passing to other places $\{\mathfrak{p}'\} \subset \Sigma_K$ where this computation is easier.
To realize this idea we have to find global geometric objects over K with given reduction modulo $\mathfrak{p}$ which are arithmetically accessible. And then we need "enough" test functions φ to exploit Proposition 6.1. This leads to hard problems in global number theory, and in the moment it is totally open whether anything useful will come out of this approach for abelian varieties.
The situation is much better if we look at the classical Discrete Logarithm in the multiplicative group of $\mathbb{F}_q$. Our global geometric object is the algebraic group G_m, and we are working with the duality theorem 1.1 with $p = q = 1$ (i.e. we use evaluation pairings with Dirichlet characters).
This approach is taken in the paper of Huang and Raskind in [12]. In the light of their results a realistic hope is that one can shift the computation of discrete logarithms in roots of unity of order n in arbitrary fields $\mathbb{F}_q$ with $n \mid q-1$ to fields $\mathbb{F}_{q'}$ with q' not much larger than n.
Here we give an obvious result.

Proposition 6.2. *Let $\mathfrak{m}$ be a divisor of K. We* assume *that there is a cyclic extension L of odd degree n of K which is unramified outside of the set $T_{\mathfrak{m}}$ of places in the support of $\mathfrak{m}$.*
Let τ be a generator of $G(L/K)$. For $\mathfrak{p} \notin T_{\mathfrak{m}}$ let $\phi_{\mathfrak{p}}$ be a Frobenius automorphism at $\mathfrak{p}$ in $G(L/K)$. By $f_{\mathfrak{p}}$ we denote a number for which $\tau^{f_{\mathfrak{p}}} = \phi_{\mathfrak{p}}$ holds. For all elements $a \in K^*$ we have

$$\sum_{\mathfrak{p} \in T_{\mathfrak{m}}} \mathrm{inv}_{\mathfrak{p}}(A)_{\mathfrak{p}} \equiv -\left(\sum_{\mathfrak{p} \notin T_{\mathfrak{m}}} w_{\mathfrak{p}}(a)) f_{\mathfrak{p}} \right) \mod n$$

where $w_{\mathfrak{p}}$ is the normalized valuation in $\mathfrak{p}$ and A is the cyclic algebra $(L, \tau, a \cdot N_{L/K}(L^*))$.

6.3. *Application*

If we can compute (enough of) the numbers $f_{\mathfrak{p}}$ we can compute

- the order of the ray class group of the order in K with conductor $\mathfrak{m}$, in particular Euler's totient function $\varphi(m)$
- the discrete logarithm in $\mathbb{F}_q^*$ if $\mathfrak{m}$ is a prime with residue field $\mathbb{F}_q$,

and

- get a very subtle descriptions of cyclic extensions of K.

6.4. *Index-Calculus in Global Brauer Groups*

Motivated by Proposition 6.2 we search for (heuristic) algorithms to determine the numbers $f_{\mathfrak{p}}$ which characterize the Frobenius automorphisms at places $\mathfrak{p}$ of K related to cyclic extensions with conductor dividing an ideal $\mathfrak{m}$.
A possible method to do this (with subexponential complexity) is an index-calculus algorithm of the type one is used to see in factorization algorithms. We give a simple variant.

6.5. *Example: $K = \mathbb{Q}$*

Take $K = \mathbb{Q}$. We use the notation and assumptions of Proposition 6.2. The congruence in this proposition can be seen as solution of a system of linear equations relating the variables f_p for p prime to m and $\mathrm{inv}_p(A)$ for $p \mid m$. Note that the system is solvable modulo n since a cyclic extension unramified outside of m exists by assumption.
Let d be the smallest natural number $\geq \sqrt{m}$.
For small δ take $a_1(\delta) := d + \delta$, $a_2(\delta) := c_0 + 2\delta \cdot d + \delta^2$ with $c_0 = d^2 - m$.
Then at primes dividing m the invariants of the cyclic algebras attached to a_1^2 and a_2 are equal, and so for primes p prime to m the corresponding numbers f_p are solutions modulo n of the equations

$$L_\delta : \sum_{p \in \mathbb{P}, p \nmid m} (2w_p(a_1(\delta)) - w_p(a_2(\delta)))X_p = 0.$$

We want to get equations with coefficients equal to 0 for $p > B$ for a certain convenient bound B[‡‡], i.e. the numbers a_1 and a_2 have to be B-smooth. Let S be the number of primes $\leq B$.
Now choose a relatively small number L and search $\delta \leq L$ (using sieves) yielding such smooth pairs $(a_1(\delta), a_2(\delta))$.
Assume that we have found a system $\mathcal{L}$ of S $\mathbb{Z}$-independent equations.

Proposition 6.3. *$det(\mathcal{L})$ is a multiple of $\varphi(m)$.*

In general this multiple will be rather big.
In a master thesis in Essen (2006) A. Timofeev did many experiments.

[‡‡] B has to be large enough so that we can expect that not all primes $\leq p$ are split in L

The nice result was that after applying the algorithm twice in all experiments the gcd of the two determinants was equal to $\varphi(m)$.

6.6. *Construction of Elements in the Brauer Group of Global Fields*

Motivated by the reciprocity laws and index-calculus, we are looking for more methods to construct elements in the Brauer group of global fields. The theoretical background for the success (or failure) is the duality theorem of Tate-Poitou. We can try to use

- Pairings with Dirichlet Characters [12]
- Pairings with Principal Homogenous Spaces with abelian varieties instead of using the multiplicative group. The arithmetic of abelian varieties predicts that this is much more difficult than using characters. Perhaps Euler systems attached to Heegner points could be interesting sources.
- As variant one could study Cassel's Pairing using Tate-Shafarevich groups and ending in the second cohomology group of the idele class group which is in fact the right global object from the point of view of class field theory.
- Instead of Brauer groups of curves one could try to use Brauer groups of higher dimensional varieties. Interesting beginnings for this can be found in [14].

References

1. R. Avanzi, H. Cohen, C. Doche, G. Frey, T. Lange, K. Nguyen, and F. Vercauteren. *The Handbook of Elliptic and Hyperelliptic Curve Cryptography*. CRC, 2005.
2. P. S. L. M. Barreto, B. Lynn, and M. Scott. *Constructing elliptic curves with prescribed embedding degrees*. In Security in Communication Networks – SCN 2002, volume 2576 of Lecture Notes in Comput. Sci., pages 257–267. Springer-Verlag, Berlin, 2003.
3. P. S. L. M. Barreto and M. Naehrig. *Pairing-friendly elliptic curves of prime order*. preprint, 2005.
4. D. Boneh and M. Franklin. *Identity based encryption from the Weil pairing*. SIAM J. Comput., 32(3):586–615, 2003.
5. D. Boneh, B. Lynn, and H. Shacham. *Short signatures from the Weil pairing*. In Advances in Cryptology – Asiacrypt 2001, volume 2248 of Lecture Notes in Comput. Sci., pages 514–532. Springer-Verlag, Berlin, 2002.
6. G. Frey. *Applications of arithmetical geometry to cryptographic constructions*. In Finite fields and applications (Augsburg, 1999), pages 128–161. Springer, Berlin, 2001.

7. G. Frey. *On the relation between Brauer groups and discrete logarithms*. Tatra Mt. Math. Publ., 33: 199-227, 2006
8. G. Frey and T. Lange. *Mathematical background of public key cryptography*. In Séminaires et Congrès SMF: AGCT 2003, num.11 (2005), pages 41-74.
9. G. Frey and T. Lange. *Fast Bilinear Maps from the Tate-Lichtenbaum Pairing on Hyperelliptic Curves*. In Proc. ANTS VII, LNCS 4076, pages 466-479. Springer, Berlin 2006.
10. G. Frey, M. Müller, and H. G. Rück. *The Tate pairing and the discrete logarithm applied to elliptic curve cryptosystems*. IEEE Trans. Inform. Theory, 45(5):1717–1719, 1999.
11. G. Frey and H. G. Rück. *A remark concerning m-divisibility and the discrete logarithm problem in the divisor class group of curves*. Math. Comp., 62:865–874, 1994.
12. M.-D. Huang and W. Raskind. *Signature calculus and discrete logarithm problems*. In Proc. ANTS VII, LNCS 4076. Springer, Berlin 2006.
13. A. Joux. *A one round protocol for tripartite Diffie–Hellman*. In Proc.ANTS IV, LNCS 1838, pages 385–394. Springer, Berlin 2000.
14. A. Kresch and Y. Tschinkel. *On the Arithmetic of Del Pezzo Surfaces of Degree 2*. Proc. LMS (3), 89, pages 545-569, 2004.
15. S. Lichtenbaum. *Duality theorems for curves over p-adic fields*. Invent. Math., 7, pages 120–136, 1969.
16. B. Mazur. *Notes on étale cohomology of number fields*. Ann. sci. ENS t.6, n^o 4 ,pages 521-552, 1973.
17. V.C. Miller. *The Weil Pairing, and Its Efficient Calculation*. J. Cryptology, 17:235–261, 2004.
18. D. Mumford. *Abelian Varieties*. Oxford University Press 1970.
19. Jürgen Neukirch. *Algebraic number theory*. Springer, 1999.
20. K. Nguyen. *Explicit Arithmetic of Brauer Groups, Ray Class Fields and Index Calculus*. PhD thesis, University Essen, 2001.
21. J.P. Serre. *Groupes algébriques et corps de classes* Hermann, Paris, 1959.
22. J.P. Serre. *Corps locaux*. Hermann, Paris, 1962.
23. H. Stichtenoth. *Algebraic Function Fields and Codes*. Springer, 1993.
24. J. Tate. *WC-groups over* $\mathfrak{p}$*-adic fields*. Séminaire Bourbaki, Exposé 156, vol. 13, 1958.

On generators of the group $\widehat{H}^{-1}(\mathrm{Gal}(L/K), E_L)$ in some abelian p-extension L/K

Emmanuel Hallouin

Laboratoire Emile Picard,
Institut de Mathématiques de Toulouse, France.
E-mail : halloin@univ-tlse2.fr

Marc Perret

Laboratoire Emile Picard,
Institut de Mathématiques de Toulouse, France.
E-mail : perret@univ-tlse2.fr

À Gille Lachaud, en l'honneur de ses soixante ans

Introduction

The main motivation of this work is Shafarevich theorem on class fields towers, as in the spirit of [4], Chap I, §4.4. Let L/K be a unramified (here, unramifiedness refers also to the infinite primes throughout) Galois extension of number fields whose Galois group G is a finite p-group (p a prime integer). We know that:

$$d_pH^3(G,\mathbb{Z}) = d_pH^2(G,\mathbb{Z}/p\mathbb{Z}) - d_pH^1(G,\mathbb{Z}/p\mathbb{Z}).$$

where $d_p\mathcal{G}$ denotes the p-rank of a finite p-group $\mathcal{G}$. If moreover the class number of L is not divisible by p then:

$$d_pH^3(G,\mathbb{Z}) \leq r_1 + r_2 \tag{1}$$

where (r_1, r_2) is the signature of the number field K. Briefly, the proof works as follows. Let C_L be the idèle class group of L and E_L its unit group, then:

$$\forall q \in \mathbb{Z}, \qquad \widehat{H}^q(G,C_L) \simeq \widehat{H}^{q+1}(G,E_L) \quad \text{and} \quad \widehat{H}^q(G,C_L) \simeq \widehat{H}^{q-2}(G,\mathbb{Z}).$$

The first isomorphism follows from the fact that L has a class number not divisible by p while the second one is part of class field theory. Thus:

$$\widehat{H}^{q+1}(G,E_L) \simeq \widehat{H}^{q-2}(G,\mathbb{Z}). \tag{2}$$

The inequality (1) comes from the specialization at $q = -1$ of this isomorphism since the rank of $\widehat{H}^0(G, E_L) = E_K/N_{L/K}(E_L)$ is easily bounded thanks to Dirichlet's unit theorem.

Together with Golod-Shafarevich inequality, which states that

$$d_p H^2(G, \mathbb{Z}/p\mathbb{Z}) > (d_p H^1(G, \mathbb{Z}/p\mathbb{Z}))^2/4,$$

inequality (1) implies that:

$$\frac{(d_p H^1(G, \mathbb{Z}/p\mathbb{Z}))^2}{4} - d_p H^1(G, \mathbb{Z}/p\mathbb{Z}) < r_1 + r_2.$$

A famous consequence is the following: if a number field K satisfies the quadratic (in $d_p \operatorname{Cl}(K)$) inequality:

$$(d_p \operatorname{Cl}(K))^2/4 - d_p \operatorname{Cl}(K) \geq r_1 + r_2,$$

then its p-class field tower is infinite.

A cubic (in $d_p \operatorname{Cl}(K)$) infinitness criterion of the p-class field tower over a field k already exists (see [2], proof of corollary 10.8.11, chapter 10). Unfortunately, it works only if there is an action of $Gal(k/k_0)$ for a quadratic subfield k_0 of k. In order to find an unconditional cubic analogue of this criterion, one can specialize the isomorphism (2) at $q = -2$. This yields the following equality:

$$d_p \widehat{H}^{-1}(G, E_L) = d_p H^3(G, \mathbb{Z}/p\mathbb{Z}) - d_p H^2(G, \mathbb{Z}/p\mathbb{Z}) + d_p H^1(G, \mathbb{Z}/p\mathbb{Z}).$$

Hence, it is crucial as a first step to find an upper bound for the p-rank $d_p \widehat{H}^{-1}(G, E_L)$ when $\operatorname{Cl}(L)$ is trivial. In this paper, we prove results about generators of this group in some special cases. More precisely, we compute the p-rank and exhibit an explicit basis of $\widehat{H}^{-1}(G, E_L)$ when L/K is an unramified abelian p-extension whose Galois group has exactly two generators..

Notation — Let K be a number field. We denote by Σ_K the set of its finite places, $\operatorname{Div}(K)$ its ideal group and $\operatorname{Cl}(K)$ its ideal class group. To each finite place $v \in \Sigma_K$ one can associate a unique prime ideal $\mathfrak{p}_v$ of K and to each $x \in K^*$, there corresponds a principal ideal $\langle x \rangle_K$ of K.

If L/K is a Galois extension of number fields, then for each $v \in \Sigma_K$, $\Sigma_{L,v}$ denotes the subset of places $w \in \Sigma_L$ above v (for short $w \mid v$) and f_v the residual degree of any $w \in \Sigma_{L,v}$ over K. The map $j_{L/K} : \operatorname{Div}(K) \to \operatorname{Div}(L)$ is the usual extension of ideals.

Let G be a finite group and M be a multiplicative G-module. The norm map $N_G : M \to M$ is defined by $x \mapsto \prod_{g \in G} g(x)$; its kernel is denoted

by $M[N_G]$. The augmentation ideal $I_G M = \left\langle \frac{g(x)}{x}, x \in M, g \in G \right\rangle$ is of importance. Of course, one has $I_G M \subset M[N_G]$; the quotient of these two subgroups is nothing else than the Tate cohomology group:

$$\widehat{H}^{-1}(G, M) \stackrel{\text{def.}}{=} \frac{M[N_G]}{I_G M}$$

in which we are interested (see [3] for an introduction to the negative cohomology groups). For $u \in M[N_G]$, we denote by $[u]$ the class of u in $\widehat{H}^{-1}(G, M)$.

1. The cyclic case

Let L/K be a cyclic extension with Galois group $G = \langle g \rangle$. A classical consequence of Hilbert 90 theorem states that the kernel of the norm N_G equals the augmentation ideal: $L^*[N_G] = I_G L^*$. In cohomological terms, this means that:

$$H^1(G, L^*) = \{1\} \qquad \Longrightarrow \qquad \widehat{H}^{-1}(G, L^*) = \{1\}.$$

Another easy consequence already known is that:

Proposition 1.1. *Let L/K be an unramified cyclic extension with Galois group $G = \langle g \rangle$. Then the map:*

$$\begin{array}{ccccc} \varphi_g : \mathrm{Ker}(\mathrm{Cl}(K) \to \mathrm{Cl}(L)) & \longrightarrow & \widehat{H}^{-1}(G, E_L) \\ [I] & \longmapsto & \left[\frac{g(y)}{y}\right] \end{array},$$

is a group isomorphism, where $[I]$ denotes the ideal class of I and y is any generator of the extension of I to L.

Proof. The only non-trivial assertion is the surjectivity of the map. Let $u \in E_L[N_G]$, then there exists $y \in L^*$ such that $u = \frac{g(y)}{y}$. Thus the ideal $\langle y \rangle_L$ is fixed by the action of G. The extension L/K being unramified, the ideal $\langle y \rangle_L$ is the extension to L of some ideal I of K: $j_{L/K}(I) = \langle y \rangle_L$. Then $[u] = \varphi_g([I])$. □

This proposition implies the following corollary:

Corollary 1.1. *Let K be a number field whose ideal class group is a cyclic p-group and L be its Hilbert class field. Suppose that L has class number one. Then for any generator g of $\mathrm{Gal}(L/K)$ and any generator π of a prime ideal*

of L whose Frobenius equal to g, $\widehat{H}^{-1}(G, E_L)$ is a cyclic p-group generated by the class of $\sigma(\pi)/\pi$:

$$\widehat{H}^{-1}(G, E_L) = \left\langle \left[\frac{g(\pi)}{\pi} \right] \right\rangle.$$

2. Some experiments with `magma`

With the help of `magma` and `pari/gp`, we have made some experiments and collect datas about the 2-rank of the group $\widehat{H}^{-1}(G, E_{K^i})$ in unramified finite 2-extensions K^i/K $(i = 1, 2)$. In each case, we start with a quadratic complex number field K whose class group is a 2-group; tables of such fields can be found in [1]. We compute $K^1 = K^{\text{hilb.}}$ and the group structure of $\widehat{H}^{-1}(E_{K^1}) \overset{\text{def.}}{=} \widehat{H}^{-1}(\text{Gal}(K^1/K), E_{K^1})$. If $\text{Cl}(K^1)$ is not trivial, we try to go further. We compute $K^2 = (K^1)^{hilb.}$ and the group structure of $\widehat{H}^{-1}(E_{K^2}) \overset{\text{def.}}{=} \widehat{H}^{-1}(\text{Gal}(K^2/K), E_{K^2})$.

Here is the `magma` program we used:

```
clear ;
Q := RationalField() ;
dis := -84 ;
K<x> := QuadraticField(dis) ;

"Computation of K^hilb..." ;
Khilb := AbsoluteField(HilbertClassField(K)) ;
Khilb<y> := OptimizedRepresentation(Khilb) ;

"... compuation of the unit group of K^hilb..." ;
E_Khilb, e_Khilb := UnitGroup(Khilb) ;

Gal_Khilb_Q, Aut_Khilb_Q, i := AutomorphismGroup(Khilb) ;
G := FixedGroup(Khilb, K) ;
Norm_G := map < Khilb -> Khilb | y :-> &* [i(g)(y) : g in G] > ;
N := hom < E_Khilb -> E_Khilb |
   [(e_Khilb * Norm_G * Inverse(e_Khilb))(E_Khilb.i) :
       i in [1..NumberOfGenerators(E_Khilb)]] > ;

Ker_N := Kernel(N) ;
I_G := [i(g)(u)/u : u in Generators(E_Khilb) @ e_Khilb, g in G] ;
I_G := sub < E_Khilb | I_G @@ e_Khilb > ;
assert(I_G subset Ker_N) ;
printf "... structure of H^(-1)(G, E_M) = %o\n", Ker_N / I_G ;
```

Unfortunately, because of the difficulty of computing the unit group of

a number field, only few computations achieved. In the following table, the notation $2 \cdot 4$ means that the group is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$.

$\mathrm{dis}(K)$	$\mathrm{Cl}(K)$	$\mathrm{Cl}(K^1)$	$\widehat{H}^{-1}(E_{K^1})$	$\mathrm{Cl}(K^2)$	$\widehat{H}^{-1}(E_{K^2})$
-84	$2 \cdot 2$	1	$2 \cdot 2 \cdot 2$		
-120	$2 \cdot 2$	2	4	1	8
-260	$2 \cdot 4$	2	$2 \cdot 4$	1	$2 \cdot 8$
-280	$2 \cdot 2$	4	4	1	16
-308	$2 \cdot 4$	1	$2 \cdot 2 \cdot 4$		
-399	$2 \cdot 8$	1	$2 \cdot 2 \cdot 8$		
-408	$2 \cdot 2$	2	$2 \cdot 2 \cdot 2$	1	$2 \cdot 2 \cdot 4$
-420	$2 \cdot 2 \cdot 2$	$2 \cdot 2$	$2 \cdot 2 \cdot 2 \cdot 4$	1	unkown
-456	$2 \cdot 4$	1	$2 \cdot 2 \cdot 4$		

In the following section, we will explain why $d_2\widehat{H}^{-1}(E_{K^1}) = 3$ when $d_2\,\mathrm{Cl}(K) = 2$ and $d_2\,\mathrm{Cl}(K^1) = 1$. In all the remaining known cases, we point out that $d_2\widehat{H}^{-1}(E_{K^1}) = d_2\widehat{H}^{-1}(E_{K^2})$.

3. When the Galois group has two generators

The goal of this section is to extend the results of §1 to the case of extensions whose Galois group is an abelian group generated by two elements.

First, we need to investigate the cohomology group with values in M^*. We still have:

Theorem 3.1. *Let K be a number field and M/K be an unramified abelian extension whose Galois group G is a p-group generated by two elements. Then $\widehat{H}^{-1}(G, M^*) = 1$.*

Proof. Since M/K is an unramified abelian extension, there exists a subgroup G' of $\mathrm{Cl}(K)$ such that $G \simeq \mathrm{Cl}(K)/G'$. Let $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ be primes of K whose classes generate G'. If $G \simeq \mathbb{Z}/p^\alpha\mathbb{Z} \times \mathbb{Z}/p^\beta\mathbb{Z}$ with $\alpha \leq \beta$, we complete these primes by choosing $\mathfrak{p}, \mathfrak{q}$ primes of K such that their decomposition groups in M/K satisfy $D(\mathfrak{p}) = \langle(1,1)\rangle$ and $D(\mathfrak{q}) = \langle(0,1)\rangle$. Adjoining $\mathfrak{p}, \mathfrak{q}$ to the $\mathfrak{p}_i$'s leads to a system of generators of $\mathrm{Cl}(K)$.

Let $H = \langle(1,0)\rangle$. Then H and G/H are cyclic and, by construction, the decomposition groups in M/K satisfy:

$$\forall 1 \leq i \leq r, \quad D(\mathfrak{p}_i) \cap H = \{\mathrm{id}\}, \qquad D(\mathfrak{p}) \cap H = \{\mathrm{id}\}, \qquad D(\mathfrak{q}) \cap H = \{\mathrm{id}\}.$$

Theorem 3.1 is implied by the two following lemmas. □

Lemma 3.1. *Let H be a normal cyclic subgroup of G. Then:*

$$\widehat{H}^{-1}(G, M^*) = \{1\} \Longleftrightarrow \widehat{H}^{-1}(G/H, N_H(M^*)) = \{1\}.$$

Proof. Suppose that $\widehat{H}^{-1}(G, M^*) = \{1\}$. If $y \in N_H(M^*)[N_{G/H}]$, then there exists $z \in M^*$ such that $y = N_H(z)$ and $N_G(z) = N_{G/H}(N_H(z)) = N_{G/H}(y) = 1$. Thus, by hypothesis, $z \in M^*[N_G] = I_G M^*$:

$$\exists z_i \in M,\ g_i \in G, \quad z = \frac{g_1(z_1)}{z_1} \times \cdots \times \frac{g_r(z_r)}{z_r}.$$

Hence:

$$y = N_H(z) = \frac{g_1(N_H(z_1))}{N_H(z_1)} \times \cdots \times \frac{g_r(N_H(z_r))}{N_H(z_r)}.$$

Therefore $y \in I_{G/H} N_H(M^*)$.

Conversely, suppose that $\widehat{H}^{-1}(G/H, N_H(M^*)) = \{1\}$. If $z \in M^*[N_G]$ then $1 = N_G(z) = N_{G/H}(N_H(z))$ and thus $N_H(z) \in N_H(M^*)[N_{G/H}]$. By hypothesis, there exist $z_1, \ldots, z_r \in M^*$ and $g_1, \ldots g_r \in G$ such that:

$$N_H(z) = \frac{g_1(N_H(z_1))}{N_H(z_1)} \times \cdots \times \frac{g_r(N_H(z_r))}{N_H(z_r)} = N_H\left(\frac{g_1(z_1)}{z_1} \times \cdots \times \frac{g_r(z_r)}{z_r}\right).$$

It follows that:

$$z \in I_G M^* \times M^*[N_H] = I_G M^* \times I_H M^* = I_G M^*,$$

because, H being cyclic, one has $M^*[N_H] = I_H M^*$. □

Lemma 3.2. *Let H be a cyclic subgroup of G such that G/H is also cyclic. If $\mathrm{Cl}(K)$ can be generated by primes whose decomposition groups intersect H trivially, then $\widehat{H}^{-1}(G/H, N_H(M^*)) = \{1\}$.*

Proof. Let h be a generator of H and $g \in G$ such that $G = \langle g, h \rangle$. Let $L = M^H$ so that $\mathrm{Gal}(L/K) = \langle g \rangle$.

Let $y \in N_H(M^*)[N_{G/H}]$. Since G/H is cyclic generated by g, there exists $b \in L$ such that $y = \frac{g(b)}{b}$.

Since $y \in N_H(M^*)$, it is a norm everywhere locally:

$$\begin{aligned}\forall w \in \Sigma_L,\ w(y) \equiv 0 \pmod{f_w} &\Longrightarrow \forall w \in \Sigma_L,\ w \circ g(b) \equiv w(b) \pmod{f_w} \\ &\Longrightarrow \forall v \in \Sigma_K,\ \forall w, w' \in \Sigma_{L,v}, \\ &\qquad w'(b) \equiv w(b) \pmod{f_w}.\end{aligned}$$

Note that there is no condition at infinity since infinite places are unramified by assumption. The last assertion implies that the ideal J of L defined by:

$$J = \prod_{w \in \Sigma_L} \mathfrak{p}_w^{-w(b) \bmod f_w} \qquad (\text{for } x \in \mathbb{Z}, \text{ we choose } x \bmod f_w \in [0..f_w - 1]),$$

is the extension to L of the ideal I of K defined by:

$$I = \prod_{v \in \Sigma_K} \mathfrak{p}_v^{-w(b) \bmod f_w} \qquad (\text{for each } v \in \Sigma_K, \text{ we choose } w \text{ a place of } \Sigma_{L,v}).$$

By hypothesis, $\mathrm{Cl}(K)$ can be generated by prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ of K whose decomposition groups satisfy $D(\mathfrak{p}_i) \cap H = \{\mathrm{id}\}$. This means that all primes of L above each $\mathfrak{p}_i$ split totally in M. There exists $a \in K$ and $e_1, \ldots, e_r \in \mathbb{N}$ such that $\langle a \rangle = I \times \prod_i \mathfrak{p}_i^{e_i}$. By construction, the ideal ab of L has support on primes of L which split totally in M.

Now, recall that the local-global principle holds form norm equations in cyclic extensions. Thus, we deduce that $ab \in N_H(M^*)$. Finally, because $a \in K$, we have:

$$y = \frac{g(b)}{b} = \frac{g(ab)}{ab} \in I_{G/H} N_H(M^*),$$

which was to be proved. □

As in the cyclic case, the triviality of the -1 cohomological group with values in M^* implies something on the -1 cohomological group with values in E_M. To begin with, let us state the following easy proposition:

Proposition 3.1. *Let K be a number field and M/K be an unramified abelian extension with Galois group G a p-group of p-rank d. If M is principal, then $d_p \widehat{H}^{-1}(G, E_M) = \frac{d(d^2+5)}{6}$.*

Proof. In [4] §4.4, thanks to class field theory, it is proved that:

$$\forall q \in \mathbb{Z},\ \widehat{H}^{q+1}(G, E_M) \simeq \widehat{H}^{q-2}(G, \mathbb{Z}).$$

Hence, for $q = -2$, we obtain:

$$\widehat{H}^{-1}(G, E_M) \simeq \widehat{H}^{-4}(G, \mathbb{Z}).$$

By duality, it is enough to compute the p-rank of $H^4(G, \mathbb{Z})$. To this end, we start with the exact sequence of G-modules (trivial action)

$$0 \to \mathbb{Z} \xrightarrow{p} \mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} \to 0$$

and we consider the long cohomology exact sequence:

$$\begin{aligned} 0 \to H^1(G,\mathbb{Z}/p\mathbb{Z}) \to & H^2(G,\mathbb{Z}) \xrightarrow{p} H^2(G,\mathbb{Z}) \to H^2(G,\mathbb{Z}/p\mathbb{Z}) \to \\ & H^3(G,\mathbb{Z}) \xrightarrow{p} H^3(G,\mathbb{Z}) \to H^3(G,\mathbb{Z}/p\mathbb{Z}) \to \\ & H^4(G,\mathbb{Z})[p] \to 0. \end{aligned}$$

The logarithm of the product of the orders of these groups equals 0, therefore:

$$d_pH^4(G,\mathbb{Z}) = d_pH^3(G,\mathbb{Z}/p\mathbb{Z}) - d_pH^2(G,\mathbb{Z}/p\mathbb{Z}) + d_pH^1(G,\mathbb{Z}/p\mathbb{Z})$$

(recall that in a finite abelian p-group A, one has: $\#A[p] = p^{d_pA}$). It is now easy to conclude because:

$$d_pH^2(G,\mathbb{Z}/p\mathbb{Z}) = \frac{d(d+1)}{2} \qquad \text{and} \qquad d_pH^3(G,\mathbb{Z}/p\mathbb{Z}) = \frac{d(d+1)(d+2)}{6}$$

as it can be proved using Knneth's formula (see [2], exercice 7, page 96). □

Remark – The isomorphism of the beginning of this proof for $q = -1$ is a key step in the proof of Golod-Shafarevich's theorem.

Let us return to the case where $d_p(G) = 2$. Then, due to proposition 3.1, one has $d_p(G, E_M) = 3$. As in corollary 1.1, one can be more precise and exhibit a basis of $\widehat{H}^{-1}(G, E_M)$.

Proposition 3.2. *Let K be a number field and M/K an unramified abelian extension with Galois group G. If M has class number one and if $\widehat{H}^{-1}(G, M^*) = \{1\}$ then:*

$$\widehat{H}^{-1}(G, E_M) = \left\langle \left[\frac{\sigma_\pi(\pi)}{\pi}\right],\ \pi \text{ a prime element of } M \right\rangle.$$

where σ_π denotes the Frobenius at π.

Proof. Let π be a prime element of M and $g, g' \in G$ such that $g \equiv g' \bmod D(\pi)$, where $D(\pi)$ denotes the decomposition group of the ideal $\langle\pi\rangle_M$. Then there exists $\alpha \in \mathbb{N}$ such that $g^{-1}g' = \sigma_\pi^\alpha$ and thus:

$$\frac{g'(\pi)}{g(\pi)} = g\left(\frac{g^{-1}g'(\pi)}{\pi}\right) = g\left(\frac{\sigma_\pi^\alpha(\pi)}{\pi}\right) \equiv$$

$$\frac{\sigma_\pi^\alpha(\pi)}{\pi} \equiv \left(\frac{\sigma_\pi(\pi)}{\pi}\right)^\alpha \pmod{I_GE_M}.$$

For every $v \in \Sigma_K$, we choose a generator π_v of one of the primes of M above $\mathfrak{p}_v$. We fix a section $\sigma \mapsto \widetilde{\sigma}$ of the cononical projection map $G \to$

$G/D(\pi_v)$. The elements $\widetilde{\sigma}(\pi_v)$, when v runs in Σ_K and $\sigma \in G/D(v)$, describe a system of prime elements of M. Then every $z \in M$ factorizes into:

$$z = u \prod_{v \in \Sigma_K} \left(\prod_{\sigma \in G/D(v)} \widetilde{\sigma}(\pi_v)^{e_{v,\sigma}} \right) \qquad \Longrightarrow$$

$$g(z) = g(u) \prod_{v \in \Sigma_K} \left(\prod_{\sigma \in G/D(v)} g\widetilde{\sigma}(\pi_v)^{e_{v,\sigma}} \right)$$

for every $g \in G$. Of course $g\widetilde{\sigma} \equiv \widetilde{g\sigma} \bmod D(\pi_v)$, therefore there exists $\alpha_{v,\sigma} \in \mathbb{N}$ such that:

$$g\widetilde{\sigma}(\pi_v) = \left(\frac{\sigma_v(\pi_v)}{\pi_v} \right)^{\alpha_{v,\sigma}} \widetilde{g\sigma}(\pi_v) \Longrightarrow$$

$$g(z) \in \langle g(u) \rangle \left\langle \frac{\sigma_\pi(\pi)}{\pi}, \pi \text{ prime of } M \right\rangle \langle \widetilde{\sigma}(\pi_v), v \in \Sigma_K, \sigma \in G/D(v) \rangle \,.$$

Now start with $u \in E_M[N_G]$. By hypothesis, we know that $\widehat{H}^{-1}(G, M^*) = \{1\}$, i.e. $M^*[N_G] = I_G M^*$. Hence, if $G = \langle g_1, \ldots, g_r \rangle$, there exists $z_1, \ldots, z_r \in M^*$ such that $u = \frac{g_1(z_1)}{z_1} \cdots \frac{g_r(z_r)}{z_r}$. Factorizing $z_1, \ldots, z_r$ into primes of M of the form $\widetilde{\sigma}(\pi_v)$, one shows that:

$$u \in I_G E_M \left\langle \frac{\sigma_\pi(\pi)}{\pi},\ \pi \text{ a prime element of } M \right\rangle \langle \widetilde{\sigma}(\pi_v),\ v \in \Sigma_K, \sigma \in G/D(v) \rangle \,;$$

But, in this decomposition, since u is invertible, the element in the third group must be equal to 1. □

Theorem 3.2. *Let K be a number field whose ideal class group is a p-group of rank two and M/K its Hilbert class field. Suppose that M has class number one. Then for any generators g_1, g_2 of $\mathrm{Gal}(M/K)$ and any generators π_1, π_2, π_{12} of prime ideals of M with Frobenius equal to g_1, g_2 and $g_1 g_2$ respectively, $\widehat{H}^{-1}(G, E_M)$ is generated by the classes of $g_1(\pi_1)/\pi_1$, $g_1(\pi_2)/\pi_2$ and $g_1 g_2(\pi_{12})/\pi_{12}$:*

$$\widehat{H}^{-1}(G, E_M) = \left\langle \left[\frac{g_1(\pi_1)}{\pi_1} \right], \left[\frac{g_2(\pi_2)}{\pi_2} \right], \left[\frac{g_1 g_2(\pi_{12})}{\pi_{12}} \right] \right\rangle .$$

Proof. For any prime element π of M, we denote its Frobenius by σ_π. By theorem 3.1, we have $\widehat{H}^{-1}(G, M^*) = \{1\}$ and thanks to the preceding result the group $\widehat{H}^{-1}(G, E_M)$ is generated by the classes of the elements $\frac{\sigma_\pi(\pi)}{\pi}$.

Therefore, we only have to prove that the class modulo $I_G E_M$ of the element $u = \frac{\sigma_\pi(\pi)}{\pi}$ is contained in the subgroup generated by the $\frac{g_i(\pi_i)}{\pi_i}$ for $i = 1, 2, 12$.

To this end, put $H = \langle g_{12} \rangle$, $L = M^H$ and $\mathfrak{p} = \langle \pi \rangle_M \cap K$, $\mathfrak{p}_1 = \langle \pi_1 \rangle_M \cap K$, $\mathfrak{p}_2 = \langle \pi_2 \rangle_M \cap K$.

There exits $\alpha_1, \alpha_2 \in \mathbb{N}$ such that $\sigma_\pi = g_1^{\alpha_1} g_2^{\alpha_2}$ and, by Artin map, $\mathfrak{p} = a\mathfrak{p}_1^{\alpha_1}\mathfrak{p}_2^{\alpha_2}$ with $a \in K^*$. Since $\langle \sigma_i \rangle \cap H = \{\mathrm{Id}\}$ for $i = 1, 2$, the primes $\mathfrak{p}_i$, $i = 1, 2$, totally split between L and M. Thus:

$$\begin{cases} j_{L/K}(\mathfrak{p}) = \langle N_H(\pi) \rangle_L \\ j_{L/K}(\mathfrak{p}_\mathrm{i}) = \langle N_H(\pi_i) \rangle_L \,,\ i = 1, 2 \end{cases} \quad \Longrightarrow N_H(\pi) = avN_H(\pi_1)^{\alpha_1} N_H(\pi_2)^{\alpha_2},$$

where $v \in E_L$. Hence:

$$N_H(u) = N_H\left(\frac{\sigma_\pi(\pi)}{\pi}\right) = \frac{\sigma_\pi\left(N_H(\pi)\right)}{N_H(\pi)} =$$

$$\frac{\sigma_\pi(a)}{a}\frac{\sigma_\pi(v)}{v} N_H\left(\frac{\sigma_\pi(\pi_1)}{\pi_1}\right)^{\alpha_1} N_H\left(\frac{\sigma_\pi(\pi_2)}{\pi_2}\right)^{\alpha_2}.$$

Let us study the four terms in the right hand product. The first one is equal to 1 because $a \in K$. Since local-global principal occurs in cyclic extensions and since M/L is unramified, there exists $w \in E_M$ such that $v = N_H(w)$. Thus the second term $\frac{\sigma_\pi(v)}{v}$ equals $N_H\left(\frac{\sigma_\pi(w)}{w}\right)$. The thirst and fourth terms go in the same way: since g_1, g_2 generate G, the elements g_1 and $g_1 g_2$ also generate G and there exists $\beta_1, \beta_2 \in \mathbb{N}$ such that $\sigma_\pi = g_1^{\beta_1}(g_1 g_2)^{\beta_2}$. It follows that:

$$N_H\left(\frac{\sigma_\pi(\pi_1)}{\pi_1}\right) = N_H\left(\frac{g_1^{\beta_1}(\pi_1)}{\pi_1}\right) = N_H\left(\frac{g_1(w_1)}{w_1}\left(\frac{g_1(\pi_1)}{\pi_1}\right)^{\beta_1}\right)$$

where $w_1 \in E_M$.

In conclusion, u satisfies:

$$N_H(u) = N_H\left(\frac{\sigma_\pi(w)}{w}\frac{g_1(w_1)}{w_1}^{\alpha_1}\frac{g_2(w_2)}{w_2}^{\alpha_1}\left(\frac{g_1(\pi_1)}{\pi_1}\right)^{\alpha_1\beta_1}\left(\frac{g_2(\pi_2)}{\pi_2}\right)^{\alpha_2\beta_2}\right)$$

hence

$$u \times \left(\frac{\sigma_\pi(w)}{w}\frac{g_1(w_1)}{w_1}^{\alpha_1}\frac{g_2(w_2)}{w_2}^{\alpha_2}\left(\frac{g_1(\pi_1)}{\pi_1}\right)^{\alpha_1\beta_1}\left(\frac{g_2(\pi_2)}{\pi_2}\right)^{\alpha_2\beta_2}\right)^{-1} \in E_M[N_H].$$

Finally, due to the cyclic case, we know that $E_M[N_H] = I_H E_M \left\langle \frac{g_1 g_2(\pi_{12})}{\pi_{12}} \right\rangle$ and thus:

$$u \bmod I_G E_M \in \left\langle \frac{g_1(\pi_1)}{\pi_1}, \frac{g_2(\pi_2)}{\pi_2}, \frac{g_1 g_2(\pi_{12})}{\pi_{12}} \right\rangle,$$

which was to be proved. □

Remark – All these results hold in the function field case for S-units where S is any non-empty finite set of places.

References

1. Franz Lemmermeyer. A survey on class field towers. `http://www.rzuser.uni.heidelberg.de/~hb3/cft.html`.
2. Jurgen Neukirch, Alexander Schmidt, and Kay Wingberg. *Cohomology of Number Fields*, volume 323 of *A Series of Comprehensive Studies in Mathematics*. Springer, 2000.
3. Jean-Pierre Serre. *Corps locaux*. Hermann, troisième edition, 1968.
4. Jean-Pierre Serre. *Galois Cohomology*. Springer, 1994.

On the semiprimitivity of cyclic codes

Yves Aubry

Institut de Mathématiques de Toulon
Université du Sud Toulon-Var, France
E-mail: yaubry@univ-tln.fr

Philippe Langevin

Institut de Mathématiques de Toulon
Université du Sud Toulon-Var, France
E-mail: langevin@univ-tln.fr

We prove, without assuming the Generalized Riemann Hypothesis, but with at most one exception, that an irreducible cyclic code $c(p, m, v)$ with v prime and p of index 2 modulo v is a two-weight code if and only if it is a semiprimitive code or it is one of the six sporadic known codes. The result is proved without any exception for index-two irreducible cyclic $c(p, m, v)$ codes with v prime not congruent to 3 modulo 8. Finally, we prove that these two results hold true in fact for irreducible cyclic code $c(p, m, v)$ such that there is three p-cyclotomic cosets modulo v.

1. Introduction

Irreducible cyclic codes are extensively studied in the literature. They can be defined by three parameters p, m and v and are denoted $c(p, m, v)$ (see section 2 for a precise definition). Such codes with only few different (Hamming) weights are highly appreciated, especially those with exactly two non-zero weights, called two-weight codes. The classification of two-weight codes is a classical problem in coding theory (see [3]); it is still an open problem but recent progress has been made. An infinite family, namely the semiprimitive codes (i.e. when -1 is a power of p modulo v), and eleven sporadic examples are known. Schmidt and White in [9] provided evidence to conjecture that this is the whole story:

Conjecture 1.1. *An irreducible cyclic code $c(p, m, v)$ is a two-weight code if and only if it is a semiprimitive code or it is one of the eleven sporadic known codes.*

They proved their conjecture, conditional on the Generalized Riemann Hypothesis (G.R.H.), for index-two codes, that is when p has index 2 modulo v. Note that semiprimitive codes have two non-zero weights and thus only the "only if" part had to be proved.

We considered in [1] the conjecture in the binary case and we proved it in a particular case without assuming G.R.H.. Our main result here is a proof of this conjecture without assuming G.R.H. but with at most one exception in the case where p has index 2 and v is prime. We prove before, using near-primitive root densities and conditionally on G.R.H., that for any prime number p there are infinitely many such codes namely index-two irreducible cyclic codes $c(p, m, v)$ with v prime.

We prove the conjecture without any exception (and without assuming G.R.H.) in the case where p has index 2 and v is a prime not congruent to 3 modulo 8. Finally, we remark that the results hold true in fact for irreducible cyclic codes $c(p, m, v)$ with v an integer such that there is three p-cyclotomic cosets modulo v.

2. Irreducible cyclic codes and McEliece weight-formula

Let us introduce irreducible cyclic codes over a prime finite field (indeed, it is enough for our purpose, namely the classification of two-weight irreducible cyclic codes, to consider such codes over prime fields, as remarked in [9]).

Let p be a prime number and consider the finite field K with p elements. Let L be the extension of degree m of K, consider a divisor n of $p^m - 1$ and write $v = (p^m - 1)/n$ (thus v and p are coprime). Let ζ be a primitive n-th root of unity in L (i.e. ζ is a generator of the cyclic subgroup of order n of the multiplicative group L^*). We define the $c(p, m, v)$ code to be the image of the following map Φ_m:

$$\begin{aligned} \Phi_m \colon L &\longrightarrow K^n \\ t &\longmapsto \left(\mathrm{Tr}_{L/K}(t\zeta^{-i})\right)_{i=0}^{n-1} \end{aligned}$$

where $\mathrm{Tr}_{L/K}$ is the trace of the field L over K.

It is a code of length n and dimension $\mathrm{ord}_n(p)$, the multiplicative order of p modulo n. Every irreducible cyclic code over K can be viewed as a $c(p, m, v)$ code (see [9]), so we can take $c(p, m, v)$ as the definition of irreducible cyclic codes over K of length n. The $c(p, m, v)$ codes are known to be projective or saturated according to whether $\gcd(n, p-1) = 1$ or $\gcd(n, p-1) = p-1$. As remarked in [9], we may assume the saturated situation.

Now we are interested in the weight $w(t)$ of a codeword $\Phi_m(t)$ of such a code, for $t \in L^*$. Let χ be a character of the multiplicative group L^* and let

$$\tau_L(\chi) = -\sum_{x \in L^*} \chi(x) e^{\frac{2i\pi}{p} \mathrm{Tr}_{L/K}(x)} \tag{1}$$

be the Gauss sum associated with χ.

Let V be the subgroup of L^* of index v and let Γ be the subgroup of characters of L^* which are trivial both on V and K^*. Note that the order of Γ is equal to $v \gcd(n, p-1)/(p-1)$ which is just equal to v in the saturated situation. We have the following McEliece formula:

Proposition 2.1. *For any $t \in L^*$, the weight $w(t)$ of the codeword $\Phi_m(t)$ is given by:*

$$w(t) = \frac{p-1}{pv}\big(p^m + \sum_{\chi \in \Gamma \setminus \{1\}} \tau_L(\chi)\bar{\chi}(t)\big). \tag{2}$$

And, conversely by Fourier inversion

$$\tau_L(\chi) = \frac{p}{p-1} \sum_{t \in L^*/V} w(t)\chi(t). \tag{3}$$

One says that p is semiprimitive modulo v when -1 is in the group generated by p in $(\mathbb{Z}/v\mathbb{Z})^*$, i.e. when $\mathrm{ord}_v(p)$ is even. Note that in this case all the Gauss sums are rational and a $c(p, m, v)$ code is a two-weight code. In the paper we investigate the reciprocal: besides some sporadic known examples, is any two-weight irreducible cyclic code semiprimitive ?

3. The case v small

Before going further let us treat the case where v is small, i.e. $v = 2$ or 3. We know that a $c(p, m, 2)$ code is a two-weight code, and that the weights can be expressed in term of quadratic Gauss sum (see [7]). In the same way, the weights of a $c(p, m, 3)$ code can be expressed by means of cubic Gauss sums. However, it is hard to give the exact values of the cubic Gauss sums (see [6]), and thus also the weights of such a code. Nevertheless, we have the following characterization:

Proposition 3.1. *A $c(p, m, 3)$ code has two weights if and only if it is semiprimitive (that is here, if and only if $p \equiv 2 \mod 3$).*

Proof. Let χ be a multiplicative character of L of order 3. The number of weights of a $c(p, m, 3)$ code is equal to the number of distincts values taken by the mapping:

$$L^* \ni t \mapsto f(t) = \tau_L(\chi)\chi(t) + \tau_L(\bar{\chi})\bar{\chi}(t).$$

Let $1 \neq j$ be a cubic root of unity. Let t be such that $\chi(t) = j$. It is easy to see that $f(1) = f(t)$ implies $\tau_L(\bar{\chi}) = j\tau_L(\chi)$, that $f(t) = f(t^2)$ implies $\tau_L(\bar{\chi}) = \tau_L(\chi)$ and that $f(1) = f(t^2)$ implies $\tau_L(\bar{\chi}) = j^2\tau_L(\chi)$. Therefore, the code has two weights if and only if there exists a cubic root of unity ω such that

$$\tau_L(\bar{\chi}) = \omega\tau_L(\chi). \tag{4}$$

In particular, since $\tau_L(\chi)^3$ is an algebraic integer of degree 2 and norm p^{3m}, we deduce that $\tau_L(\chi)^6 = \tau_L(\bar{\chi})^6 = p^{6m}$. Hence the Gauss sums $\tau_L(\chi)$ are pure Gauss sums (see [7] for a definition of a pure Gauss sum). It follows by a theorem of Baumert, Mills and Ward (see Theorem 11.6.4 of [7] for example) that p is semiprimitive modulo 3. □

4. Infinitely many index-two $c(p, m, v)$ codes with v prime

For the study of $c(p, m, v)$ codes with v prime and p of index two modulo v, we are interested in primitive and near-primitive root densities.

In 1927, Emil Artin made the following conjecture (called now the Artin's primitive root conjecture): for any integer $\alpha \neq \pm 1$ not a square, the natural density

$$\lim_{x \to +\infty} \frac{\sharp\{v \text{ prime} \mid v \leq x \text{ and } \alpha \text{ generates } \mathbf{F}_v^*\}}{\sharp\{v \text{ prime} \mid v \leq x\}}$$

exists and is positive. In 1967, Hooley proved this conjecture under the assumption of G.R.H.. In particular, he proved that if α is neither ± 1 nor a perfect square, then there are infinitely many primes v for which α is a primitive root modulo v.

If we ask α to generate only the squares of $\mathbf{F}_v^*$ and not the whole group $\mathbf{F}_v^*$, i.e. to have index 2 and not index 1 modulo v, we come to the notion of near-primitive roots. Precisely, fix $\alpha \neq \pm 1$ not a perfect power and let v be a prime and t be an integer such that $v \equiv 1 \pmod{t}$. Consider

$$N_{\alpha,t}(x) = \sharp\{v \text{ prime} \mid v \leq x \text{ and } v \not| \alpha \text{ and } \operatorname{ind}_v(\alpha) = t\}.$$

Notice that for $t = 1$ this quantity is just the previous one studied by Artin and Hooley. In 2000, Moree introduced in [8] a weighting function

depending on α and t and gave an estimation of $N_{\alpha,t}(x)$ assuming G.R.H.. In particular, for $\alpha = p$ a prime number and $t = 2$, he proved that

$$N_{p,2}(x) = \sum_{\substack{v \text{ odd prime} \\ v \leq x}} \frac{\varphi(\frac{v-1}{2})}{v-1} + O(\frac{x \log\log x}{\log^2 x}).$$

This implies that there exist infinitely many primes v such that p has index 2 modulo v.

In particular, we have:

Proposition 4.1. *Conditionally on G.R.H., for any prime number p there are infinitely many index-two irreducible cyclic codes $c(p, m, v)$ with v prime.*

5. Necessary conditions on two-weight codes

The irreducible cyclic codes $c(p, m, v)$, with v a prime number and with p of index 2 modulo v, range in two families: the first one with $v \equiv 1 \pmod 4$ and the second one with $v \equiv 3 \pmod 4$. If $v \equiv 1 \pmod 4$, then -1 is a square modulo v and since p generates the squares modulo v, we are reduced to the semiprimitive case. This lead us to consider the second case, where -1 is not a square modulo v. Moreover, in view of Proposition 3.1, we can suppose that v is greater than 3.

Hence, let us consider a prime number p and an integer v satisfying the following ($\sharp$) conditions:

(a) v is a prime greater than 3,
(b) $\mathrm{ord}_v(p) = (v-1)/2$ i.e. p has index 2 modulo v,
(c) $v \equiv 3 \pmod 4$.

Let f be the multiplicative order of p modulo v. Note that f divides m, and we set $s = m/f$. It is shown in [4] that if a $c(p, m, v)$ code with v satisfying the ($\sharp$) conditions has two weights then:

$$\frac{v+1}{4} = p^{hs}. \tag{5}$$

We give, now, a more precise result:

Theorem 5.1. *If a $c(p, m, v)$ code with v satisfying the ($\sharp$) conditions is a two-weight code then we have:*

$$m = \mathrm{ord}_v(p).$$

Proof. Since p has index 2 modulo v, then p is a square modulo v, and

$$(p) = PP'$$

splits in the extension $\mathbf{Q}(\sqrt{-v})/\mathbf{Q}$. We have that the norm

$$N_{\mathbf{Q}(\sqrt{-v})/\mathbf{Q}}(P) = p$$

and that $P^h = (\alpha)$ is a principal ideal (since h is the ideal class number of $\mathbf{Q}(\sqrt{-v})$), with $\alpha = (a + b\sqrt{-v})/2$ (with $a, b \in \mathbf{Z}$) an algebraic integer of $\mathbf{Q}(\sqrt{-v})$. Taking norms, we obtain $p^h = (a^2 + vb^2)/4$ and since a and b cannot be zero in this situation, we conclude that

$$\frac{v+1}{4} \leq p^h.$$

But by (5) a $c(p, m, v)$ code with v satisfying the ($\sharp$) conditions has two weights if and only if

$$\frac{v+1}{4} = p^{hs}. \tag{6}$$

Thus, $p^{hs} \leq p^h$ and $s = 1$. □

Then, the previously defined parameter s appearing in [4] and [9] is equal to 1 under the ($\sharp$) conditions. In particular, we have:

Corollary 5.1. *If a $c(p, m, v)$ code with v satisfying the ($\sharp$) conditions is a two-weight code then*

$$\frac{v+1}{4} = p^h. \tag{7}$$

where h is the class number of the imaginary quadratic number field $\mathbf{Q}(\sqrt{-v})$. In particular, such a code is completely defined by the parameter v.

Furthermore, we have the following necessary condition on p for two-weight $c(p, m, v)$ code with v satisfying the ($\sharp$) conditions:

Corollary 5.2. *If a $c(p, m, v)$ code with v satisfying the ($\sharp$) conditions has two weights, then p is the least prime which totally splits in the extension $\mathbf{Q}(\sqrt{-v})/\mathbf{Q}$, i.e. p is the least prime which is a square modulo v.*

Proof. Indeed, if ℓ is a prime which totaly splits in $\mathbf{Q}(\sqrt{-v})/\mathbf{Q}$, then the previous proof implies that $\ell^h \geq \frac{v+1}{4} = p^h$ which gives $\ell \geq p$. □

6. Main results

Using the previous section, we can state the following result which can also be derived from the proof of lemma 4.1. of [4].

Theorem 6.1. *There is no two-weight $c(p, m, v)$ code with v satisfying the* ($\sharp$) *conditions and with $v \equiv 7 \pmod 8$. Hence, Conjecture 1 holds true for index-two irreducible cyclic $c(p, m, v)$ codes with v a prime not congruent to 3 modulo 8.*

Proof. Since $v \equiv 7 \pmod 8$, it follows that 2 is a square modulo v, and the ideal (2) splits in the extension $\mathbf{Q}(\sqrt{-v})/\mathbf{Q}$. By Corollary 5.2, we conclude that $p = 2$. But we proved in [1] that there exists no two-weight binary $c(p, m, v)$ code with v satisfying the ($\sharp$) conditions, so we get the non-existence of such codes. Hence, this proves the conjecture since the case $v \equiv 1 \pmod 4$ is trivial, as quoted in the previous section, and the last case $v \equiv 3 \pmod 4$ is divided in two subcases : when $v \equiv 7 \pmod 8$, which is now solved, and when $v \equiv 3 \pmod 8$ which is the remainder case. □

Actually, we will consider now a more general approach using the identity of Corollary 5.1 but with at most one exception.

If a $c(p, m, v)$ code with v satisfying the ($\sharp$) conditions has two weights then we have the following relation

$$\frac{v+1}{4} = p^h,$$

where h is the class number of the imaginary quadratic number field $\mathbf{Q}(\sqrt{-v})$ (see Corollary 5.1).

In 1935, Siegel gave a non-effective lower bound on the residue at $s = 1$ of the L-function $L(s, \chi_v)$ associated to the primitive odd Dirichlet character χ_v of $\mathbf{Q}(\sqrt{-v})$. Tatuzawa, in 1951, proved an effective lower bound of $L(1, \chi_v)$ but with at most one exception (see [10] and see also [5] for a simple proof): if $0 < \varepsilon < 1/2$ and $v \geq \max(e^{1/\varepsilon}, e^{11.2})$, then

$$L(1, \chi_v) \geq 0.655\varepsilon v^{-\varepsilon}.$$

Since the class number h of $\mathbf{Q}(\sqrt{-v})$ with $-v \equiv 1 \pmod 4$ is linked to $L(1, \chi_v)$ by the following Dirichlet class number formula:

$$L(1, \chi_v) = \frac{\pi h}{\sqrt{v}},$$

we can use Tatuzawa theorem to get an upper bound on v.

Proposition 6.1. *There exists at most one two-weight $c(p, m, v)$ code with $v \geq 10^8$ satisfying the ($\sharp$) conditions.*

Proof. Let $\varepsilon = 1/\log(10^8) \in (0, 1/2)$. For $v \geq \max(e^{1/\varepsilon}, e^{11.2}) = 10^8$, we have, with at most one exception:

$$L(1, \chi_v) \geq 0.655\varepsilon v^{-\varepsilon} = 0.035v^{-0.054}.$$

Now, $\frac{v+1}{4} = p^h \geq 2^h$ implies that $\log \frac{v+1}{4} \geq h \log 2$. By the Dirichlet class number formula, we get:

$$\log \frac{v+1}{4} \geq \frac{\sqrt{v}L(1, \chi_v)}{\pi} \log 2.$$

But, for $v \geq 10^8$, we have on one hand $\log \frac{v+1}{4} \geq 17.03$ and on the other hand $\frac{\sqrt{v}L(1,\chi_v)}{\pi} \log 2 > 28.55$ by Tatuzawa theorem. Thus, there exists no $v \geq 10^8$ such that $\frac{v+1}{4} = p^h$, with at most one exception. □

Now, we make an exhaustive research of the primes $v \leq 10^8$ such that $(v+1)/4$ is a power of a prime p. Then, for such primes v, we check whether $(v+1)/4 = p^{h(v)}$ holds true or not, with $h(v)$ the class number of $\mathbf{Q}(\sqrt{-v})$. Actually, we recover the following sporadic known examples of Table 1.

Table 1. Sporadic $c(p, m, v)$ codes with v satisfying the ($\sharp$) conditions and $v \leq 10^8$.

v	11	19	67	107	163	499
p	3	5	17	3	41	5
h	1	1	1	3	1	3

Thus, we have proved the following theorem:

Theorem 6.2. *Any two-weight irreducible cyclic $c(p, m, v)$ code where p has index two modulo a prime v and which is not one of the six sporadic examples of Table 1 is semiprimitive, with at most one exception. Hence, Conjecture 1 is true, with at most one exception, for all index-two $c(p, m, v)$ codes with v prime.*

7. Cyclotomic cosets

Let p be a prime. For any integer v prime to p, consider on the ring $\mathbf{Z}/v\mathbf{Z}$ the equivalence relation given by: for $a, b \in \mathbf{Z}/v\mathbf{Z}$, we set $a \sim b$ if and only if there exists $t \in \mathbf{Z}$ such that $a = bp^t$. The equivalence classes for this equivalence relation are the so-called p-cyclotomic cosets modulo v.

Recall that the order $\mathrm{ord}_v(g)$ of an element g of the multiplicative group $(\mathbf{Z}/v\mathbf{Z})^*$ divides the order $\varphi(v)$ of this group, where φ is the Euler function. We denote by $\mathrm{ind}_v(g)$ the index of g modulo v i.e.

$$\mathrm{ind}_v(g) = \frac{\varphi(v)}{\mathrm{ord}_v(g)}.$$

Then $\mathrm{ind}_v(g) = [(\mathbf{Z}/v\mathbf{Z})^* : \langle g\rangle]$ where $\langle g\rangle$ denotes the subgroup of $(\mathbf{Z}/v\mathbf{Z})^*$ generated by g. But the number $\gamma(p,v)$ of p-cyclotomic cosets modulo v is also equal to the number of irreducibles polynomials in the decomposition of the polynomial $X^v - 1$ over $\mathbf{F}_p$, thus it is equal to

$$\gamma(p,v) = \sum_{d|v} \frac{\varphi(d)}{\mathrm{ord}_d(p)} = \sum_{d|v} \mathrm{ind}_d(p) \tag{8}$$

with the convention that $\mathrm{ind}_1(p) = 1$. For example, the condition $\gamma(p,v) = 2$ is equivalent to $\mathrm{ind}_v(p) = 1$, that is p is a primitive root modulo v.

Proposition 7.1. *Let v be an integer. The ring $\mathbf{Z}/v\mathbf{Z}$ contains exactly 3 p-cyclotomic cosets if and only if one of the following holds:*

(i) v is a prime and p has index 2 mod v;
(ii) v is the square of a prime and p has index 1 mod v.

Proof. By (8) we have $\gamma(p,v) = 3$ if and only if $\mathrm{ind}_v(p) = 2$ and v has no proper divisor, or $\mathrm{ind}_v(p) = 1$ and v has a unique proper divisor. The proposition is then proved. □

Proposition 7.2. *Let v be an integer. If the ring $\mathbf{Z}/v\mathbf{Z}$ contains exactly three p-cyclotomic cosets then any $c(p,m,v)$ code has at most three non-zero weights.*

Proof. The result is in fact much general: the number of weights is less or equal than the number of cyclotomic cosets. It follows from the fact that the weight of a codeword of a $c(p,m,v)$ code is invariant under $t \mapsto t\zeta$ and under $t \mapsto t^p$; see Theorem 2.5 of [2] for a detailed proof. □

The case (ii) of Proposition 7.1 falls into the semiprimitive case since p generates the whole group $(\mathbf{Z}/v\mathbf{Z})^*$ and thus contains -1.

Finally, we have proved the following result:

Theorem 7.1. *If v is an integer such that there is three p-cyclotomic cosets modulo v then any two-weight irreducible cyclic code $c(p,m,v)$ which is not*

one of the six sporadic examples of Table 1 is semiprimitive, with at most one exception. Hence, Conjecture 1 holds true, with at most one exception, for all $c(p, m, v)$ codes with v an integer such that there is three p-cyclotomic cosets modulo v.

Proof. If a binary irreducible cyclic code with three-cyclotomic cosets has two weights then it is semiprimitive. Indeed, by Proposition 7.1, an irreducible cyclic code with three-cyclotomic cosets leads to two cases. The first one leads $c(p, m, v)$ codes with v a square of a prime and p of index 1 modulo v which gives a semiprimitive code.

The other case leads to $c(p, m, v)$ codes with v a prime and p of index 2 modulo v (the so-called index-two codes). When $v \equiv 1 \pmod 4$, we saw that we obtain a semiprimitive code. When $v \equiv 3 \pmod 4$, we obtain $c(p, m, v)$ codes with v satisfying the ($\sharp$) conditions. In the case where $p = 2$, i.e. the binary case, we found in [1] that there is no two-weight codes. When $p \neq 2$, theorem 6.2 gives the result. □

References

1. Y. Aubry and P. Langevin, On the weights of binary irreducible cyclic codes, *Coding and Cryptography*, Springer Lecture Notes in Computer Science, vol. **3969**, 46-54 (2006).
2. R. W. Fitzgerald and J. L. Yucas, Sums of Gauss sums and weights of irreducible codes, *Finite Field and Their Applications* **11** (2005), 89-110.
3. R. Calderbank, W. M. Kantor, The geometry of two-weight codes, Technical report, Bell Laboratory, 1978.
4. Ph. Langevin, A new class of two weight codes, Finite fields and their applications, Glasgow 1995, *London Math. Soc.* Lecture Note Ser. **233**, 181-187 (1996).
5. S. Louboutin, Simple proofs of the Siegel-Tatuzawa and Brauer-Siegel theorems, *Colloq. Math.* **108** (2007), no. 2, 277-283.
6. C. R. Matthews, Gauss sums and elliptic functions I. The Kummer sums, *Invent. Math.*, **52** (1979), 163-185.
7. B. C. Berndt, R. J. Evans and K. S. Williams, Gauss and Jacobi Sums, Wiley-Interscience, N. Y., 1998.
8. P. Moree, Asymptotically exact heuristics for (near) primitive roots, *J. Number Theory* **83** (2000), no. 1, 155-181.
9. B. Schmidt and C. White, All two-weight irreducible cyclic codes ?, *Finite Fields and Their Applications* **8** (2002), 1-17.
10. T. Tatuzawa, On a theorem of Siegel, *Jap. J. Math.* **21** (1951), 163-178 (1952).

Decoding of scroll codes

George H. Hitching

Høgskolen i Vestfold
Boks 2243
N-3103 Tønsberg
Norway
E-mail: george.h.hitching@hive.no

Trygve Johnsen

Dept. of Mathematics
University of Bergen
Johs. Brunsgt 12
N-5008 Bergen
Norway
E-mail: johnsen@math.uib.no

We define and study a class of codes obtained from scrolls over curves of any genus over finite fields. These codes generalize Goppa codes in a natural way, and the orthogonal complements of these codes belong to the same class. We show how syndromes of error vectors correspond to certain vector bundle extensions, and how decoding is associated to finding destabilizing bundles.

Keywords: curves, scrolls, principal parts, linear codes, decoding, vector bundle extensions

1. Introduction

For linear error-correcting codes over a finite field $\mathbb{F}_q$ one is always interested in the process of decoding, that is: determining which codeword was sent, even if some small error was made during transmission. The process of decoding may admit different mathematical interpretations, depending on how the codes in question were constructed. One way to produce such codes is to pick a geometric object X in the projective space $\mathbf{P}^{K-1}$, and let some or all of the points of X be represented by elements of $\mathbb{F}_q^K$. If one uses these K-tuples as the columns of a generator matrix, one defines a code, whose minimum distance d is the maximal number of points of X in a hyperplane in $\mathbf{P}^{K-1}$. The choice of representative for each point and the ordering of

the points do not affect the equivalence class of the code, and hence not the parameters either. Another, dual, way is to apply the same n points as the columns of a parity check matrix for a (dual) code. Then the minimum distance is the smallest number d^* of columns that are dependent (spanning only a $\mathbf{P}^{d^*-2}$ in $\mathbf{P}^{K-1}$).

The geometric object X most commonly used is a non-singular algebraic curve, and the equivalence class of the code thus obtained is that of a Goppa code. See for example [14], page 55.

It can also be interesting to define and investigate codes from higher-dimensional geometric objects. These usually contain more points than curves, so one can produce longer codes. Examples are algebraic surfaces, Grassmann varieties, m-fold embeddings of projective spaces in bigger spaces, and products of such embedded varieties. The objects we will study in the present article are scrolls over curves or, more abstractly, projective bundles $\mathbf{P}E$, where E is a rank r vector bundle over some complete algebraic curve X of genus g. This is a fibration over X with fibres isomorphic to $\mathbf{P}^{r-1}$, and $\mathbf{P}E$ is mapped into $\mathbf{P}^{K-1}$ in such a way that the fibres become linear subspaces. The resulting image T of $\mathbf{P}E$ is called a scroll.

In this paper we will not pick all points of each fibre from T to produce codes (as for example in [4], [10], and [11]). Instead we will pick exactly r independent points from each fibre, and use representatives of these points as columns of a generator matrix. The special case $r = 1$ corresponds to traditional Goppa codes.

In [7], the second author gave a geometric interpretation of decoding for the case $r = 1$, in terms of vector bundle extensions. In this paper we give a generalization of this interpretation for arbitrary natural numbers r. The key point is that the syndromes of transmitted messages correspond to the cohomology classes of vector bundle extensions

$$0 \to O_X \to W \to H \to 0$$

over X, where H is a certain vector bundle of rank r, and an essential part of error location corresponds to finding a certain quotient bundle of W.

This viewpoint has been utilized and studied for $r = 1$ through a series of papers [1], [2], [3], so we believe it may have some interest also for higher r.

Here is a summary of the article. In Section 2 we recall some facts about scrolls and vector bundles which will be needed. In §3, we define "SAGS codes", a type of evaluation code which generalizes Goppa's SAG codes

to scrolls. In particular, these have the property that their dual codes can again be interpreted as evaluation codes. In §4, we recall or prove some facts about the geometry of vector bundle extensions, and in §5 we apply this to decoding and error correction on SAGS codes. In the final section, we make brief remarks about the applicability of these results to scroll codes which are not necessarily evaluation codes.

An important tool is the use of bundle-valued principal parts to define the codes. We believe this makes transparent the connection between syndromes, bundles and geometry.

Acknowledgements: The first author is grateful to the Deutsche Forschungsgemeinschaft Schwerpunktprogramm "Globale Methoden in der komplexen Geometrie" and Leibniz Universität Hannover for support during the writing of this paper. He also thanks the University of Bergen for financial support and hospitality.

2. Scrolls and vector bundles

In this section we introduce the objects with which we will be working. Firstly, we fix some notation.

We will work over the algebraic closure of a finite field $\mathbb{F}_q$. Later we will specialize to $\mathbb{F}_q$ when suitable. We denote vector bundles over the curve X with Roman letters E, W, O_X, K_X etc., and their sheaves of sections with the corresponding script letters $\mathcal{E}$, $\mathcal{W}$, $\mathcal{O}_X$, $\mathcal{K}_X$ etc. If V is a vector space, then $\mathbf{P}V$ is the projective space of codimension one linear subspaces in V. Similarly, for a vector bundle $E \to X$ we define $\mathbf{P}E$ to be the bundle whose fiber at $x \in X$ is the projective space of codimension one linear subspaces of $E|_x$. Given a line bundle Υ over a variety Y, we write $|\Upsilon|$ for the projective space $\mathbf{P}H^0(Y, \Upsilon)$. If $|\Upsilon|$ is nonempty, we have a natural map $Y \dashrightarrow |\Upsilon|$.

If V is a vector space and $g \in V^*$ a nonzero linear form, we denote $\langle g \rangle$ the line in V^* spanned by g, and also the point in $\mathbf{P}V$ defined by g. We use the same notation for points of projectivized vector bundles.

Any vector bundle $E \to X$ gives rise to a short exact sequence of $\mathcal{O}_X$-modules

$$0 \to \mathcal{E} \to \underline{\mathrm{Rat}}(E) \to \underline{\mathrm{Prin}}(E) \to 0$$

where $\underline{\mathrm{Rat}}(E)$ is the sheaf of rational sections of E and $\underline{\mathrm{Prin}}(E)$ the sheaf

of principal parts* with values in E. Taking global sections, we obtain

$$0 \to H^0(X, E) \to \mathrm{Rat}(E) \to \mathrm{Prin}(E) \to H^1(X, E) \to 0. \tag{1}$$

We will work with situations where the bundles E are defined over finite fields $\mathbb{F}_q$, and the cohomology spaces have the same dimensions over $\mathbb{F}_q$ and its algebraic closure. We denote $\overline{\alpha}$ the principal part of a global rational section α of E, and we write $[p]$ for the cohomology class of a principal part $p \in \mathrm{Prin}(E)$. See for example [8] for further information.

Definition 2.1. Let $\mathcal{E}$ be a locally free sheaf of rank $r \geq 1$ on a curve X, chosen in such a way that the linear system $\Upsilon = \mathcal{O}_{\mathbf{P}E}(1)$ on the corresponding $\mathbf{P}^{r-1}$-bundle $\mathbf{P}E$ over X is very ample, and $h^1(\Upsilon) = 0$. We map $\mathbf{P}E$ into $\mathbf{P}^{k-1}$ with the complete linear system $H^0(\Upsilon)$. The image T is by definition a smooth scroll, and isomorphic to $\mathbf{P}E$.

In particular, if $X = \mathbf{P}^1$, and $\mathcal{E} = \mathcal{O}_{\mathbf{P}^1}(e_1) \oplus \cdots \oplus \mathcal{O}_{\mathbf{P}^1}(e_r)$, with $e_1 \geq \ldots \geq e_r \geq 1$ then $\deg E = f = e_1 + \cdots + e_r \geq 2$. In this case $k = f + r$, and the image T is by definition a rational normal scroll of type $(e_1, \ldots, e_r)$.

Remark 2.1. If the locally free sheaf $\mathcal{E}$ satisfies certain stability conditions, then the dimension k is equal to $\deg(E)+r(1-g)$ also for non-rational curves (twisting E with a large enough multiple of the line bundle corresponding to a fiber F if necessary). In general (see [11], Proposition 2.1),

$$h^0\left(\mathcal{O}_{\mathbf{P}E}(b_1) \otimes \mathcal{O}_{\mathbf{P}E}(b_2 F)\right) = \binom{b_1 + r - 1}{r - 1}(\mu b_1 + b_2 + 1 - g)$$

in this case, where μ is the slope $\frac{\deg(\mathcal{E})}{r}$ of E. Here we only study the case $b_1 = 1$. From the proof of this result it also follows that $h^1(X, E) = 0$ under these stability conditions, and this gives $h^1(\mathbf{P}E, \Upsilon) = 0$.

Next, we will describe some codes which can be produced from these objects.

3. Matrix description

In this section we will define a generalization of the strongly algebraic geometric (SAG) codes considered in [7].

*Note that this is a different object from the "principal part sheaf" $\mathcal{P}^k_X(\mathcal{E})$ considered by for example Laksov and Thorup [9].

3.1. *Strongly algebraic geometric scroll codes*

Let C be a code over a finite field $\mathbb{F}_q$ defined as follows. Start with a scroll $\mathbf{P}E$ over a curve X, which is embedded in $\mathbf{P}^{k-1}$ as described above, and let E be defined over $\mathbb{F}_q$. Suppose γ is the number of $\mathbb{F}_q$-rational points on X; if $X = \mathbf{P}^1$ then $\gamma = q + 1$. Then we recall that T contains

$$n = \gamma(q^{r-1} + q^{r-2} + \cdots + q + 1) \tag{2}$$

points over $\mathbb{F}_q$. Choose s of the γ fibers of $\mathbf{P}E$ over X, and in each fiber we pick at least r points, such that these points span the fiber. Altogether we have then chosen v points $P_1, \ldots, P_v$, and $sr \leq v \leq n$. Let Υ be the linear system on $\mathbf{P}E$ described above, and look at the map $\phi\colon H^0(\mathbf{P}E, \Upsilon) \to (\mathbb{F}_q)^v$ over F_q defined by $\phi(f) = (f(P_1), \ldots, f(P_v))$. The code C is the image of ϕ.

Let M be the divisor on $\mathbf{P}E$ corresponding to the s fibers spanned by the P_i, so M is numerically (linearly if $X = \mathbf{P}^1$) equivalent to sF on $\mathbf{P}E$, where F is the class of a fiber. Recall that Υ is the bundle associated to the hyperplane system $\mathcal{O}_{\mathbf{P}E}(1)$. Look at the exact sequence of sheaves

$$0 \to \Upsilon(-M) \to \Upsilon \to \frac{\Upsilon}{\Upsilon(-M)} \to 0.$$

This induces an exact cohomology sequence

$$0 \to H^0(\Upsilon(-M)) \to H^0(\Upsilon) \to (\mathbb{F}_q)^{sr} \to H^1(\Upsilon(-M)) \to H^1(\Upsilon) \to 0 \tag{3}$$

over $\mathbb{F}_q$. In turn this induces a sequence of maps

$$0 \to H^0(\Upsilon(-M)) \to H^0(\Upsilon) \to (\mathbb{F}_q)^{sr} \to (\mathbb{F}_q)^v,$$

where each function on the union of the s chosen fibers is evaluated at the v points by the last map of the sequence. We denote this map by g. Of course we claim no exactness of the last sequence at $(\mathbb{F}_q)^{sr}$. We see from this that we can regard the linear code C as the image of the quotient space $\frac{H^0(\Upsilon)}{H^0(\Upsilon(-M))}$ over $\mathbb{F}_q$. In a special case considered by many authors one picks all $\mathbb{F}_q$-rational points in all fibers, so $s = \gamma$ and we pick $q^{r-1} + \cdots + q + 1$ points in each fiber, and then $v = n = \gamma(q^{r-1} + \cdots + q + 1)$. The last two sequences above are simplified if $H^0(\Upsilon(-M)) = 0$. For $X = \mathbf{P}^1$ this happens if $s \geq e_1 + 1$, and such an s can be chosen if $q \geq e_1$.

We now look at a special case:
Here we pick instead exactly r points in each of the s fibers, and we also pick them such that they span the fibers. Write D for the sum of the points of X over which the divisor M on $\mathbf{P}E$ is supported. Clearly this is of the form

$x_1 + \cdots + x_s$ for distinct $\mathbb{F}_q$-rational points $x_i \in X$. For each $i = 1, \ldots, s$ we will denote the points in the fiber over x_i by $P_{i,1}, \ldots, P_{i,r}$.

Then $v = sr$, the map g described above is an isomorphism of vector spaces, and we may identify the spaces $(\mathbb{F}_q)^{sr}$ and $(\mathbb{F}_q)^v$ of the last sequence, and regard the map $H^0(\Upsilon) \to (\mathbb{F}_q)^{sr} = (\mathbb{F}_q)^v$ of the long exact cohomology sequence as an evaluation map in the $v = sr$ points.

Now the cohomology $H^0(\Upsilon)$ and $H^1(\Upsilon(-M))$ can be identified with cohomology spaces of bundles on X. We have

$$H^0(\mathbf{P}E, \Upsilon) = H^0(X, E)$$

and

$$\begin{aligned} H^1(\mathbf{P}E, \Upsilon(-M)) &= H^1(X, E \otimes \pi_*(\mathcal{O}_{\mathbf{P}E}(-M))) \\ &= H^1(X, E \otimes \mathcal{O}(-D)) \\ &= H^0(X, K_X(D) \otimes E^*)^* \text{ by Serre duality} \\ &= H^0(\mathbf{P}E_1, \Upsilon_1)^* \end{aligned}$$

where Υ_1 is a suitable line bundle on a scroll $\mathbf{P}E_1$. Here $E_1 = K_X(D) \otimes E^*$, and Υ_1 is $\mathcal{O}_{\mathbf{P}E_1}(1)$ for this locally free sheaf of rank r on X. We also get

$$H^1(T, \Upsilon) = H^1(X, E) = H^0(X, K \otimes E^*)^* = H^0(T_1, \Upsilon_1(-M))^*;$$

here and in the sequel, we denote by T_1 the image of $\mathbf{P}E_1$ by the linear system $\mathcal{O}(1)$. For $X = \mathbf{P}^1$, this becomes

$$\begin{aligned} &H^1(\mathbf{P}E, \Upsilon(-sF)) = H^1(\mathbf{P}^1, \mathcal{O}(e_1 - s) \oplus \mathcal{O}(e_2 - s) \oplus \cdots \oplus \mathcal{O}(e_r - s)) \\ &= H^0(\mathbf{P}^1, \mathcal{O}(s - e_1 - 2) \oplus \mathcal{O}(s - e_2 - 2) \oplus \cdots \oplus \mathcal{O}(s - e_r - 2))^* = H^0(T_1, \Upsilon_1)^*. \end{aligned}$$

In fact Υ_1 is $\mathcal{O}(1)$ on $\mathbf{P}E_1$, where $\mathcal{E}_1 = \mathcal{O}(s - e_d - 2) \oplus \mathcal{O}(s - e_{d-1} - 2) \oplus \cdots \oplus \mathcal{O}(s - e_1 - 2)$ on $\mathbf{P}^1$.

The identifications of the H^0-spaces follows from [13], p. 110. Moreover, this and the identification of the H^1-spaces follows from a straightforward generalization of Lemma V, 2.4 of [5]: Clearly $H^i(\Upsilon(-M))_x = H^i(\Upsilon_x) = 0$, for all $i > 0$ and all points $x \in X$, since $\Upsilon|_x = \mathcal{O}_{\mathbf{P}^{d-1}}(1)$. Therefore $R^i(\pi_* \Upsilon(-M)) = 0$ for $i > 0$. See [5], Chapter III, Ex. 11.8., and Chapter III, Ex. 8.4.

We note that $\mathbf{P}E_1 \cong \mathbf{P}E^*$, since $E_1 = E^* \otimes K_X(D)$. Hence the long exact cohomology sequence (3) becomes:

$$\begin{aligned} 0 \to H^0(\mathbf{P}E, \Upsilon(-M)) \to H^0(\mathbf{P}E, \Upsilon) \to (\mathbb{F}_q)^{sr} \\ \to H^0(\mathbf{P}E^*, \Upsilon_1)^* \to H^0(\mathbf{P}E^*, (\Upsilon_1) - M)^* \to 0, \quad (4) \end{aligned}$$

over $\mathbb{F}_q$, which simplifies to

$$0 \to H^0(\mathbf{P}E, \Upsilon) \to (\mathbb{F}_q)^{sr} \to H^0(\mathbf{P}E^*, \Upsilon_1)^* \to 0 \tag{5}$$

if $h^0(\mathbf{P}E, \Upsilon(-M)) = h^1(\mathbf{P}E, \Upsilon) = 0$. Dualizing, we get

$$\begin{aligned} 0 \to H^0(\mathbf{P}E^*, \Upsilon_1(-M)) \to H^0(\mathbf{P}E^*, \Upsilon_1) \to (\mathbb{F}_q)^v \\ \to H^0(\mathbf{P}E, \Upsilon)^* \to H^0(\mathbf{P}E, \Upsilon(-M))^* \to 0 \end{aligned} \tag{6}$$

which simplifies to

$$0 \to H^0(\mathbf{P}E^*, \Upsilon_1) \to (\mathbb{F}_q)^{sr} \to H^0(\mathbf{P}E, \Upsilon)^* \to 0 \tag{7}$$

under the conditions stated. This motivates the following, generalizing the definition of a SAG code (see [7], §2).

Definition 3.1. A scroll code C defined as above by evaluation of sections of Υ at exactly r independent points of s fibers is called a *strongly algebraic geometric scroll code* or *SAGS code* if $h^0(\mathbf{P}E, \Upsilon(-M)) = h^1(\mathbf{P}E, \Upsilon) = 0$.

The sequences (4) and (5) give that we obtain a generator matrix for C by evaluating sections in $H^0(\mathbf{P}E, \Upsilon)$ at the v points. On the other hand, (6) and (7) show that we get a generator matrix for (a code equivalent to) C^*, that is, a parity check matrix for C, by evaluating sections in $H^0(\mathbf{P}E^*, \Upsilon_1)$ at some v "dual" points (all of them in fibers corresponding to the same s points over $\mathbf{P}^1$. We will say more about this in the next section.

Remark 3.1. Recall that a Goppa code $C(D, G)$ is strongly algebraic geometric if $2g - 2 < \deg G < s$. In analogy with this, we notice that C is a SAGS code if E is semistable and the following inequality holds:

$$r(2g - 2) < deg(E) < rs. \tag{8}$$

For example, suppose $s \geq 2g$. By [11], Remark 2.1, there exist semistable (in fact, even so-called p-semistable) bundles of degree zero and rank r on X for X, r and q "general enough". Twisting such a bundle by an effective divisor of degree strictly between $2g - 2$ and s, we get an E which defines a SAGS.

3.2. *Another description of the codes*

Here we give another way of defining the codes C and C^* which will be useful for our work later with extensions.

At each $x \in X$, a section t of $O_{\mathbf{P}E}(1) \to \mathbf{P}E$ restricts to a linear form $t(x)$ on the projective space $\mathbf{P}E|_x$; that is, a vector in $E|_x$. Evaluation of t at $P = \langle e^* \rangle \in \mathbf{P}E|_x$ is simply restriction of $t(x)$ to the line in $E^*|_x$ spanned by e^*. The points $P_{1,1}, \ldots, P_{s,r}$ come from covectors $e^*_{1,1}, \ldots, e^*_{s,r} \in E^*$ which form a basis of each of the fibers of E^* over the points of D. Thus there exist unique $e_{1,1}, \ldots, e_{s,r} \in E$ such that $e^*_{i,j}(e_{i',j'}) = \delta_{j,j'}$, when this contraction makes sense (that is, when $x_{i'} = x_i$). For each (i,j), we have $\langle e^*_{i,j} \rangle^* = \langle e_{i,j} \rangle$, and restriction of $t(x)$ to $\langle e^*_{i,j} \rangle$ yields

$$(\text{coefficient of } e_{i,j} \text{ in } t(x_i)) \cdot e_{i,j}$$

which is well defined since the set of all the $e_{i,j}$ includes a basis of each of the chosen fibers. We write $\lambda_{i,j}$ for this coefficient. Identifying $\mathbb{F}_q^{sr}$ with $\bigoplus_{i,j} \mathbb{F}_q \cdot e_{i,j}$, we see that t is sent to the sr-tuple $(\lambda_{1,1}, \ldots, \lambda_{s,r})$. If we write this more suggestively as

$$((\lambda_{1,1}, \ldots, \lambda_{1,r}), \ldots (\lambda_{s-1,1}, \ldots, \lambda_{s,r}))$$

and consider t now as a section of the vector bundle $E \to X$, then we see that the r-tuple $(\lambda_{(i,1}, \ldots, \lambda_{i,r})$ is just the expression of $t(x_i)$ in terms of our chosen basis of $E|_{x_i}$. We have natural identifications

$$E|_D = \bigoplus_{i,j} \mathbb{F}_q \cdot e_{i,j} = \bigoplus_{i,j} O_{\mathbf{P}E}(1)|_{\langle e^*_{i,j} \rangle}$$

allowing us to pass between the interpretations of t as a section of $E \to X$ and of $\mathcal{O}_{\mathbf{P}E}(1) \to \mathbf{P}E$. Thus the sequence (5) is identified with $0 \to H^0(X,E) \to E|_D \to H^1(X, E(-D)) \to 0$.

We now set $H := E^*(D)$. Note that $H = \pi_* (O_{\mathbf{P}E^*}(1) \otimes M)$. Now $E|_D = H^*(D)|_D$, which can be viewed as (the global sections of) the subsheaf of $\underline{\mathrm{Prin}}(H^*)$ of principal parts supported on D with at most simple poles. For each (i,j), let $p_{i,j} \in \mathrm{Prin}(H^*)$ be the principal part defined by $e_{i,j}$. (Of course, this is supported at x_i with a simple pole.) Then we have $H^*(D)|_D = \bigoplus_{i,j} \mathbb{F}_q \cdot p_{i,j}$ and the sequence (5) becomes

$$0 \to H^0(X, H^*(D)) \xrightarrow{\rho} H^*(D)|_D \xrightarrow{\nu} H^1(X, H^*) \to 0$$

where ρ and ν are induced by the principal part map[†] and the coboundary map in (1) respectively. Explicitly, ρ sends a rational section of H^* with

[†]Since the poles are all simple, we could also think of this as the sum of the residue maps over the points of D.

poles bounded by D to its principal part, and ν sends a principal part $\lambda_{1,1}p_{1,1} + \cdots + \lambda_{s,r}p_{s,r}$ to the cohomology class $\left[\sum_{i,j} \lambda_{i,j}p_{i,j}\right]$.

Thus the code C is identified with the subspace of $H^*(D)|_D$ of elements occurring as principal parts of global rational sections of H^*, and the syndrome of an element in $\mathbb{F}_q^{sr}$ corresponds to the obstruction to lifting it to a global rational section of H^*.

3.3. *Generator and parity check matrices*

‡ In [7], generator and parity check matrices are given for the codes C and C^* when $r = 1$, that is, $\mathbf{P}E$ is the curve X. Here we generalize this approach to the present situation.

Let $t_1, \ldots, t_l$ be a basis for $H^0(X, E)$. For each $m = 1, \ldots, l$ and each (i, j), write $\lambda_{m,(i,j)}$ for the coefficient of $e_{i,j}$ in $t_m(x_i)$. Then, by the discussion in the last paragraph, the evaluation map sends t_m to the principal part $\lambda_{m,(1,1)}p_{1,1} + \cdots + \lambda_{m,(s,r)}p_{s,r}$, so the matrix of ρ with respect to the bases $\{t_m\}$ and $\{p_{i,j}\}$ is

$$\begin{pmatrix} \lambda_{1,(1,1)} & \cdots & \lambda_{l,(1,1)} \\ \vdots & & \vdots \\ \lambda_{1,(s,r)} & \cdots & \lambda_{l,(s,r)} \end{pmatrix} =: S.$$

In order to find a matrix for ν, in fact we will find one for ${}^t\nu\colon H^1(X, H^*)^* \to \left(\mathbb{F}_q^{sr}\right)^*$ and dualize. We recall explicitly the Serre duality pairing

$$H^0(X, K_X \otimes H) \times H^1(X, H^*) \to H^1(X, K_X) = \mathbb{F}_q.$$

Let p be an H^*-valued principal part and $[p]$ its cohomology class; by (1), every class in $H^1(X, H^*)$ is of this form. Let u be a global section of $K_X \otimes H$. Then $u(p) \in \mathrm{Prin}(K_X)$ and the contraction of u and $[p]$ is simply $[u(p)]$. Hence ${}^t\nu(u)$ is the linear form given by $(p \mapsto [u(p)])$.

Now, for each i, let z_i be a local coordinate on X centered at x_i. We fix an isomorphism $\mathbb{F}_q \xrightarrow{\sim} H^1(X, K_X)$ and let c be the image of 1. We describe a basis of $\left(\mathbb{F}_q^{sr}\right)^*$ dual to the basis $p_{(1,1)}, \ldots, p_{(s,r)}$ of $\mathbb{F}_q^{sr}$. For each (i, j), let $h_{i,j} \in H$ be such that $\langle h_{i,j} \rangle$ is the image of $\langle e_{i,j}^* \rangle$ under the natural isomorphism $\mathbf{P}E = \mathbf{P}(H^*(D)) \xrightarrow{\sim} \mathbf{P}H^*$. We define a linear form $\overline{h_{i,j}}$ on $\mathbb{F}_q^{sr} = \bigoplus_{i,j} \mathbb{F}_q \cdot p_{i,j}$ by $p \mapsto [dz_i \otimes h_{i,j}(p)]$. By construction, $\overline{h_{i,j}}(p_{i',j'})$ is nonzero if and only if $j = j'$ and $i = i'$. Multiplying the $h_{i,j}$ by nonzero

‡This subsection is logically independent of the rest.

scalars if necessary, we can assume that $\overline{h_{i,j}}(p_{i',j'}) = c \cdot \delta_{i,i'}\delta_{j,j'}$, so we obtain the required basis.

Now let $u \in H^0(X, K_X \otimes H)$. For each i, j we have

$$^t\nu(u)(p_{i,j}) = [p_{i,j}(u)] = c \cdot (\text{coefficient of } dz_i \otimes h_{i,j} \text{ in } u(x_i)) \qquad (9)$$

Let us view u as a section of the line bundle $\pi^* K_X \otimes O_{\mathbf{P}H}(1)$. As we did for E and E^*, for each (i, j), let $h^*_{i,j}$ be the dual basis vector of $h_{i,j}$ in H^*. To evaluate u at the point $\langle h^*_{i,j}\rangle \in \mathbf{P}H$, we restrict $u(x_i) \in (K_X \otimes H)|_{x_i}$ to the line $\langle h^*_{i,j}\rangle$ in $H^*|_{x_i}$. This gives an element of $K_X|_{x_i} \otimes \langle h^*_{i,j}\rangle^* = \mathbb{F}_q \cdot (dz_i \otimes h_{i,j})$, and the coefficient is the same as that of c in (9).

Thus, if $u(x_i) = dz_i \otimes (\mu_{i,1}h_{i,1} + \cdots + \mu_{i,r}h_{i,r})$ for each i, then

$$^t\nu(u) = \mu_{1,1}\overline{h_{1,1}} + \cdots + \mu_{s,r}\overline{h_{s,r}}$$

is the expression of $^t\nu(u)$ with respect to the basis $\overline{h_{1,1}}, \ldots, \overline{h_{s,r}}$. Thus we can view $^t\nu(u)$ as the evaluation of u at each of the points $\langle h^*_{i,j}\rangle \in \mathbf{P}H$.

Let now $u_1, \ldots, u_{l'}$ be a basis for $H^0(X, K_X \otimes H)$. For each $n = 1, \ldots, l'$, write $\mu_{n,(i,j)}$ for the coefficient of $dz_i \otimes h_{i,j}$ in $u_n(x_i)$. Then the matrix of $^t\nu$ with respect to our chosen bases is

$$\begin{pmatrix} \mu_{1,(1,1)} & \cdots & \mu_{l',(1,1)} \\ \vdots & & \vdots \\ \mu_{1,(s,r)} & \cdots & \mu_{l',(s,r)} \end{pmatrix} =: {}^tR.$$

The rows of this matrix give generators for C^*. But the $(ri + j)$th row represents the values of each of the u_n at $\langle h^*_{i,j}\rangle$, so we see explicitly how C^* is also an evaluation code. The matrix of ν with respect to $\{p_{1,1}, \ldots, p_{s,r}\}$ and the basis of $H^1(X, H^*)$ dual to $\{u_1, \ldots, u_{l'}\}$ is R. By exactness, $RS = 0$ and $^tS^tR$ are zero, and ^tS and R are parity check matrices for C^* and C respectively.

Note: As we have defined them, C and C^* belong to different vector spaces. However, since we have the mutually dual bases $\{p_{i,j}\}$ and $\{\overline{h_{i,j}}\}$, we can view both codes as subspaces of $\mathbb{F}_q^{rs}$ via the vector space isomorphism $\mathbb{F}_q^{rs} \xrightarrow{\sim} \left(\mathbb{F}_q^{rs}\right)^*$ sending each $\overline{h_{i,j}}$ to $p_{i,j}$.

Remark 3.2. It follows from the discussion above that the orthogonal complements (or duals) of SAGS codes are (code equivalent to) SAGS codes in general, just like for the traditional case $r = 1$. Hence the description above lends itself just as well to make parity check matrices as to make generator matrices. This is one of the virtues of Goppa codes (based on curves),

which it has been hard to reproduce for codes produced from varieties of higher dimension. If one picks all points of for example Grassmannians or scrolls, then the coordinates of these points are suitable for producing columns of generator matrices of codes that are interesting. But if one tries to use the same points as columns of parity check matrices, then because of the existence of linear spaces inside the varieties (lines), one cannot exceed minimum distance 3. Hence, in order to get essentially self-dual classes of codes, like for Goppa codes, one must revise the way one picks points.

3.4. *The link with extensions*

Since the column vectors of the parity check matrix of C are described through coordinates of points of $\mathbf{P}E^*$ embedded by the complete linear system Υ_1, we see that the (projectivized) syndrome space of C in a natural way is identified with

$$\mathbf{P}H^0(\mathbf{P}E^*, \Upsilon_1) = \mathbf{P}H^0(X, K(D) \otimes E^*).$$

If $X = \mathbf{P}^1$ and $\Upsilon = \mathcal{O}(e_1) \oplus \cdots \mathcal{O}(e_d)$, then this is

$$\mathbf{P}H^0(\mathbf{P}E^*, \Upsilon_1) = \mathbf{P}H^0(\mathbf{P}^1, \mathcal{O}(s-e_1-2) \oplus \mathcal{O}(s-e_2-2) \oplus \cdots \oplus \mathcal{O}(s-e_d-2))$$

The syndrome space can also be identified with

$$H^1(X, E \otimes M^*) = H^1(X, H^*),$$

where as before $H^* = E(-D)$. For $X = \mathbf{P}^1$ and $\Upsilon = \mathcal{O}(e_1) \oplus \cdots \mathcal{O}(e_d)$, this is

$$H^1(\mathbf{P}^1, \mathcal{O}(e_1 - s)) \oplus \mathcal{O}(e_2 - s)) \oplus \cdots \oplus \mathcal{O}(e_d - s)) = H^1(X, H^*),$$

where $H = \mathcal{O}(s - e_1) \oplus \mathcal{O}(s - e_2) \oplus \cdots \oplus \mathcal{O}(s - e_d)$.

Now $H^1(X, H^*) = Ext^1(\mathcal{O}_X, H^*)$ can be identified with isomorphism classes of extensions

$$0 \to H^* \to W \to O_X \to 0.$$

In the next section, we will relate the geometry of the space $\mathbf{P}H^1(X, H^*)^*$ to the behaviour of these extensions.

4. Geometry of extension spaces

Henceforth, to allow for slightly greater generality, instead of the bundle H^* we will work with $\mathrm{Hom}(F_2, F_1)$ for bundles F_1 and F_2 over X. Recall that the *decomposable locus* of $\mathrm{Hom}(F_2, F_1)$ is the locus of maps of rank one.

This is a determinantal subvariety of $\mathrm{Hom}(F_2, F_1)$, defined by the vanishing of all 2×2 minors of the maps. We denote Δ the locus defined by these (homogeneous) polynomials in $\mathbf{P}\,\mathrm{Hom}(F_2, F_1)^*$.

Example 4.1. If F_1 and F_2 are both of rank two then Δ is a bundle of smooth quadrics in the $\mathbf{P}^3$-bundle $\mathbf{P}\,\mathrm{Hom}(F_2, F_1)^* \to X$. Of course, if either one is a line bundle then $\Delta = \mathbf{P}\,\mathrm{Hom}(F_2, F_1)^*$.

4.1. *Embeddings of scrolls*

Here we give another description of the map from $\mathbf{P}H$ into the projectivized syndrome space $\mathbf{P}H^1(X, H^*)^*$, which will be adapted for our study of extensions.

Let $V \to X$ be any vector bundle. For any $\mathbb{F}_q$-rational point $x \in X$, we have a short exact sequence

$$0 \to \mathcal{V} \to \mathcal{V}(x) \to \frac{\mathcal{V}(x)}{\mathcal{V}} \to 0$$

whose cohomology sequence includes

$$\cdots \to H^0(X, V(x)) \to V(x)|_x \to H^1(X, V) \to \cdots$$

Since $\mathbf{P}(V^*(-x))|_x$ is canonically isomorphic to $\mathbf{P}V^*|_x$, the projectivized coboundary map gives rise to a map $\psi_x \colon \mathbf{P}V^*|_x \dashrightarrow \mathbf{P}H^1(X, V)^*$. We define a map $\psi \colon \mathbf{P}V^* \dashrightarrow \mathbf{P}H^1(X, V)^*$ by taking the product of all the ψ_x.

Now by Serre duality and the projection formula, we have an identification

$$H^1(X, V) \xrightarrow{\sim} H^0(\mathbf{P}V^*, \pi^* K_X \otimes O_{\mathbf{P}V^*}(1))^*. \tag{10}$$

Lemma 4.1. *([6], §2) Via the above identification, ψ coincides with the standard map $\mathbf{P}V^* \dashrightarrow |\pi^* K_X \otimes O_{\mathbf{P}V^*}(1)|$. Moreover, ψ is an embedding if and only if for all $x, y \in X$, we have $h^0(X, K_X(-x-y) \otimes V^*) = h^0(X, K_X \otimes V^*) - 2r$.*

Remark 4.1. The key feature of this interpretation is that ψ sends $\langle v \rangle \in \mathbf{P}V^*|_x$ to the projectivized cohomology class of a V-valued principal part supported at x with a simple pole in the direction v. This will allow us to use the alternative construction of the code C in §3.2 to understand the geometry of the syndromes.

We recall a definition:

Definition 4.1. Let $F_2 \to X$ be a vector bundle. Then an *elementary transformation of* F_2 is a vector bundle defined by a locally free subsheaf of $\mathcal{F}_2$ of rank equal to the rank of F_2.

Such subsheaves can be defined using principal parts. If F_1 is another vector bundle over X, then any $\mathrm{Hom}(F_2, F_1)$-valued principal part naturally defines a map $\mathcal{F}_2 \to \underline{\mathrm{Prin}}(F_1)$. Then the kernel of such a map defines an elementary transformation of F_2. Moreover, any elementary transformation of F_2 is of this form (although not in a unique way).

We will need the following technical result on extension classes:

Lemma 4.2. *([6], §4.1) Let W be an extension of F_2 by F_1. An elementary transformation of $\mathcal{G}$ of $\mathcal{F}_2$ lifts to a vector subbundle of W if and only if the class $\delta(W)$ of the extension can be defined (cf. (1)) by a principal part $p \in \mathrm{Prin}(\mathrm{Hom}(F_2, F_1))$ such that $\mathcal{G} = \ker(p\colon \mathcal{F}_2 \to \underline{\mathrm{Prin}}(F_1))$.*

Now we can give the main result of this section.

Theorem 4.1. *Let $0 \to F_1 \to W \to F_2 \to 0$ be a nontrivial extension. Then $\langle \delta(W) \rangle$ belongs to the linear span of at most h independent points of $\Delta|_D$ if and only if W has a subbundle lifting from an elementary transformation of F_2 of the form*

$$0 \to \mathcal{G} \to \mathcal{F}_2 \to \tau \to 0 \tag{11}$$

where $\tau \subset \underline{\mathrm{Prin}}(F_1)$ is a skyscraper sheaf of length at most h supported on D and with at most simple poles.

Proof. Suppose $\langle \delta(W) \rangle$ belongs to the linear span of at most h independent points of $\Delta|_D$ in $\mathbf{P}H^1(X, \mathrm{Hom}(F_2, F_1))^*$. Then by the alternative definition of ψ given above, $\delta(W)$ can be defined by $p \in \mathrm{Prin}(\mathrm{Hom}(F_2, F_1)$ of the form $\sum_{j=1}^{h} p_j$ where each p_j is a principal part supported at one point of D with a simple pole along some rank one map. Thus we have a short exact sequence

$$0 \to \mathcal{G} \to \mathcal{F}_2 \xrightarrow{p} \tau \to 0$$

where $\tau \subset \underline{\mathrm{Prin}}(F_1)$ is a skyscraper sheaf of length at most h supported on D and with at worst simple poles. By Lemma 4.2, the sheaf $\mathcal{G}$ lifts to a subbundle of W.

Conversely, suppose an elementary transformation $\mathcal{G}$ of $\mathcal{F}_2$ of the stated type lifts to a subbundle of W. By Lemma 4.2, the class $\delta(W)$ can be defined

by some $p \in \mathrm{Prin}(\mathrm{Hom}(F_2, F_1))$ which, viewed as a map $\mathcal{F}_2 \to \underline{\mathrm{Prin}}(F_1)$, has kernel $\mathcal{G}$. From the sequence (11) we deduce that p is supported along D and has at most simple poles. We write $p = \sum_{x \in D} p_x$, where each p_x is a principal part supported at one point x, and then write each p_x as a sum of rank one homomorphisms, of minimal length. Since τ is of length at most h, any such expression for p contains at most h independent such rank one homomorphisms. By the alternative definition of ψ, the point $\langle \delta(W) \rangle$ is contained in the span of the corresponding at most h rank one points of $\mathbf{P}\,\mathrm{Hom}(F_2, F_1)^*$. □

Remark 4.2. Suppose $F_1 = H^*$ and $F_2 = O_X$, so we are in the situation of the last section. Then this theorem shows that the extension $0 \to H^* \to W \to O_X \to 0$ can be "quasi-inverted" to a short exact sequence

$$0 \to O_X(-A) \to W \to (H^*)' \to 0$$

for some effective divisor $A \leq D$ and some bundle $(H^*)'$, if $\langle \delta(W) \rangle$ lies in the linear span of some points of $\mathbf{P}H^*$ all lying over the support of the divisor $A \leq D$. In this case the rank of each p_x can be at most 1.

Remark 4.3. When $\mathbb{F}_q$ is replaced with the complex number field, a generalization of Theorem 4.1 is proven in [6], Theorem 4. For example, p may have poles of higher order. However, the above proof suffices for the situation we are considering.

Now we give some alternative ways of viewing the condition of Theorem 4.1. Firstly, it is equivalent to saying that $\delta(W)$ belongs to

$$\begin{aligned}\ker(H^1(X, \mathrm{Hom}(\mathcal{O}_X, H^*)) \to H^1(X, \mathrm{Hom}(\mathcal{O}_X(-A), H^*))) \\ = \ker(H^1(X, H^*) \to H^1(X, H^*(A))) \\ = \ker(H^0(X, K_X \otimes H)^* \to H^0(X, K_X \otimes H(-A))^*).\end{aligned}$$

We look instead at the (isomorphic) space of extensions of type

$$0 \to \mathcal{O}_X \to W \to H \to 0.$$

Then it follows from a dual version, as in [12], Lemma 3.2, that there is a surjection $W \to \mathcal{O}_X(A) \to 0$ if and only if $\delta(W)$ belongs to

$$\begin{aligned}\ker(H^1(X, \mathrm{Hom}(H, \mathcal{O}_X) \to H^1(X, \mathrm{Hom}(H, \mathcal{O}_X(A))) \\ = \ker(H^1(X, H^*) \to H^1(X, H^*(A)).\end{aligned}$$

We see that the two kernels are the same, and we have two alternative descriptions.

In the first description we may view $\mathcal{O}_X(-A)$ as a special case of a locally free sheaf G with a sheaf injection $\phi : G \to \mathcal{O}_X$, such that ϕ factors via a map $G \to W$.

In the second description $\mathcal{O}_X(A)$ is a special case of a locally free sheaf G with a sheaf homomorphism $\phi : \mathcal{O}_X \to G$, such that ϕ extends to a homomorphism $W \to G$.

The common kernel can also be viewed as that of a map

$$Ext(\mathcal{O}_X, H^*) \to Ext(\mathcal{O}_X, H^*(A)),$$

using Proposition 6.3 of [5]. Then by Proposition 6.7 of [5], we interpret this as the kernel of a map

$$Ext(H, \mathcal{O}_X) \to Ext(H(-A), \mathcal{O}_X) = Ext(H, \mathcal{O}_X(A)).$$

Example 4.2. An easy case to handle is when $X = \mathbf{P}^1$ and we pick all r points in each fiber along the directrix curves. Assume $s = q + 1$. It is clear that the natural subscroll of type $(e_2, \ldots, e_r)$ is contained in a hyperplane, and that any hyperplane containing this subscroll intersects the first directrix in e_1 points, and for some such hyperplane they can taken to be rational. One easily sees that this hyperplane contains $e_1 + (r-1)(q+1)$ points, which is largest possible, and that the minimum distance of the code then is $q+1-e_1$. Working dually, with the scroll T_1 of type $(q-1-e_r, \ldots q-1-e_1)$, we see that we get a linear dependency between $q+1-e_1$ points on the directrix curve of smallest degree $(q-1-e_1)$, which again indicates minimum distance $q+1-e_1$. Furthermore the higher weights d_i increase by one until we reach a value of i such that no codimension i-space contains the subscroll of $(e_2, \ldots, e_r)$. This space contains $e_1 + e_2 + (r-2)(q+1)$ points, so $d_i = 2q + 2 - e_1 - e_2$. We leave it to the reader to make the remaining calculations to the determine the complete weight hierarchy. It is a sad fact that a code with such a nice description has such bad code-theoretical properties.

5. Error correction

We return to the code C. Here we give a geometric condition for the correctability of the error, in terms of the image of the embedding of $\mathbf{P}\operatorname{Hom}(F_2, F_1)^*$ in the syndrome space.

Important hypothesis: We will assume that the $\operatorname{Hom}(F_2, F_1)$-valued principal parts $p_{1,1}, \ldots, p_{s,r}$ are all along directions corresponding to rank

one homomorphisms. This is possible since each fiber is spanned by such maps.

5.1. *A geometric condition for correctability*

The following is analogous to [7], Theorem 3.4.

Theorem 5.1. *Suppose a codeword* $\mathbf{x} \in C$ *is transmitted, and* $\mathbf{y} = \mathbf{x} + \mathbf{e}$ *is received. Let* W *be the extension of* F_2 *by* F_1 *defined by the syndrome class* $\nu(\mathbf{y}) = \nu(\mathbf{e})$*. Then the error* $\mathbf{e}$ *has weight at most* h *only if* W *has a subbundle, necessarily of degree at least* $\deg(F_2) - h$*, lifting from an elementary transformation of* F *of the form*

$$0 \to \mathcal{G} \to \mathcal{F}_2 \to \tau \to 0$$

where $\tau \subset \underline{\mathrm{Prin}}(F_1)$ *is a skyscraper sheaf of length at most* h *supported on* D *and with at most simple poles.*

Proof. Suppose $\mathbf{e}$ has weight h. Then $\mathbf{e}$ is a principal part of the form

$$\lambda_{i_1,j_1} p_{i_1,j_1} + \cdots + \lambda_{i_h,j_h} p_{i_h,j_h}$$

for some nonzero $\lambda_{i_1,j_1}, \ldots, \lambda_{i_h,j_h} \in \mathbb{F}_q$, with the (i_h, j_h) all distinct. Then by Lemma 4.2, the kernel of the map of $\mathcal{O}_X$-modules $\mathbf{e}\colon \mathcal{F}_2 \to \underline{\mathrm{Prin}}(F_1)$ lifts to a subbundle of W. Since $\mathbf{e}$ is supported along D and is a sum of h rank one elements of $\mathrm{Hom}(F_2, F_1)$ with at most simple poles along D, this subbundle is an elementary transformation of the stated type. □

Remark 5.1. As in Theorem 3.4 of [7], we can give a geometric interpretation of this situation. By Theorem 4.1, there are at most h errors in $\mathbf{y}$ only if the class $\langle \nu(\mathbf{e}) \rangle$ belongs to an h-secant plane to $\Delta|_D$ spanned by at most h distinct points. Moreover, also as in [7], both the errors and the points of the scroll $\mathbf{P}H^*$ are defined over $\mathbb{F}_q$.

5.2. *Error location*

Assume we have the code C constructed in section 3, and that we have picked exactly r points in each of s fibers, where s typically is the number of all $\mathbb{F}_q$-rational points on X. Assume a codeword is sent, and the syndrome calculated. If this is zero, there is no problem. Otherwise, look at the corresponding point in the projectivized syndrome space $\mathbf{P}Ext^1(\mathcal{O}_X, H^*)^* = \mathbf{P}Ext^1(H, \mathcal{O}_X)^*$. This is the space where the scroll $\mathbf{P}H \cong \mathbf{P}E^* \cong T_1$ is embedded.

Now we think of error location in two steps. In Step 1 we find the various fibers of T_1 such that the syndrome is a linear combination of points of these fibers. In Step 2 we find the individual points in these fibers such that the syndrome contribution from each given fiber is a linear combination of syndromes from these individual points. It is only in Step 1 that we can use the direct analogue with the situation studied in [1], [2], [3] and [7]. On the other hand, Step 2 is basically only a linear algebra problem: the syndrome component from this fiber is a point of this fiber. Find which linear combination it is, of the r (dual) points that we picked in this fiber in the first place.

Hence we focus on Step 1. View the syndrome as an extension

$$0 \to \mathcal{O}_X \to W \to H \to 0.$$

Error location is then to find the (hopefully) unique line bundle $\mathcal{O}_X(A)$ of lowest degree such that there is a surjection of W onto $\mathcal{O}_X(A)$. Having found the divisor class, one must find the effective divisor A' in the class such that the syndrome is spanned by the fibers of T_1 corresponding to points on the divisor A'. This part of the process is described in Chapter 4 of [1] for $r = 1$.

Definition 5.1. For a rank r bundle V on a curve X we set $s_1(V) = \deg V - r \max\{\deg L\}$, for L a line subbundle of V on X.

Then we have:

Proposition 5.1. *For a given syndrome point $\nu(\mathbf{y})$, interpreted as an extension of type*

$$0 \to H^* \to W \to \mathcal{O}_X \to 0,$$

we have

$$s_1(W) \leq (r+1)a - rs + \deg \mathcal{E},$$

where a is the number of different fibers of T (over X) we must use to pick points such that errors in the positions corresponding to these points give rise to the syndrome.

Proof. We use points from a different fibers to span the syndrome, where these fibers correspond to the points $x_{i_1}, \cdots x_{i_a}$ on the curve, and we denote by $\mathcal{A}$ the sheaf $\mathcal{O}_X(x_{i_1} + \cdots x_{i_a})$. Then $\mathcal{A}^*$ is a subsheaf of rank one of $\mathcal{W}$, and

$s_1(W) \leq \deg W + (r+1)\deg A = \deg H^* + (r+1)a = \deg \mathcal{E} - rs + (r+1)a$. □

Remark 5.2. For $r = 1$ and Goppa codes this gives $s(W) \leq 2a - \deg(D - G) = 2a - d$, where d is the designed minimum distance, and we see that if the number of errors is less than designed-correctable, then W is unstable.

What does it take to ensure that the fibers spanning a point can be uniquely chosen? Assume there are errors in two fibers. If the syndrome point $\langle \nu(\mathbf{e}) \rangle$ is also in the span of two other fibers, then the span of the first two fibers has a common point with the span of the second group of two fibers, and the span of all four fibers is less than the "expected" value which is $4r - 1$. Hence a sufficient condition for this not to happen is

$$h^0(T_1, \Upsilon_1(-F_1 - F_2 - F_3 - F_4)) = h^0(X, E_1(-x_1 - x_2 - x_3 - x_4)) = h^0(X, E_1) - 4r$$

for all choices of x_1, x_2, x_3, x_4.

In general, syndromes from a fibers can be uniquely traced back to a fibers if

$$h^0(X, E_1(-B)) = h^0(X, E_1) - 2ar$$

for all choices of effective divisors B of degree $2a$ (compare with Lemma 4.1).

For $X = \mathbf{P}^1$ this happens if $s - e_1 - 2 - 2a \geq -1$. For curves of higher genus we have:

Proposition 5.2. *Let E be a stable bundle of rank $r > 1$ on X. Then errors in*

$$a \leq \frac{\mu(H)}{2} = \frac{s - \mu(\mathcal{E})}{2} = \frac{rs - \deg \mathcal{E}}{2r}$$

different fibers can be traced back to a unique choice of a fibers.

Proof. By Riemann–Roch, we have

$$h^0(X, E_1(-B)) = \deg E_1 - 2ra + r(1 - g) + h^0(X, K_X(B) \otimes E_1^*).$$

We show that $h^0(X, K_X(B) \otimes E_1^*) = h^0(X, H^*(B)) = 0$. Since E is stable, so is H^*, and one obtains $h^0(X, H^*(B)) = 0$ unless $2a > \mu(H)$. (By the same argument, $H^0(X, E_1) = \deg E_1 + r(1 - g)$, so we obtain the desired conclusion.) □

Remark 5.3. Since one can correct errors from a fibers if $a \leq \frac{\mu(H)-1}{2}$, it is tempting to conclude that one can correct up to $t = r\left(\frac{\mu(H)-1}{2}\right) = \frac{\deg H - r}{2}$ errors (which holds for $r = 1$, where $\deg H$ is the designed minimum

distance), since there are r points in each fiber. But the discussions so far only makes this true for t errors if they are clustered in as few as $\frac{\mu(H)-1}{2}$, fibers. If they are "spread out" on more fibers, the discussion above does not ensure unique decoding of more than $\frac{\mu(H)-1}{2}$ errors. See also Remark 5.6.

Proposition 5.3. *Suppose E is a stable bundle on X. Then the syndrome of an error which can be traced uniquely back to a choice of a fibers in the sense of the previous result, that is, where the number a of fibers where errors are made is at most $\frac{\mu(H)}{2}$, defines an extension $0 \to H^* \to W \to \mathcal{O}_X \to 0$ with*

$$s_1(W) < \frac{(r-1)(\deg \mathcal{E} - sr)}{2r}.$$

Proof. Insert $a = \frac{\mu(H)}{2} = \frac{rs - \deg \mathcal{E}}{2r}$ in the statement of Proposition 5.1. □

We see that the right hand side is strictly negative if $r > 1$ and C is a SAGS code.

Remark 5.4. What can be said about the code parameters of the SAGS codes? Certainly the word length is sr. If enough fibers are chosen (s big enough) that the chosen fibers span the projective space $\mathbb{P}^{k-1}$ in which T lies, then the dimension of the code is k. It is much harder to find the true minimum distance of the codes. All we can say is that it depends on the choice of the points in each fiber. The case studied in Example 4.2 obviously represents bad choices. To illustrate the problem of choosing points conveniently, look at the simplest case of a code which is not a Goppa code. We choose $X = \mathbf{P}^1$, and $L = \mathcal{O}(1) \oplus \mathcal{O}(1)$. Hence T is a quadric in $X = \mathbf{P}^3$. How can we choose 2 points on each line on one of the two families, such that as few as possible among the $2q + 2$ points are contained in the same plane. The worst case involves $q + 2$ points in a plane (take two lines L_1 and L_2 on the quadric, meeting in a point P), and choose all $q+1$ points on L_2, one additional point on L_1, and q additional points on lines parallel to L_1). But there clearly exist better choices, unless q is very small. In general, the minimum distance guaranteed by the method, is $s - 2 - e_1$ for codes from rational curves (Example 4.2) and $\mu(H)$ for codes from curves of higher genus if E is stable (Proposition 5.2). But one can hope for much better true values with good choices of points.

Remark 5.5. The considerations above involve no direct decoding algorithm. Neither did [7]. Nevertheless the principles from [7] were used to

approach concrete decoding in the series of papers [1], [2], [3]. We feel that the generalization presented in the present paper of the line of thoughts in [7], should lend itself to a corresponding generalization of the results of these other papers.

Remark 5.6. We observe that this interpretation of decoding is most likely to be useful with a channel which is not symmetric. This is because correctability depends on the number of fibers over which the error is distributed, and not a priori on the size of the error. More precisely: if we knew that errors were concentrated in a fibers, where $a < \mu(H)/2$, then we could correct up to r errors in each of these fibers as easily as just one. (This is the Step 2 described above.) Thus one would expect a decoding algorithm based on this interpretation to be best adapted to a channel where the probability of an error is higher in a fiber where an error has already been made.

6. Syndrome decoding of other codes

Throughout this work we have insisted on picking exactly r points in each fiber when defining the codes. This is because we have defined the codes as evaluation codes, thus starting with a natural generator, and not a parity check matrix. To obtain the exact sequence

$$0 \to H^0(\mathbf{P}E, \Upsilon) \to (\mathbb{F}_q)^{sr} \to H^0(\mathbf{P}E^*, \Upsilon_1)^* \to 0$$

and its dual counterpart

$$0 \to H^0(\mathbf{P}E^*, \Upsilon_1) \to (\mathbb{F}_q)^{sr} \to H^0(\mathbf{P}E, \Upsilon)^* \to 0,$$

we had to choose r points in each fiber, spanning it. As in §3, the last sequence defines a parity check matrix via evaluation of the dual points in each fiber.

An easier approach for our purposes would of course have been to start with T_1 and its complete linear system Υ_1 in the first place, and define a code C by a parity check matrix obtained from evaluation of sections of Υ_1 in some more arbitrarily chosen points on the fibers of T_1. As an extreme case we could have picked all $\mathbb{F}_q$-rational points of all fibers. Everything said above about Step 1 of the last section would then have been unaltered, but in Step 2 we would be far from having unique decoding, unless we knew for some reason that at most a single error could be made in each individual fiber. (As mentioned in Remark 3.2, the minimum distance would be 3 in this extreme case).

Nevertheless, if one defines codes from scrolls via parity check matrices instead of via generator matrices (as evaluation codes), then one obtains a larger class of codes, for which one can interpret decoding as described above via vector bundle manipulations.

References

1. T. Bouganis and D. Coles, *A Geometric View of Decoding AG Codes*, Proc. AAECC-15. May, 2003.
2. D. Coles, *Vector Bundles and Codes on the Hermitian Curve*, IEEE Transactions on Information Theory, **51**, no. 6. June, 2005.
3. D. Coles, *On Constructing AG Codes without Basis Functions for Riemann-Roch Spaces*, Proc. AAECC-16. February, 2006.
4. S. H. Hansen, *Error-Correcting Codes from Higher-Dimensional Varieties*, Finite Fields and their applications, **7**, 530-552 (2001).
5. R. Hartshorne, *Algebraic Geometry*, Graduate Text in Mathematics **52**, Springer Verlag (1977).
6. G. Hitching, *Geometry of vector bundle extensions and applications to the generalised theta divisor*, math.AG **0610970** (2006).
7. T. Johnsen, *Rank two bundles on algebraic curves and decoding of Goppa Codes*, International Journal of Pure and Applied Mathematics, **4**, No. 1, 33-45 (2003). See also alg-geom **9608018**.
8. G. Kempf, *Abelian integrals*, Monografías del Instituto de Matemáticas 13. Universidad Nacional Autónoma de México, Mexico, 1983.
9. D. Laksov, A. Thorup, *Weierstrass points on schemes*, Journal reine und angewandte Mathematik **460** (1995), 127–164.
10. C. C. Lomont, *Error-correcting Codes on Algebraic Surfaces*, math.NT **0309123**, (2003).
11. T. Nakashima, *Error-correcting codes on projective bundles*, Finite Fields and their applications, **12**, 222-231 (2006).
12. M. S. Narasimhan, S. Ramanan, *Moduli of vector bundles on a Compact Riemann Surface*, The Annals of Mathematics, Ser. 1, **89** (1),14-51 (1969).
13. F. O. Schreyer, *Syzygies of canonical curves and special linear series*, Math. Ann. **275**, 105-137 (1986).
14. J. H. van Lint, G. van der Geer, *Introduction to Coding Theory and Algebraic Geometry*, DMV Seminar **12**, Birkhauser (1988).

List decoding using syndromes

Peter Beelen

Technical University of Denmark
Department of Mathematics
Building 303, room 150 Denmark
E-mail : P.Beelen@mat.dtu.dk

Tom Høholdt

Technical University of Denmark
Department of Mathematics
Building 303, room 150 Denmark
E-mail : T.Hoeholdt@mat.dtu.dk

In [5] a method is described to reformulate Sudan's algorithm for list decoding (see [7]) into solving a linear system of equations involving (generalized) syndromes. The advantage of this description is that one has to deal with a smaller system, as well as that one can apply faster algorithms (e.g. Sakata's algorithm [6]). In this paper we investigate how to do the same for the more general Guruswami-Sudan algorithm for list decoding. We concentrate our investigation on RS-codes and one-point Hermitian codes.

Keywords: List decoding, syndrome, Guruswami-Sudan algorithm, RS-codes, Hermitian codes, Sakata's algorithm

1. Introduction

Let $\mathbb{F}$ be a finite field and $C \subset \mathbb{F}^n$ a code of length n. If $c \in C$ and $h = (h_1, h_2, \ldots, h_n)$ is a row of a parity check matrix of C, then it is clear that $h \cdot c = 0$. Such parity checks can therefore be used to test if a given word $w \in \mathbb{F}^n$ is an element of C, but it turns out that the expressions $h \cdot w$ (usually called syndromes of w) can be useful in decoding algorithms as well. The link between decoding and syndromes is an old one. Indeed the first known algorithm for the decoding of Reed-Solomon codes (Peterson's algorithm) uses syndromes. Suppose that the finite field $\mathbb{F}$ has q elements and let $\mathcal{P} = \{x_1, \ldots, x_n\}$ be a subset of $\mathbb{F}$ consisting of n distinct elements. We can see an RS-code of dimension $k \leq n$ as the set of all n-tuples that arise by evaluating all polynomial $f(x)$ of degree less than or equal to

$k-1$ in the points $x_1, \ldots, x_n$. The syndromes of a word w that are used in Peterson's decoding algorithm are the following:

$$S_\lambda(w) = \sum_{i=1}^{n} x_i^\lambda w_i.$$

After Sudan's algorithm for list decoding of RS-codes was discovered [7], again a reformulation in terms of certain generalized syndromes turned out to be useful [5]. These generalized syndromes can be seen as syndromes of words $w^e := (w_1^e, \ldots, w_n^e)$. Although the theoretically fastest, currently known decoding algorithms (see [1]) do not use syndromes, it turns out that for many practical parameters, the algorithm in [5] and variations of it is still the most desirable. Although generalized syndromes appear for the first time to describe list decoding algorithms, they can also be used to describe interesting decoding algorithms for RS-codes (see [2]).

All in all, it is clear that the usage of generalized syndromes is an ongoing and fruitful proces. It is therefore surprising that the more general Guruswami-Sudan algorithm for list decoding has not been reformulated in terms of syndromes, especially in the case of RS-codes. The goal of this paper is to fill this gap in the literature. We will do this for the case of RS-codes of length $q-1$ defined over the finite field with q elements and for one-point Hermitian codes of length q^3 defined over the finite field with q^2 elements. The paper is organized as follows: in Section 2, we develop the necessary tools to reformulate the Guruswami-Sudan algorithm in terms of syndromes. In Section 3 we do the same for one-point Hermitian codes. The key-tool in both cases turns out to be the residue theorem for differentials on algebraic curves.

2. RS-codes

In this section we will denote by $\mathbb{F}$ a finite field with q elements and characteristic p. Further we denote by $\mathbb{F}[x]$ the polynomial ring over $\mathbb{F}$ in one variable. For $f \in \mathbb{F}[x]$, we denote by $D_x^{(n)}(f)$ the n-th Hasse derivative of f with respect to x. The Hasse derivative can be calculated for polynomials by linearly extending the following definition for monomials:

$$D_x^{(n)}(x^m) = \binom{m}{n} x^{m-n}.$$

This formula is valid for all natural numbers m and n if we define

$$\binom{m}{n} = 0, \text{ if } n > m.$$

The well-known Leibniz rule for the n-th derivative of a product can be formulated in terms of Hasse derivatives as follows:

$$D_x^{(n)}(f \cdot g) = \sum_{i=0}^{n} D_x^{(i)}(f) D_x^{(n-i)}(g).$$

One of the advantages of Hasse derivatives is, that this formula is also valid in positive characteristic.

Let β be a primitive element of the finite field $\mathbb{F}$, i.e., suppose that $\langle\beta\rangle = \mathbb{F}\backslash\{0\}$. Further write $n := \#\mathbb{F} - 1$. For $1 \leq i \leq n$, we define $x_i := \beta^i$ and $\mathcal{P} := \{x_1, \ldots, x_n\}$. The RS-code $C_L(\mathcal{P}, (k-1)P_\infty)$ is an $[n, k, n-k+1]$ code defined over $\mathbb{F}$ obtained by evaluating all polynomials of degree less than or equal to $k-1$ in $x_1, \ldots, x_n$. For any $\alpha \in \mathbb{F}$ and $f \in \mathbb{F}[x]$, we denote by $f|_\alpha$ the evaluation of the polynomial f in the value α and by $f|_{\mathcal{P}}$ the row vector $(f|_{x_1} \cdots f|_{x_n})$.

Now we turn to list decoding of RS-codes. Suppose that a word $w = (w_1, \ldots, w_n)$ has been received. We wish to determine all codewords $c \in C_L(\mathcal{P}, (k-1)P_\infty)$ closest to w. Equivalently, we wish to find all polynomials $f_\lambda \in \mathbb{F}[x]$ of degree less than or equal to $k-1$ giving rise to these codewords. The idea in the Guruswami-Sudan algorithm [4,7] is to find a bivariate polynomial

$$Q(x, y) = \sum_{b=0}^{l} \sum_{a=0}^{l_b} x^a y^b Q_{b,a}$$

that is divisible by $y - f_\lambda$ for all the polynomials f_λ. Once one has found such a $Q(x, y)$, the polynomials f_λ can then be found by factoring $Q(x, y)$.

The polynomial $Q(x, y)$ can be found using the Guruswami-Sudan algorithm. This algorithm finds $Q(x, y)$ by solving a system of linear equations in the equations of $Q(x, y)$. More specifically: let $s \geq 1$ be a natural number and τ the number of errors one wishes to correct. For any $1 \leq i \leq n$, any $0 \leq t \leq s-1$, and $0 \leq r \leq t$ the coefficients $Q_{b,a}$ should satisfy the following system of linear equations (with $l_b := s(n-\tau) - 1 - b(k-1)$):

$$\sum_{b=0}^{l} \sum_{a=t-r}^{l_b} D_y^{(r)}(y^b)|_{w_i} D_x^{(t-r)}(x^a)|_{x_i} Q_{b,a} = 0.$$

Equivalently, one can write

$$\sum_{b=0}^{l} D_y^{(r)}(y^b)|_{w_i} \left(D_x^{(t-r)}(1)|_{x_i} \cdots D_x^{(t-r)}(x^{l_b})|_{x_i} \right) \begin{pmatrix} Q_{b,0} \\ \vdots \\ Q_{b,l_b} \end{pmatrix} = 0. \quad (1)$$

We will now transform this system of linear equations into another system involving only syndromes by multiplying from the left with a matrix. Our result generalizes a procedure described in [5], where it is always assumed that $s = 1$. If $s = l = 1$, one obtains the Peterson decoding method for RS-codes (see for example [8]). Like in [5], it will turn out that the multiplied system is easier and faster to solve than the original system.

We can write the equations in (1) as follows (where $0 \leq t \leq s-1$ and $0 \leq r \leq t$):

$$\sum_{b=r}^{l} \binom{b}{r} \begin{pmatrix} D_x^{(t-r)}(1)|_{x_1} w_1^{b-r} & \cdots & D_x^{(t-r)}(x^{l_b})|_{x_1} w_1^{b-r} \\ \vdots & & \vdots \\ D_x^{(t-r)}(1)|_{x_n} w_n^{b-r} & \cdots & D_x^{(t-r)}(x^{l_b})|_{x_n} w_n^{b-r} \end{pmatrix} \begin{pmatrix} Q_{b,0} \\ \vdots \\ Q_{b,l_b} \end{pmatrix} = 0. \quad (2)$$

By grouping together equations for a fixed r, we can equivalently write:

$$\sum_{b=r}^{l} \binom{b}{r} \begin{pmatrix} D_x^{(0)}(1)|_{x_1} w_1^{b-r} & \cdots & D_x^{(0)}(x^{l_b})|_{x_1} w_1^{b-r} \\ \vdots & & \vdots \\ D_x^{(0)}(1)|_{x_n} w_n^{b-r} & \cdots & D_x^{(0)}(x^{l_b})|_{x_n} w_n^{b-r} \\ D_x^{(1)}(1)|_{x_1} w_1^{b-r} & \cdots & D_x^{(1)}(x^{l_b})|_{x_1} w_1^{b-r} \\ \vdots & & \vdots \\ D_x^{(1)}(1)|_{x_n} w_n^{b-r} & \cdots & D_x^{(1)}(x^{l_b})|_{x_n} w_n^{b-r} \\ \vdots & & \vdots \\ D_x^{(s-1-r)}(1)|_{x_1} w_1^{b-r} & \cdots & D_x^{(s-1-r)}(x^{l_b})|_{x_1} w_1^{b-r} \\ \vdots & & \vdots \\ D_x^{(s-1-r)}(1)|_{x_n} w_n^{b-r} & \cdots & D_x^{(s-1-r)}(x^{l_b})|_{x_n} w_n^{b-r} \end{pmatrix} \begin{pmatrix} Q_{b,0} \\ \vdots \\ Q_{b,l_b} \end{pmatrix} = 0, \quad (3)$$

with $0 \leq r \leq s-1$.

We will now investigate matrices of the type occurring in equation (3). For convenience we therefore first give the following definitions.

$$A_v(m) := \begin{pmatrix} D_x^{(0)}(1)|_{x_1} & \cdots & D_x^{(0)}(x^m)|_{x_1} \\ \vdots & & \vdots \\ D_x^{(0)}(1)|_{x_n} & \cdots & D_x^{(0)}(x^m)|_{x_n} \\ D_x^{(1)}(1)|_{x_1} & \cdots & D_x^{(1)}(x^m)|_{x_1} \\ \vdots & & \vdots \\ D_x^{(1)}(1)|_{x_n} & \cdots & D_x^{(1)}(x^m)|_{x_n} \\ \vdots & & \vdots \\ D_x^{(v-1)}(1)|_{x_1} & \cdots & D_x^{(v-1)}(x^m)|_{x_1} \\ \vdots & & \vdots \\ D_x^{(v-1)}(1)|_{x_n} & \cdots & D_x^{(v-1)}(x^m)|_{x_n} \end{pmatrix}.$$

The matrix $A_v(m)$ has dimension $vn \times (m+1)$. For future convenience we also define

$$\Delta_{x,v}^{(j)} := \sum_{\alpha \geq 0} \binom{v-1+\alpha}{\alpha} D_x^{(j-\alpha(q-1))},$$

where we interpret $D_x^{(j-\alpha(q-1))}$ as the zero function if $j < \alpha(q-1)$. Finally we also define the matrix

$$B_v(\mu) := \begin{pmatrix} \Delta_{x,v}^{(v-1)}(x^v)|_{\mathcal{P}} & \Delta_{x,v}^{(v-2)}(x^v)|_{\mathcal{P}} & \cdots & \Delta_{x,v}^{(0)}(x^v)|_{\mathcal{P}} \\ \vdots & \vdots & & \vdots \\ \Delta_{x,v}^{(v-1)}(x^{\mu+s})|_{\mathcal{P}} & \Delta_{x,v}^{(v-2)}(x^{\mu+v})|_{\mathcal{P}} & \cdots & \Delta_{x,v}^{(0)}(x^{\mu+v})|_{\mathcal{P}} \end{pmatrix}.$$

This is an $(\mu+1) \times vn$ matrix.

Proposition 2.1. *Let v, μ, m be natural numbers such that $\mu, m \leq vn-1$. Then the matrix $A_v(m)$ has rank $m+1$. The matrix $B_v(\mu)$ has rank $\mu+1$. If $m < vn-1$, then the columns of $A_v(m)$ form a basis of the null space of $B_v(vn-m-2)$.*

Proof. We will investigate the matrix $(m_{ij}) = B_v(vn-1)A_v(vn-1)$. We

have

$$m_{ij} = \sum_{k=1}^{v} \Delta_{x,v}^{(v-k)}(x^{v+i-1})|_{\mathcal{P}} \cdot D_x^{(k-1)}(x^{j-1})|_{\mathcal{P}},$$

where $\cdot$ here denotes the inner product. Using the Leibniz rule we obtain

$$\begin{aligned} m_{ij} &= \textstyle\sum_{e=1}^{n} \Delta_{x,v}^{(v-1)}(x^{v+i+j-2})|_{x_e} \\ &= \textstyle\sum_{\alpha \geq 0} \binom{v-1+\alpha}{\alpha} \binom{v+i+j-2}{v-1-\alpha(q-1)} \sum_{e=1}^{n} x_e^{i+j-1+\alpha(q-1)}. \end{aligned} \tag{4}$$

Our first claim is that $m_{ij} = 0$ for any i and j satisfying $i + j \leq vn$. Indeed if $n = q - 1$ does not divide $i + j - 1$, then $\sum_{e=1}^{n} x_e^{i+j-1+\alpha(q-1)} = 0$. On the other hand, if n divides $i + j - 1$ and $i + j - 1 < vn$, then $\sum_{e=1}^{n} x_e^{i+j-1+\alpha(q-1)} = -1$ for all α, and the claim follows from equation (4) and Lemma 2.1.

Again using equation (4) and Lemma 2.1, we see that if $i + j - 1 = vn$, one has

$$m_{ij} = \sum_{\alpha \geq 0} \binom{v-1+\alpha}{\alpha} \binom{v+vn-1}{v-1-\alpha(q-1)} \sum_{e=1}^{n} 1 \equiv (-1)^v \pmod{p}.$$

The above implies that $A_v(vn-1)B_v(vn-1)$ is a regular matrix. Therefore the same holds for the matrices $A_v(vn-1)$ and $B_v(vn-1)$. Note that the matrix $A_v(m)$ (resp. $B_v(\mu)$) can be obtained from $A_v(vn-1)$ (resp. $B_v(vn-1)$) by deleting some columns (resp. rows). Therefore the matrices $A_v(m)$ and $B_v(\mu)$ have full rank.

Since $m_{ij} = 0$ if $i + j \leq vn$, we see that $B_v(\mu)A_s(m) = 0$ as long as $m + \mu < vn - 1$. Using the above calculated ranks of the matrices $A_v(m)$ and $B_v(\mu)$, we also see that for any $m < vn - 1$, the columns of $A_v(m)$ form a basis of the null space of $B_v(vn - m - 2)$. □

The following lemma is needed to conclude the proof of the previous proposition.

Lemma 2.1. *Let $q = p^e$ and p a prime. Let j, v be positive integers satisfying $0 < j < v$. Then*

$$\sum_{\alpha \geq 0} \binom{v-1+\alpha}{\alpha} \binom{v-1+j(q-1)}{v-1-\alpha(q-1)} \equiv 0 \pmod{p}.$$

On the other hand

$$\sum_{\alpha\geq 0}\binom{v-1+\alpha}{\alpha}\binom{v-1+v(q-1)}{v-1-\alpha(q-1)}\equiv(-1)^{v-1}\pmod{p}.$$

Proof. First we define the differential form

$$\omega_{j,v}:=\frac{(T)^{v-1+j(q-1)}}{(T-T^q)^v}dT.$$

For $a\in\mathbb{F}_q\cup\{\infty\}$, we denote by $\mathrm{Res}_a(\omega_{j,v})$ the residue of $\omega_{j,v}$ at a. Since $j>0$, we have $\mathrm{Res}_0(\omega_{j,v})=0$. A direct calculation using the power series expansion

$$\frac{1}{(1-t)^v}=\sum_{\alpha\geq 0}\binom{v-1+\alpha}{\alpha}t^\alpha$$

shows that for $a\in\mathbb{F}_q\backslash\{0\}$ one has

$$\mathrm{Res}_a(\omega_{j,v})=\sum_{\alpha\geq 0}\binom{v-1+\alpha}{\alpha}\binom{v-1+j(q-1)}{v-1-\alpha(q-1)}.$$

To calculate the residue at infinity, note that

$$\omega_{j,v}=-\frac{S^{(q-1)(v-j)-1}}{(S^{q-1}-1)^v}dS,$$

with $S=1/T$. Therefore one has

$$\mathrm{Res}_\infty(\omega_{j,v})=\begin{cases}0 & \text{if } 1\leq j<v,\\ (-1)^{v-1} & \text{if } j=v.\end{cases}$$

Using the residue theorem (see for example [11, Cor. IV.3.3]), the lemma follows. □

We now explain how the system of equations (3) can be simplified. For a fixed r, we have $(s-r)n$ equations. If $r=0$, we multiply this subsystem from the left with the matrix $B_s(sn-l_0-3)$. If r is between 1 and $s-1$, we multiply this subsystem from the left with the (regular) matrix $B_{s-r}((s-r)n-1)$. In the first place note that any solution to this "multiplied" system gives rise to a unique solution of system (3), since the columns of $A_s(l_0)$ form a basis for the null space of $B_s(sn-l_0-3)$ by Proposition 2.1 and the matrices $B_{s-r}((s-r)n-1)$ are regular. In this sense the systems are equivalent. This way of reasoning is the same as in [8] p. 58-59. Note that the "multiplied" system has (l_0+1) variables and equations less, thus making the determination of a Q-polynomial less complex. Moreover, as we will

see in a moment, the new system has more structure and can be described using syndromes. To this end, we will need the following syndromes (with $\lambda \geq 0$ and $e \geq 0$):

$$S_\lambda^{(e)}(w) := \sum_{i=1}^{n} x_i^\lambda w_i^e$$

and the following linear combinations of them (with $r \leq b \leq l$):

$$\Sigma_{\lambda,r}^{(b-r)}(w) := \sum_{\alpha \geq 0} \binom{s-r-1+\alpha}{\alpha} \binom{s-r-1+\lambda}{s-r-1-\alpha(q-1)} S_\lambda^{(b-r)}(w).$$

We usually omit the w in the notation and simply write $S_\lambda^{(e)}$ and $\Sigma_{\lambda,r}^{(b-r)}$.

Proposition 2.2. *The system of linear equations (3) is equivalent to the following systems of linear equations:*

$$\sum_{b=1}^{l} \begin{pmatrix} \Sigma_{1,0}^{(b)} & \Sigma_{2,0}^{(b)} & \cdots & \Sigma_{l_b+1,0}^{(b)} \\ \Sigma_{2,0}^{(b)} & & & \vdots \\ \vdots & & & \\ \Sigma_{sn-l_0-1,0}^{(b)} & \cdots & & \Sigma_{sn+l_b-l_0-1,0}^{(b)} \end{pmatrix} \begin{pmatrix} Q_{b,0} \\ \vdots \\ Q_{b,l_b} \end{pmatrix} = 0, \qquad (5)$$

and (with $1 \leq r \leq s-1$)

$$\sum_{b=r}^{l} \binom{b}{r} \begin{pmatrix} \Sigma_{1,r}^{(b-r)} & \Sigma_{2,r}^{(b-r)} & \cdots & \Sigma_{l_b+1,r}^{(b-r)} \\ \Sigma_{2,r}^{(b-r)} & & & \vdots \\ \vdots & & & \\ \Sigma_{(s-r)n,r}^{(b-r)} & \cdots & & \Sigma_{(s-r)n+l_b,r}^{(b-r)} \end{pmatrix} \begin{pmatrix} Q_{b,0} \\ \vdots \\ Q_{b,l_b} \end{pmatrix} = 0, \qquad (6)$$

Proof. Fix some $b \geq 0$ and some r between 0 and $s-1$. For convenience we denote by $A_{s-r}^{(b-r)}(l_b)$ the matrix occurring as the $(b-r+1)$-th summand in equation (3). Following the remark just before this proposition, we only need to investigate the product of the matrices $B_{s-r}((s-r)n-1)A_{s-r}^{(b-r)}(l_b)$

We now consider the matrix $(m_{ij}^{(b-r)}) = B_{s-r}((s-r)n-1)A_{s-r}^{(b-r)}(l_b)$. It is an $((s-r)n) \times (l_b+1)$ matrix and we have

$$m_{ij}^{(b-r)} = \sum_{k=1}^{s-r} \sum_{e=1}^{n} \Delta_{x,s-r}^{(s-r-k)}(x^{s-r+i-1})|_{x_e} D_x^{(k-1)}(x^{j-1})|_{x_e} w_e^{b-r},$$

which implies by the Leibniz rule that

$$m_{ij}^{(b-r)} = \sum_{e=1}^{n} \Delta_{x,s-r}^{(s-r-1)}(x^{s-r+i+j-2})|_{x_e} w_e^{b-r} = \Sigma_{i+j-1,r}^{(b-r)}.$$

Here we used that for any $\alpha \geq 0$ we have $S_\lambda^{(b)} = S_{\lambda+\alpha(q-1)}^{(b)}$. □

Note that all matrices occurring in this proposition are Hankel matrices. This allows one to solve the system in a faster way than using Gaussian elimination as we will see in the next section.

3. One-point codes from the Hermitian curve

In this section we will deduce a syndrome form of the Gurswami-Sudan algorithm for one-point codes from the Hermitian curve defined over $\mathbb{F}_{q^2}$, the finite field with q^2 elements. Here q is a power of a prime number p. The Hermitian curve has q^3+1 rational points and genus $g := q(q-1)/2$. Let $\mathcal{P} = \{P_1, \ldots, P_{q^3}\}$ be the collection of the $n := q^3$ rational points different from the point at infinity P_∞. We consider codes of the form $C_L(\mathcal{P}, (k+g-1)P_\infty)$. This is an $[n, \geq k]$ code defined over $\mathbb{F}_{q^2}$. The precise parameters of such codes are well known (see [9,10,12]). In the remainder of this section we will write $\mathbb{F}$ for the finite field $\mathbb{F}_{q^2}$.

The linear equations that appear in the Guruswami-Sudan algorithm involve Hasse derivatives, just like in the case of RS-codes. Therefore, the first task is to compute Hasse derivatives in the Hermitian function field. This function field can be described as $\mathbb{F}(x,y)$, with $y^q + y = x^{q+1}$. We will now give a procedure to compute the quantity $D_x^{(n)}(f)$ for any natural number n and any $f \in \mathbb{F}[x,y] \subset \mathbb{F}(x,y)$.

The Leibniz rule also holds for a Hasse derivative in a function field (see for example [3, Section 1.3]). It will be convenient to describe this Leibniz rule in a different way. For $f \in \mathbb{F}[x,y]$ we define the formal sum

$$G_f(T) := \sum_{n \geq 0} D_x^{(n)}(f) T^n.$$

The Leibniz rule then implies that for any $f, g \in \mathbb{F}[x,y]$ one has

$$G_f(T) G_g(T) = G_{fg}(T).$$

In particular one has $G_{f^p}(T) = G_f(T)^p$. Another way to formulate this is the following

$$D_x^{(n)}(f^p) = \begin{cases} 0 & \text{if } p \text{ does not divide } n, \\ D_x^{(n/p)}(f)^p & \text{otherwise.} \end{cases} \tag{7}$$

Since the Hasse derivative $D_x^{(n)}(f)$ is linear, it suffices to calculate it in case f is a monomial in x and y. By the Leibniz rule, it suffices to calculate $D_x^{(n)}(f)$ in case f is a power of x (which is easy) or a power of y (which deserves some attention). Applying the Leibniz rule again, we see that the crucial point is to calculate $D_x^{(n)}(y)$ for any natural number n. The following lemma addresses this.

Lemma 3.1. *Let $\mathbb{F}(x, y)$ be the Hermitian function field defined by the equation $y^q + y = x^{q+1}$. Then one has*

$$D_x^{(n)}(y) = \begin{cases} y & \text{if } n = 0, \\ x^q & \text{if } n = 1, \\ (-1)^i (x^{q^2} - x)^{q^{i-1}} & \text{if } n = q^i, \text{ with } i \geq 1, \\ (-1)^i & \text{if } n = (q+1)q^i, \text{ with } i \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. For $n = 0$, the statement is clear, so from now on we suppose that $n > 0$. By the defining equation of the Hermitian curve, we know that

$$D_x^{(n)}(y) = D_x^{(n)}(x^{q+1}) - D_x^{(n)}(y^q) = \binom{q+1}{n} x^{q+1-n} - D_x^{(n)}(y^q).$$

Applying equation (7) iteratively, we see that if q does not divide n, then $D_x^{(n)}(y^q) = 0$, while otherwise $D_x^{(n)}(y^q) = D_x^{(n/q)}(y)^q$. Since $\binom{q+1}{n} = \binom{q}{n-1} + \binom{q}{n}$, we deduce that if $n \notin \{1, q, q+1\}$, we have $D_x^{(n)}(x^{q+1}) = 0$, while otherwise $D_x^{(n)}(x^{q+1}) = x^{q+1-n}$. Putting all this together, the result follows. □

Now we return to list decoding. We denote by $L(mP_\infty)$ the $\mathbb{F}$-linear space of functions on the Hermitian curve having a pole of order at most m at P_∞ and no poles at other places. It is well known that a basis of this space is given by $\{x^i y^j \mid qi + (q+1)j \leq m, 0 \leq i, 0 \leq j < q\}$. Some care has to be taken to find a basis of the codes themselves, because some functions can evaluate to the same codeword (for example the functions x and x^{q^2}). A basis of the codes can be found by evaluating all monomials in the set

$\mathcal{B}_1 := \{x^i y^j \,|\, qi + (q+1)j \leq m, 0 \leq i < q^2, 0 \leq j < q\}$. For list decoding we also need to consider the following more general sets of monomials:

$$\mathcal{B}_v := \{x^i y^j \,|\, 0 \leq i < vq^2, 0 \leq j < q\}.$$

It is convenient to order the set $\mathcal{B}_v$ of monomials as $f_0, f_1, \ldots, f_{vq^3-1}$ in such a way that $v_{P_\infty}(f_i) > v_{P_\infty}(f_j)$ if $i < j$. The mapping $f_k \mapsto x^{vq^2-1}y^{q-1}/f_k$ is a bijection on $\mathcal{B}_v$ that reverses this ordering. This implies that $x^{vq^2-1}y^{q-1}/f_k = f_{vq^3-1-k}$ or in other words that

$$f_k \cdot f_{vq^3-1-k} = f_{vq^3-1}. \tag{8}$$

Note that it is not true for all i and j that $f_i \cdot f_j = f_{i+j}$.

In the Guruswami-Sudan list decoding algorithm for the code $C_L(\mathcal{P}, (k+g-1)P_\infty)$ one looks for a suitable polynomial $Q(x, y, z) = \sum_{b=0}^{l} \sum_{a=0}^{l_b} f_a z^b Q_{b,a} \in \mathbb{F}[x, y][z]$. The Hermitian code analogue of equation (1) is the following: for a monomial $x^i y^j$, we define $\rho(x^i y^j) = iq + j(q+1)$. Let $s \geq 1$ be a natural number and τ the number of errors one wishes to correct. For any $1 \leq i \leq n$, any $0 \leq t \leq s-1$, and $0 \leq r \leq t$ the coefficients $Q_{b,a}$ should satisfy the following system of linear equations (where l_b satisfies the inequality $\rho(f_{l_b}) \leq s(n-\tau) - b(k+g-1)$:

$$\sum_{b=0}^{l} D_z^{(r)}(z^b)|_{w_i} \left(D_x^{(t-r)}(f_0)|_{P_i} \cdots D_x^{(t-r)}(f_{l_b})|_{P_i} \right) \begin{pmatrix} Q_{b,0} \\ \vdots \\ Q_{b,l_b} \end{pmatrix} = 0. \tag{9}$$

Here $f_a \in \mathcal{B}_s$. We can again group all equations together in a structured

way similar as in equation (3). One then obtains the following system:

$$\sum_{b=r}^{l} \binom{b}{r} \begin{pmatrix} D_x^{(0)}(f_0)|_{P_1} w_1^{b-r} & \cdots & D_x^{(0)}(f_{l_b})|_{P_1} w_1^{b-r} \\ \vdots & & \vdots \\ D_x^{(0)}(f_0)|_{P_n} w_n^{b-r} & \cdots & D_x^{(0)}(f_{l_b})|_{P_n} w_n^{b-r} \\ D_x^{(1)}(f_0)|_{P_1} w_1^{b-r} & \cdots & D_x^{(1)}(f_{l_b})|_{P_1} w_1^{b-r} \\ \vdots & & \vdots \\ D_x^{(1)}(f_0)|_{P_n} w_n^{b-r} & \cdots & D_x^{(1)}(f_{l_b})|_{P_n} w_n^{b-r} \\ \vdots & & \vdots \\ D_x^{(s-1)}(f_0)|_{P_1} w_1^{b-r} & \cdots & D_x^{(s-1)}(f_{l_b})|_{P_1} w_1^{b-r} \\ \vdots & & \vdots \\ D_x^{(s-1)}(f_0)|_{P_n} w_n^{b-r} & \cdots & D_x^{(s-1)}(f_{l_b})|_{P_n} w_n^{b-r} \end{pmatrix} \begin{pmatrix} Q_{b,0} \\ \vdots \\ Q_{b,l_b} \end{pmatrix} = 0, \tag{10}$$

with $0 \leq r \leq s-1$. The similarity with the systems obtained in case of RS-codes is obvious. Following this similarity further, we now define

$$A_v^H(m) := \begin{pmatrix} D_x^{(0)}(f_0)|_{P_1} & \cdots & D_x^{(0)}(f_m)|_{P_1} \\ \vdots & & \vdots \\ D_x^{(0)}(f_0)|_{P_n} & \cdots & D_x^{(0)}(f_m)|_{P_n} \\ D_x^{(1)}(f_0)|_{P_1} & \cdots & D_x^{(1)}(f_m)|_{P_1} \\ \vdots & & \vdots \\ D_x^{(1)}(f_0)|_{P_n} & \cdots & D_x^{(1)}(f_m)|_{P_n} \\ \vdots & & \vdots \\ D_x^{(v-1)}(f_0)|_{P_1} & \cdots & D_x^{(v-1)}(f_m)|_{P_1} \\ \vdots & & \vdots \\ D_x^{(v-1)}(f_0)|_{P_n} & \cdots & D_x^{(v-1)}(f_m)|_{P_n} \end{pmatrix}$$

Now we describe the analogue of the matrix $B_v(\mu)$.

$$B_v^H(\mu) := \begin{pmatrix} \Delta_{x,v}^{(v-1)}(f_0)|_{\mathcal{P}} & \Delta_{x,v}^{(v-2)}(f_0)|_{\mathcal{P}} & \cdots & \Delta_{x,v}^{(0)}(f_0)|_{\mathcal{P}} \\ \vdots & \vdots & & \vdots \\ \Delta_{x,v}^{(v-1)}(f_\mu)|_{\mathcal{P}} & \Delta_{x,v}^{(v-2)}(f_\mu)|_{\mathcal{P}} & \cdots & \Delta_{x,v}^{(0)}(f_\mu)|_{\mathcal{P}} \end{pmatrix},$$

where

$$\Delta_{x,v}^{(j)}(f) := \sum_{\alpha \geq 0} \binom{v-1+\alpha}{\alpha} D_x^{(j-\alpha(q^2-1))}(f).$$

We have the following analogue of Proposition 2.1:

Proposition 3.1. *Let v, μ, m be natural numbers such that $\mu, m \leq vn-1$. Then the matrix $A_v^H(m)$ has rank $m+1$. The matrix $B_v^H(\mu)$ has rank $\mu+1$. If $m < vn-1$, then the columns of $A_v^H(m)$ form a basis of the null space of $B_v^H(vn-m-2)$.*

Proof. We investigate the matrix $(m_{ij}) = B_v^H(vn-1)A_v^H(vn-1)$. We have

$$m_{ij} = \sum_{k=1}^{v} \Delta_{x,v}^{(v-k)}(f_{i-1})|_{\mathcal{P}} \cdot D_x^{(k-1)}(f_{j-1})|_{\mathcal{P}},$$

where $\cdot$ here denotes the inner product. Using the Leibniz rule we obtain

$$m_{ij} = \sum_{e=1}^{n} \Delta_{x,v}^{(v-1)}(f_{i-1}f_{j-1})|_{P_e}.$$

Note that $\Delta_{x,v}^{(v-1)}(f_{i-1}f_{j-1})|_{P_e}$ is the coefficient of $(x-x(P_e))^{v-1}$ in the power series expansion of $f_{i-1}f_{j-1}/(1-(x-x(P_e))^{q^2-1})^v$ in the parameter $x-x(P_e)$. Therefore, if one defines

$$\omega_{i,j,v} := \frac{f_{i-1}f_{j-1}}{(x-x^{q^2})^v} dx,$$

then one has $\Delta_{x,v}^{(v-1)}(f_{i-1}f_{j-1})|_{P_e} = \mathrm{Res}_{P_e}(\omega_{i,j,v})$. Therefore, using the residue theorem, we can write

$$m_{ij} = \sum_{e=1}^{n} \mathrm{Res}_{P_e}(\omega_{i,j,v}) = -\mathrm{Res}_{P_\infty}(\omega_{i,j,v}). \tag{11}$$

We will show that the differential $\omega_{i,j,v}$ is regular at P_∞ if $i+j-2 < vq^3-1$ and that it has a pole of order one (and hence a non-zero residue)

if $i+j-2 = vq^3-1$. A local parameter at P_∞ is given by x/y. One has $dx = -y^{2-q}d(x/y)$ and

$$\omega_{i,j,v} = \frac{-f_{i-1}f_{j-1}}{y^{q-2}(x-x^{q^2})^v} d\left(\frac{x}{y}\right).$$

If $i+j-2 < vq^3-1$, then $v_{P_\infty}(f_{i-1}f_{j-1}) > v_{P_\infty}(f_{i-1}f_{vq^3-i}) = v_{P_\infty}(f_{vq^3-1}) = -(vq^3+q^2-q-1)$. Here we used the ordering of the f_k and equation (8). On the other hand $v_{P_\infty}(y^{q-2}(x-x^{q^2})^v) = -(vq^3+q^2-q-2)$. Therefore $v_{P_\infty}(\omega_{i,j,v}) \geq 0$ which implies by equation (11) that $m_{ij} = 0$. Now suppose that $i+j-2 = vq^3-1$. Similarly as above we now deduce that $v_{P_\infty}(\omega_{i,j,v}) = -1$. In fact (using equation (8)) one has

$$\omega_{i,vq^3-i+1,v} = \frac{-(x/y)^{-1}}{(1/x^{q^2-1}-1)^v} d\left(\frac{x}{y}\right).$$

Therefore, using equation (11), we see that $m_{i,vq^3-i+1} = (-1)^v$.

We can now reason along similar lines as in the proof of Proposition 2.1. □

Using the matrices $B_v^H(vn-1)$, we can again reformulate the Guruswami-Sudan list decoding algorithm in terms of syndromes. We again describe the syndromes and the linear combinations of syndromes that occur in this set up (with $0 \leq i \leq q^3-1$ and $e \geq 0$):

$$S_i^{H,(e)}(w) := \sum_{m=1}^{n} f_i(P_m) w_m^e$$

and (with $i,j > 0$ and $0 \leq r \leq s-1$):

$$\Sigma_{i,j,r}^{H,(e)}(w) := \sum_{m=1}^{n} \Delta_{x,s-r}^{s-r-1}(f_{i-1}f_{j-1})|_{P_m} w_m^e.$$

Again we will in the following omit the w in the notation. Note that the expression $\Sigma_{i,j,r}^{H,(e)}$ can be written as a linear combination of syndromes of the form $S_i^{H,(e)}$ using Lemma 3.1, the Leibniz rule for Hasse derivatives, and the defining equation of the Hermitian curve. Similarly to Proposition 2.2 we now obtain the following system of linear equations that is equivalent to system (10):

Proposition 3.2. *The system of linear equations (10) is equivalent to the following systems of linear equations:*

$$\sum_{b=1}^{l} \begin{pmatrix} \Sigma_{1,1,0}^{H,(b)} & \Sigma_{1,2,0}^{H,(b)} & \cdots & \Sigma_{1,l_b+1,0}^{H,(b)} \\ \Sigma_{2,1,0}^{H,(b)} & & & \vdots \\ \vdots & & & \\ \Sigma_{sn-l_0-1,1,0}^{H,(b)} & \cdots & & \Sigma_{sn-l_0-1,l_b-1,0}^{H,(b)} \end{pmatrix} \begin{pmatrix} Q_{b,0} \\ \vdots \\ Q_{b,l_b} \end{pmatrix} = 0, \qquad (12)$$

and (with $1 \leq r \leq s-1$*)*

$$\sum_{b=r}^{l} \binom{b}{r} \begin{pmatrix} \Sigma_{1,1,r}^{H,(b-r)} & \Sigma_{1,2,r}^{H,(b-r)} & \cdots & \Sigma_{1,l_b+1,r}^{H,(b-r)} \\ \Sigma_{2,1,r}^{H,(b-r)} & & & \vdots \\ \vdots & & & \\ \Sigma_{(s-r)n,1,r}^{H,(b-r)} & \cdots & & \Sigma_{(s-r)n,l_b+1,r}^{H,(b-r)} \end{pmatrix} \begin{pmatrix} Q_{b,0} \\ \vdots \\ Q_{b,l_b} \end{pmatrix} = 0, \qquad (13)$$

We omit the proof, since it is very similar to that of Proposition 2.2. The matrices occurring in the above proposition are no longer Hankel, but still exhibit a more structure than the matrices we had in equation (10). Moreover, as was also the case for Reed-Solomon codes, we now have $l_0 + 1$ equations and variables less to deal with.

4. Determination of $Q(x, y)$

If we solve the equations (5) and (6) using standard gaussian elimination the complexity is $O(n^3)$. However due to the special structure of the involved matrices i it is possible to use a variant of an algorithm of Sakata to reduce the complexity [6]. To see this we rewrite the equations as

$$\sum_{b=1}^{l} \sum_{i=1}^{l_b+1} \Sigma_{i+j}^{(b)} Q_{b,i-1} = 0 \qquad (14)$$

for all $j, 0 \leq j \leq sn - l_0 - 2$

and

$$\sum_{b=1}^{l}\sum_{i=1}^{l_b+1}\binom{b}{r}\Sigma_{i+j,r}^{(b-r)}Q_{b,i-1}=0 \tag{15}$$

for all j, $0 \leq j \leq (s-r)n-1$ and all r $1 \leq r \leq s-1$

This is the situation treated by S.Sakata (with proper renaming) so his algorithm finds a polynomial $Q(x,y)$ with minimal weighted degree. The complexity of this is $O(n^2)$ which is slighty more than the complexity of the method proposed by Aleknovich ([1]) but the difference does not appear to be significant for practical lengths. S. Sakata also applies the Berlekamp-Massey-Sakata algorithm ([13]) for Guruswami-Sudan list decoding of Hermitian codes, however this is not based on the syndromes.

References

1. M. Alekhnovich, Linear diophantine equations over polyomials and soft decoding of Reed-Solomon codes, IEEE Trans. Inform. Theory, vol. 51, pp. 2257–2265, July 2005.
2. G. Schmidt, V. Sidorenko, and M. Bossert, Decoding Reed-Solomon codes beyond half the minimum distance using shift-register synthesis, ISIT 2006, Seattle, USA, July 9–14, pp. 459–463, 2006.
3. D.M. Goldschmidt, Algebraic functions and projective curves, Springer, Berlin, 2003.
4. V. Guruswami and M. Sudan, Improved decoding of Reed-Solomon and algebraic-geometric codes, IEEE Trans. Inform. Theory, vol. 45, pp. 1757–1767, Sept. 1999.
5. R.M. Roth and G. Ruckenstein, Efficient decoding of Reed-Solomon codes beyond half the minimum distance, IEEE Trans. Inform. Theory, vol. 46, pp. 246–257, Jan. 2000.
6. S. Sakata, Finding a minimal polynomial vector of a vector of nD arrays, *Applied Algebra, Algebraic Algorithms and Error-correcting Codes, Proceedings of AAECC-9*, New Orleans, USA:Lecture Notes in Computer Science, 539, Springer Verlag, pp. 414-425, 1991.
7. M. Sudan, Decoding of Reed-Solomon codes beyond the error-correcting bound, J. Compl., vol. 13, pp. 180–193, 1997.
8. T. Høholdt and J. Justesen, A course in error-correcting codes, European Mathematical Society, Zürich, 2004.
9. T. Høholdt, J.H. van Lint and R. Pellikaan, Volume I, Chapter 10 of Handbook of coding theory, V.S. Pless and W.C. Huffman (eds.), Elsevier, Amsterdam, 1998.
10. P.V. Kumar and K. Yang, On the true minimum distance of Hermitian codes, AGCT-3, Lecture Notes in Math. vol. 1518, Springer, Berlin (1992), 99–107.
11. H. Stichtenoth, Algebraic function fields and codes, Springer, Berlin, 1993.
12. H. Stichtenoth, A note on Hermitian codes over GF(q^2), IEEE Trans. Inf. Theory. **34** (1988), 1345–1348.

13. S. Sakata, "On fast interpolation method for Guruswami-Sudan list decoding of one-point algebraic-geometry codes", in *Applied algebra, algebraic algorithms and error-correcting codes*, ser. LNCS, Springer, 2001, vol. 2227, pp.172-181.

A note on the tensor rank of the multiplication in certain finite fields

Stéphane Ballet

Laboratoire de Géométrie Algébrique
et Applications à la Théorie de l'Information
Université de la Polynésie Française
BP 6570, 98702 Faa'a, Tahiti, Polynésie Française
E-mail : ballet@upf.pf

In this paper, we obtain some new bounds of the tensor rank of the multiplication in any extension of non quadratic finite fields $\mathbb{F}_q$ where $q > 5$, by using the descent over $\mathbb{F}_q$ of certain towers of algebraic function fields defined over $\mathbb{F}_{q^2}$, constructed by Garcia-Stichtenoth and Garcia-Stichtenoth-Rück.

Keywords: Tensor rank, Bilinear complexity, finite fields, algebraic function fields.

1. Introduction

1.1. *Tensor rank of multiplication*

Let $\mathbb{F}_q$ be a finite field with q elements where q is a prime power and let $\mathbb{F}_{q^n}$ be a $\mathbb{F}_q$ extension of degree n. We denote by m the ordinary multiplication in the $\mathbb{F}_q$-vector space $\mathbb{F}_{q^n}$. The multiplication m is a bilinear map from $\mathbb{F}_{q^n} \times \mathbb{F}_{q^n}$ into $\mathbb{F}_{q^n}$, thus it corresponds to a linear map M from the tensor product $\mathbb{F}_{q^n} \bigotimes \mathbb{F}_{q^n}$ over $\mathbb{F}_q$ into $\mathbb{F}_{q^n}$. One can also represent M by a tensor $t_M \in \mathbb{F}_{q^n}^* \bigotimes \mathbb{F}_{q^n}^* \bigotimes \mathbb{F}_{q^n}$ where $\mathbb{F}_{q^n}^*$ denotes the dual of $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$. Hence the product of two elements x and y of $\mathbb{F}_{q^n}$ is the convolution of this tensor with $x \otimes y \in \mathbb{F}_{q^n} \bigotimes \mathbb{F}_{q^n}$. If

$$t_M = \sum_{l=1}^{\lambda} a_l \otimes b_l \otimes c_l \tag{1}$$

where $a_l \in \mathbb{F}_{q^n}^*$, $b_l \in \mathbb{F}_{q^n}^*$, $c_l \in \mathbb{F}_{q^n}$, then

$$x.y = \sum_{l=1}^{\lambda} a_l(x) b_l(y) c_l. \tag{2}$$

Every expression (2) is called a bilinear multiplication algorithm $\mathcal{U}$. The integer λ is called the multiplicative complexity $\mu(\mathcal{U})$ of $\mathcal{U}$.

Let us set

$$\mu_q(n) = \min_{\mathcal{U}} \mu(\mathcal{U}),$$

where $\mathcal{U}$ is running over all bilinear multiplication algorithms in $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$.

Then $\mu_q(n)$ corresponds to the minimum possible number of summands in any tensor decomposition of type (1), which is the rank of the tensor of the multiplication in $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$. The tensor rank $\mu_q(n)$ is also called the bilinear complexity of multiplication in $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$.

1.2. *Notations*

Let $F/\mathbb{F}_q$ be an algebraic function field of one variable of genus g, with constant field $\mathbb{F}_q$, associated to a curve X defined over $\mathbb{F}_q$. For any place P we define F_P to be the residue class field of P and O_P its valuation ring. If $\mathcal{D}$ is a divisor then $\mathcal{L}(\mathcal{D}) = \{f \in F^\star, \mathcal{D} + (f) \geq 0\} \cup \{0\}$ is a vector space over $\mathbb{F}_q$ whose dimension $dim\ \mathcal{D}$ is given by the Riemann-Roch theorem. The degree of a divisor $\mathcal{D} = \sum_P a_P P$ is defined by $deg\ \mathcal{D} = \sum_P a_P deg\ P$ where $deg\ P$ is the dimension of F_P over $\mathbb{F}_q$. The order of a divisor $\mathcal{D} = \sum_P a_P P$ in P is the number a_P denoted by $ord_P\mathcal{D}$. The support of a divisor $\mathcal{D}$ is the set $supp\ \mathcal{D}$ of the places P such that $ord_P\mathcal{D} \neq 0$.

1.3. *Known results*

The bilinear complexity $\mu_q(n)$ of multiplication in the n-degree extension of a finite field $\mathbb{F}_q$ with q elements is known for certain values of n. In particular, S. Winograd [17] and H. de Groote [12] have shown that this complexity is $\geq 2n-1$, with equality holding if and only if $n \leq \frac{1}{2}q + 1$. Using the principle of the D.V. and G.V. Chudnovsky algorithm [11] applied to elliptic curves, M.A. Shokrollahi has shown in [15] that the bilinear complexity of multiplication is equal to $2n$ for $\frac{1}{2}q + 1 < n < \frac{1}{2}(q + 1 + \epsilon(q))$ where ϵ is the function defined by:

$$\epsilon(q) = \begin{cases} \text{greatest integer} \leq 2\sqrt{q} \text{ prime to q}, & \text{if q is not a perfect square} \\ 2\sqrt{q}, & \text{if q is a perfect square.} \end{cases}$$

Moreover, M.A. Shokrollahi and U. Baum in [9] has succeded to construct effective optimal algorithms of type Chudnovsky in the elliptic case.

Recently in [1], [2], [5], [6], [7], [8] and [4] the study made by M.A. Shokrollahi has been generalized to algebraic function fields of genus g.

In particular, from the existence of towers of algebraic functions fields satisfying good properties, it was proved:

Theorem 1.1. *Let $q = p^r$ a power of the prime p. The bilinear complexity $\mu_q(n)$ of multiplication in any finite field $\mathbb{F}_{q^n}$ is linear with respect to the extension degree, more precisely:*

$$\mu_q(n) \leq C_q n$$

where C_q is the constant defined by:

$$C_q = \begin{cases} \textit{if } q = 2 & \textit{then} \quad 36 & \textit{[4]} \\ \textit{else if } q = p \geq 5 & \textit{then} \quad 3(1 + \frac{4}{q-3}) & \textit{[8]} \\ \textit{else if } q = p^{2m} \geq 9 & \textit{then} \quad 2(1 + \frac{2}{\sqrt{q}-2}) & \textit{[4]} \\ \textit{else if } q \geq 16 & \textit{then} \quad 3(1 + \frac{2p}{q-3}) & \textit{[5],[6], and [7]} \\ \textit{else if } q \geq 3 & \textit{then} \quad 6(1 + \frac{2}{q-2}) & \textit{[4].} \end{cases}$$

In order to obtain these good estimates for the constant C_q, the author has given in [1] some easy to verify conditions allowing the use of the D.V and G.V Chudnovsky algorithm. Then Ballet and Rolland [5] have improved the algorithm using places of degree 1 and 2. Let us set the last version of the theorem which in particular describes the basis of the multiplication algorithm:

Theorem 1.2. *Let*

- *$F/\mathbb{F}_q$ be an algebraic function field,*
- *Q be a degree n place of $F/\mathbb{F}_q$,*
- *$\mathcal{D}$ be a divisor of $F/\mathbb{F}_q$,*
- *$\mathcal{P} = \{P_1, ..., P_{N_1}, Q_1, ..., Q_{N_2}\}$ be a set of places of degree 1 and 2.*

We suppose that Q, $P_1, \cdots, P_{N_1}, Q_1, ..., Q_{N_2}$ are not in the support of $\mathcal{D}$ and that:

a) the application

$$Ev_Q : \mathcal{L}(\mathcal{D}) \to \mathbb{F}_{q^n} \simeq F_Q$$

is onto,

b) the application

$$Ev_{\mathcal{P}} : \begin{cases} \mathcal{L}(2\mathcal{D}) \to \mathbb{F}_q^{N_1} \times \mathbb{F}_{q^2}^{N_2} \\ f \qquad\quad \mapsto (f(P_1), ..., f(P_{N_1}), f(Q_1), ..., f(Q_{N_2})) \end{cases}$$

is injective.
Then

$$\mu_q(n) \leq N_1 + 3N_2.$$

Let us remark that the algorithm given in [11] by D.V. and G.V. Chudnovsky is the case $N_2 = 0$. The generalization introduced here is usefull. Indeed, we know good towers of function fields, with many rational points, over $\mathbb{F}_{q^2}$ and not over $\mathbb{F}_q$. So, if we want to obtain good results for the multiplication over $\mathbb{F}_q$ we need to interpolate on places of degree 1 and also on places of degree 2. From the results of [1] and the previous algorithm, we obtain (cf. [1], [5]):

Theorem 1.3. *Let q be a prime power and let n be an integer > 1. Let $F/\mathbb{F}_q$ be an algebraic function field of genus g and N_k a number of places of degree k in $F/\mathbb{F}_q$. If $F/\mathbb{F}_q$ is such that $N_n > 0$ (or $2g + 1 \leq q^{\frac{n-1}{2}}(q^{\frac{1}{2}} - 1)$) then:*

1) if $N_1 > 2n + 2g - 2$, then

$$\mu_q(n) \leq 2n + g - 1,$$

2) if there exists a non-special divisor of degree $g - 1$ and $N_1 + 2N_2 > 2n + 2g - 2$, then

$$\mu_q(n) \leq 3n + 3g,$$

3) if $N_1 + 2N_2 > 2n + 4g - 2$, then

$$\mu_q(n) \leq 3n + 6g.$$

Moreover, we have a very interesting particular case of Theorem 1.2 for $q \geq 4$ obtained in [3]:

Theorem 1.4. *Let q be a prime power such that $q \geq 4$ and let n be an integer > 1. Let $F/\mathbb{F}_q$ be an algebraic function field of genus g and N_k a number of places of degree k in $F/\mathbb{F}_q$. If $F/\mathbb{F}_q$ is such that $N_n > 0$ (or $2g + 1 \leq q^{\frac{n-1}{2}}(q^{\frac{1}{2}} - 1)$) then:*

1) if $N_1 > 2n + 2g - 2$, *then*

$$\mu_q(n) \leq 2n + g - 1,$$

2) if $N_1 + 2N_2 > 2n + 2g - 2$, *then*

$$\mu_q(n) \leq 3n + 3g.$$

Applied on a Garcia-Stichtenoth tower [13] and a Garcia-Stichtenoth-Rück, these theorems give good estimates of the constant C_q for any integer n.

However, in all the cases we can see that Theorems 1.3 and 1.4 need many places to be applied whereas only a certain number of places of degree one are really used. In fact the conditions on the number of places of degree one in Theorem 1.3 (and 1.4) seem too constraint with respect to what we can expect. These conditions are due to a certain way to construct the algorithm 1.2. Let us recall the idea used to obtain Theorem 1.3 (and 1.4). In fact, Theorem 1.3 is obtained firstly by the existence of a non-special divisor $\mathcal{D}$ of degree $g - 1$ such that the first evaluation map Ev_Q is valid with respect to Theorem 1.2. Then, it is sufficient to have a set of places $\mathcal{P} = \{P_1, ..., P_{N_1}, Q_1, ..., Q_{N_2}\}$ such that $2\mathcal{D} - (P_1 + ... + P_{N_1} + Q_1 + ... + Q_{N_2})$ is trivially non-special (i.e of negative degree) to apply finally Theorem 1.2.

By using a new way to obtain the conditions a) and b) in Theorem 1.2, under few supplementary assumptions corresponding to the simultaneous existence of two non-special divisors of degree $g - 1$ with a certain form, the author obtains in [3] the following result:

Theorem 1.5. *Let* q *be a prime power and let* n *be an integer* > 1. *Let* $F/\mathbb{F}_q$ *be an algebraic function field of genus* g *and* N_k *a number of places of degree* k *in* $F/\mathbb{F}_q$. *Let* $\mathcal{P} = \{P_1, ..., P_{N_1}, Q_1, ..., Q_{N_2}\}$ *be a set of places of degree* 1 *and* 2. *If there exists a divisor* $\mathcal{D}'$ *such that the divisor* $\mathcal{D}' - (P_1 + ... + P_{N_1} + Q_1 + ... + Q_{N_2})$ *is non-special of degree* $g - 1$ *and if there exist a divisor* $\mathcal{D}$ *of degree* $n + g - 1$ *such that* $2\mathcal{D} \subseteq \mathcal{D}'$ *and a place* Q *of degree* n *such that the divisor* $\mathcal{D} - Q$ *is non-special of degree* $g - 1$ *then:*

1) if $N_1 \geq 2n + g - 1$, *then*

$$\mu_q(n) \leq 2n + g - 1,$$

2) if $N_1 + 2N_2 \geq 2n + g - 1$, *then*

$$\mu_q(n) \leq 3n + 3g,$$

and moreover if $N_1 + 2N_2 \leq 2n + g$ *then*

$$\mu_q(n) \leq 3\left(n + \frac{g}{2}\right).$$

In fact, this last theorem replaces constraints on the number of places of degree one or two by constraints on the divisors. It enables us to obtain better results on the bilinear complexity of multiplication in the cases where these divisors exist which is only proven for $g = 0, g = 1$ [10], for a particular example of genus 2 (cf. [3]), and recently in the general case where $q > 5$ given by the following result [4]:

Theorem 1.6. *Let* $q > 5$ *be a prime power and let* n *be an integer* > 1*. Let* $F/\mathbb{F}_q$ *be an algebraic function field of genus* g *and* N_k *a number of places of degree* k *in* $F/\mathbb{F}_q$*. Let* $\mathcal{P} = \{P_1, ..., P_{N_1}, Q_1, ..., Q_{N_2}\}$ *be a set of places of degree* 1 *and* 2*. If* $F/\mathbb{F}_q$ *is such that* $N_n > 0$ *(or* $2g + 1 \leq q^{\frac{n-1}{2}}(q^{\frac{1}{2}} - 1)$*) then:*

1) if $N_1 \geq 2n + g - 1$*, then*

$$\mu_q(n) \leq 2n + g - 1,$$

2) if $N_1 + 2N_2 \geq 2n + g - 1$*, then*

$$\mu_q(n) \leq 3\left(n + \frac{g}{2}\right).$$

1.4. *New results established in this paper*

In Section 2, we recall the definitions and the properties of the towers of algebraic function fields obtained from Garcia-Stichtenoth towers [13] and from Garcia-Stichtenoth-Rck towers [14], studied in [2], [5], [6] and [8]. Then we deduce new bounds of the tensor rank of the multiplication in any extension of non-quadratic finite fields $\mathbb{F}_q$ with $q > 5$ by using the preceding towers.

2. New bounds of the tensor rank

In this section, we recall some towers of algebraic function fields. Theorem 1.6, applied to the algebraic function fields of these towers, enables us to obtain new bounds of the tensor rank of the multiplication in certain finite fields.

For any algebraic function field $F/\mathbb{F}_q$ defined over the finite field $\mathbb{F}_q$, we denote by $g(F/\mathbb{F}_q)$ the genus of $F/\mathbb{F}_q$ and by $N_k(F/\mathbb{F}_q)$ the number of places of degree k in $F/\mathbb{F}_q$.

2.1. *Upper bounds*

2.1.1. *Garcia-Stichtenoth tower of Artin-Schreier algebraic function field extensions*

We present now a modified Garcia-Stichtenoth's tower (cf. [13], [2], [5]) having good properties. Let us consider a finite field $\mathbb{F}_{q^2}$ with $q = p^r$ and r an integer. Let us consider the Garcia-Stichtenoth's elementary abelian tower T_1 over $\mathbb{F}_{q^2}$ constructed in [13] and defined by the sequence $(F_1, F_2, ...)$ where

$$F_{k+1} := F_k(z_{k+1})$$

and z_{k+1} satisfies the equation :

$$z_{k+1}^q + z_{k+1} = x_k^{q+1}$$

with

$$x_k := z_k/x_{k-1} \ in \ F_k(for \ k \geq 1).$$

Moreover $F_1 := \mathbb{F}_{q^2}(x_0)$ is the rational function field over $\mathbb{F}_{q^2}$ and F_2 the Hermitian function field over $\mathbb{F}_{q^2}$. Let us denote by g_k the genus of F_k in T_1. If $r > 1$, we consider the completed Garcia-Stichtenoth tower

$$T_2 = F_{1,0} \subseteq F_{1,1} \subseteq ... \subseteq F_{1,r} \subseteq F_{1,0} \subseteq F_{1,1} \subseteq ... \subseteq F_{1,r}...$$

considered in [2] such that $F_k \subseteq F_{k,s} \subseteq F_{k+1}$ for any integer s such that $s = 0, ..., r$, with $F_{k,0} = F_k$ and $F_{k,r} = F_{k+1}$. Let us denote by $g_{k,s}$ the genus of $F_{k,s}$ in T_2 and by $N_{k,s}$ the number of places of degree one of $F_{k,s}$ in T_2. Recall that each extension $F_{k,s}/F_k$ is Galois of degree p^s wih full constant field $\mathbb{F}_{q^2}$. Now, if $r > 1$ we consider the tower studied in [5] and [6]

$$T_3 = G_{0,0} \subseteq G_{0,1} \subseteq ... \subseteq G_{0,r} \subseteq G_{1,0} \subseteq G_{1,1} \subseteq ... \subseteq G_{1,r}, ...$$

defined over the constant fied $\mathbb{F}_q$ and related to the tower T_2 by

$$F_{k,s} = \mathbb{F}_{q^2} G_{k,s} \quad for \ all \ k \ and \ s,$$

namely $F_{k,s}/\mathbb{F}_{q^2}$ is the constant field extension of $G_{k,s}/\mathbb{F}_q$. By [5] and [6], the tower T_3 is well defined and one has:

Proposition 2.1. *Let $q = p^r$. For any integer $k \geq 1$, for any integer s such that $s = 0, ..., r$, the algebraic function field $G_{k,s}/\mathbb{F}_q$ has a genus $g(G_{k,s}) = g_{k,s}$ with $N_1(G_{k,s})$ places of degree one and $N_2(G_{k,s})$ places of degree two such that:*

1) $G_k \subseteq G_{k,s} \subseteq G_{k+1}$ *with* $G_{k,0} = G_k$ *and* $G_{k,r} = G_{k+1}$.

2) $g(G_{k,s}) \leq \frac{g(G_{k+1})}{p^{r-s}} + 1$ *with* $g(G_{k+1}) = g_{k+1} \leq q^{k+1} + q^k$.

3) $N_1(G_{k,s}) + 2N_2(G_{k,s}) \geq (q^2 - 1)q^{k-1}p^s$.

Consequently, we obtain the following result:

Theorem 2.1. *Let* $q = p^r$ *be a prime power and* r *an odd integer such that* $q > 5$. *Then, we have for any integer* n*:*

$$\mu_q(n) \leq 3n\left(1 + \frac{p}{q-2}\right)$$

Proof. Let us set $M_{k,s} = N_1(G_{k,s}) + 2N_2(G_{k,s})$. For any integer n, let k and s be the smallest integers such that $2n \leq M_{k,s} - g_{k,s} + 1$, then $2n > M_{k,s-1} - g_{k,s-1} + 1$. For any integer $k \geq 3$ and for any integer $s = 1, ..., r$, we have $g_k \leq q^k + q^{k-1}$ by Theorem 3.1 in [5] and $g_{k,s} \leq \frac{g_{k+1}}{p^{r-s}} + 1$ by Theorem 3.1 in [5]. Hence, we have $g_{k,s-1} \leq q^{k-1}p^{s-1}(q+1) + 1$. Moreover, we have $M_{k,s} \geq (q^2 - 1)q^{k-1}p^s$ by Theorem 3.2 in [5], then we obtain the following inequality $2n > (q^2 - q - 2)p^{s-1}q^{k-1}$. Thus, we have $g_{k,s} \leq \frac{2np}{q-2}$ and so $\mu_q(n) \leq 3(n + \frac{g_{k,s}}{2}) \leq 3n(1 + \frac{p}{q-2})$ by Theorem 1.6. □

2.1.2. *Garcia-Stichtenoth-Ruck tower of Kummer function field extensions*

In this section we present a Garcia-Stichtenoth-Ruck's tower (cf. [8]) having good properties. Let $\mathbb{F}_q$ be a finite field of characteristic $p \geq 3$. Let us consider the towers $T_1/\mathbb{F}_{p^2}$ and $T_2/\mathbb{F}_p$ respectively defined over $\mathbb{F}_{p^2}$ and $\mathbb{F}_p$, which are defined recursively by the same following equation, studied in [14]:

$$y^2 = \frac{x^2 + 1}{2x}.$$

The towers $T_i/\mathbb{F}_q$ with $\mathbb{F}_q = \mathbb{F}_{p^2}$ if $i = 1$ and $\mathbb{F}_q = \mathbb{F}_p$ if $i = 2$ are represented by the sequences of function fields

$$T_i/\mathbb{F}_q = (M_{i,0}, M_{i,1}, M_{i,2}, ...)$$

where $M_{i,n} = \mathbb{F}_q(x_0, x_1, ..., x_n)$ and $x_{k+1}^2 = (x_k^2 + 1)/2x_k$ holds for each $k \geq 0$. Note that $M_{i,0}$ is the rational function field over $\mathbb{F}_q$.

For any prime number $p \geq 3$, let us remark that the tower $T_1/\mathbb{F}_{p^2}$ is asymptotically optimal over the field $\mathbb{F}_{p^2}$, i.e. $T_1/\mathbb{F}_{p^2}$ reaches the Drinfeld-Vladut bound. Moreover, note that for any integer k, $M_{1,k}/\mathbb{F}_{p^2}$ in $T_1/\mathbb{F}_{p^2}$ is the constant field extension of $M_{2,k}/\mathbb{F}_p$ in $T_2/\mathbb{F}_p$. Hence, let us denote by g_k the genus of $M_{1,k}/\mathbb{F}_{p^2}$ and $M_{2,k}/\mathbb{F}_p$.

Theorem 2.2. *Let p be a prime number such that $p > 5$. Then, we have for any integer n:*

$$\mu_p(n) \leq 3n\left(1 + \frac{2}{p-2}\right)$$

Proof. Let us set $M_k = N_1(M_{2,k}/\mathbb{F}_p) + 2N_2(M_{2,k}/\mathbb{F}_p)$. For any integer n, there exists a smallest integer k such that $2n \leq M_{1,k} - g_k + 1$. Then $2n > M_{k-1} - g_{k-1} + 1$. On the other hand for any integer k, we have $M_k - g_k + 1 \geq 2^{k+1}(p-2) + 1$ by Theorem 2.1 in [8]. Then we obtain $n \geq 2^{k-1}(p-2)$, hence $g_k = 2^{k+1} \leq \frac{4n}{p-2}$. Otherwise, from Theorem 1.6, we have: $\mu_p(n) \leq 3\left(n + \frac{g(M_{2,k}/\mathbb{F}_p)}{2}\right) \leq 3\left(n + \frac{g_k}{2}\right) \leq 3n\left(1 + \frac{2}{p-2}\right)$, and the proof is complete. □

2.2. *Asymptotical upper bounds*

From the asymptotical point of view, let us recall that Shparlinski, Tsfasman, Vladut have given in [16] many interesting remarks on the algorithm of D.V. and G.V. Chudnovsky. In particular, they have obtained asymptotic bounds for the bilinear complexity by considering:

$$M_q = \limsup_{k\to\infty} \frac{\mu_q(k)}{k}$$

and

$$m_q = \liminf_{k\to\infty} \frac{\mu_q(k)}{k}.$$

Let us summarize their estimates given in [16]:

1) $q = 2$

$$3.52 \leq m_2 \leq 35/6.$$

$$M_2 \leq 27.$$

2) $q \geq 9$ is a square

$$2\left(1 + \frac{1}{q-1}\right) \leq m_q \leq 2\left(1 + \frac{1}{\sqrt{q}-2}\right).$$

$$M_q \leq 2\left(1+\frac{1}{\sqrt{q}-2}\right).$$

3) $q > 2$

$$2\left(1+\frac{1}{q-1}\right) \leq m_q \leq 3\left(1+\frac{1}{q-2}\right).$$

$$M_q \leq 6\left(1+\frac{1}{q-2}\right).$$

Moreover, from certain values of the constant C_q, the constant M_q of Shparlinski, Tsfasman, and Vladut was sensitively improved in certain cases:

for a prime $p > 5$, $M_p \leq 3\left(1+\frac{4}{p-3}\right)$ by [8]

for a prime power $q \geq 16$, $M_q \leq 3\left(1+\frac{2p}{q-3}\right)$ by [5], [6] and [7].

However, the distance between m_q and M_q is still important.

In fact, from the results of the previous section, it is easy to see that we can improve the asymptotical results:

Proposition 2.2. *Let $q = p^m$ be an odd power of a prime such that $q > 5$. Then, we have:*

$$M_p \leq 3\left(1+\frac{2}{p-2}\right)$$

and

$$M_q \leq 3\left(1+\frac{p}{q-2}\right)$$

Proof. It follows directly from Theorems 2.1 and 2.2. □

References

1. **S. Ballet.** *Curves with Many Points and Multiplication Complexity in Any Extension of $\mathbb{F}_q$.* Finite Fields and Their Applications, 5 (1999), 364-377.
2. **S. Ballet.** *Low Increasing Tower of Algebraic Function Fields and Bilinear Complexity of Multiplication in Any Extension of $\mathbb{F}_q$.* Finite Fields and Their Applications, 9 (2003), 472-478.
3. **S. Ballet.** *An improvement of the construction of the D.V. and G.V. Chudnovsky algorithm for multiplication in finite fields.* Theoretical Computer Science, 352 (2006), 293-305.

4. **S. Ballet.** *On the tensor rank of the multiplication in the finite fields.* submitted.
5. **S. Ballet, R. Rolland.** *Multiplication Algorithm in a Finite Field and Tensor Rank of the Multiplication.* Journal of Algebra, vol 272/1, (2004), 173-185.
6. **S. Ballet, D. Le Brigand, R. Rolland.** *The definition field of a tower of function fields and applications.* Arithmetic, Geometry and Coding Theory 2005 (AGCT 10), Société Mathématique de France, Séminaires et Congrés, to appear.
7. **S. Ballet, D. Le Brigand.** *On the existence of non-special divisors of degree g and $g-1$ in algebraic function fields over $\mathbb{F}_q$.* Journal of Number Theory 116 (2006), 293-310.
8. **S. Ballet, J. Chaumine.** *On the bounds of the bilinear complexity of multiplication in some finite fields.* Applicable Algebra in Engineering, Communication and Computing, 15 (2004), 205-211.
9. **Baum, U., Shokrollahi, M.A.** *An Optimal Algorithm for Multiplication in $\mathbb{F}_{256}/\mathbb{F}_4$.* Applicable Algebra in Engineering, Communication and Computing, 2 (1991), 15-20.
10. **J. Chaumine.** *On the bilinear complexity of multiplication in small finite fields.* C.R. Acad. Sci. Paris, Théorie des nombres, Ser I 343 (2006)
11. **D.V. Chudnovsky, G.V. Chudnovsky.** *Algebraic Complexities and Algebraic Curves over Finite Fields.* J. Complexity, 4 (1988), 285-316.
12. **H.F. De Groote.** *Characterization of Division Algebras of Minimal Rank and the Structure of their Algorithm Varieties.* SIAM J. Comp. 12,No.1 (1983), 101-117.
13. **A. Garcia, H. Stichtenoth.** *A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vladut bound.* Inventiones Mathematicae, 121 (1995), 211-222.
14. **A. Garcia, H. Stichtenoth, H.-G. Ruck** *On tame towers over finite fields.* Journal fur die reine und angewandte Mathematik, 557 (2003), 53-80.
15. **M.A. Shokrollahi.** *Optimal Algorithms for Multiplication in Certain Finite Fields using Algebraic Curves.* SIAM J. Comp. 21, No.6 (1992), 1193-1198.
16. **I.E. Shparlinski, M.A. Tsfsaman, S.G. Vladut.** *Curves with Many Points and Multiplication in Finite Fields.* Lectures Notes in Mathematics, 1518 (1992), 145-169, Springer-Verlag, Berlin.
17. **S. Winograd.** *On Multiplication in Algebraic Extension Fields.* Theoretical Computer Science, 8 (1979), 359-377.

Multiplication in small finite fields using elliptic curves

Jean Chaumine

Laboratoire de Géométrie Algébrique et
Applications à la Théorie de l'Information,
Université de la Polynésie française,
B.P. 6570, 98702 Faa'a, Tahiti, Polynésie française
E-Mail : jean.chaumine@upf.pf

In this paper, we improve a result of Shokrollahi who has applied the algorithm of D. V. Chudnovsky and G. V. Chudnovsky to algebraic curves of genus one. More precisely, from the abelian group structure of the set of rational points on elliptic curves, we show that, if the degre n of the extension attains the upper bound of the range given by Shokrollahi, then the bilinear complexity of the multiplication in all extensions $\mathbb{F}_{q^n}$ is still equal to $2n$. This light improvement permits the use of the elliptic curve method for practical cases not covered by the Shokrollahi results.

Keywords: Bilinear complexity, Finite field, Algebraic function field, Elliptic curve.

1. Introduction

1.1. *Bilinear complexity of the multiplication*

Let $\mathbb{F}_q$ be a finite field with q elements where $q = p^r$ is a prime power and let $\mathbb{F}_{q^n}$ be a $\mathbb{F}_q$ extension of degree n. The bilinear complexity of the multiplication in $\mathbb{F}_{q^n}$ over $\mathbb{F}_q$, denoted by $\mu_q(n)$, is the tensor rank of the multiplication ([13], [1]). It corresponds to the least possible number of summands in any tensor decomposition.

1.2. *Known results*

The bilinear complexity $\mu_q(n)$ of the multiplication in $\mathbb{F}_{q^n}$ is known for certain values of n. Winograd in 1979 [18] and De Groote in 1983 [9] have shown that this bilinear complexity is $\geq 2n-1$, with equality holding if and only if $n \leq \frac{1}{2}q + 1$. In 1988, by interpolating on algebraic curves, Chudnovsky and Chudnovsky [8] have succeeded in obtaining a principle

of construction of fast multiplication algorithms and they were the first to show the linearity of the bilinear complexity. In 1992, by applying the algorithm of Chudnovsky-Chudnovsky to well-fitted elliptic curves, Shokrollahi [11] has shown the following theorem:

Theorem 1.1. *The bilinear complexity $\mu_q(n)$ of the multiplication in $\mathbb{F}_{q^n}$ is equal to $2n$ if $\frac{1}{2}q+1 < n < \frac{1}{2}(q+1+\epsilon(q))$ where ϵ is the function defined by:*

$$\epsilon(q) = \begin{cases} \text{the greatest integer} \leq 2\sqrt{q} \text{ and prime to } q, \text{ if } q \text{ is not} \\ \quad \text{a perfect square.} \\ 2\sqrt{q}, \text{ if } q \text{ is a perfect square.} \end{cases}$$

Ballet [1], [2] and Rolland [3] have generalized the study of Shokrollahi to algebraic function fields of arbitrary genus g and they have obtained new bounds for the bilinear complexity of the multiplication in $\mathbb{F}_{q^n}$ by using towers of function fields.

In the paper, we show that in Theorem 1.1, the right strict inequality can be replaced by a large one. For example, if we take $q = 8$ and $n = 7$, the method applies, and we can design a practical algorithm to multiply in $\mathbb{F}_{8^7}$ using an elliptic curve instead of an hyperelliptic curve.

1.3. *Definitions - Notations*

Let us recall some definitions of algebraic geometry using the notations of [15]. Let $F/\mathbb{F}_q$ be an algebraic function field of one variable over $\mathbb{F}_q$ of genus g. We denote by $\mathrm{Div}(F/\mathbb{F}_q)$ the divisor group of the algebraic function field $F/\mathbb{F}_q$, i.e. the free abelian group generated by the places of $F/\mathbb{F}_q$ over $\mathbb{F}_q$. We denote by $\mathcal{P}_F$ the set of principal divisors of $F/\mathbb{F}_q$ over $\mathbb{F}_q$ and $\mathrm{Div}^0(F/\mathbb{F}_q)$ the set of divisors of $F/\mathbb{F}_q$ of degree 0 which are both subgroups of $\mathrm{Div}(F/\mathbb{F}_q)$. Let $\mathrm{Div}^n(F/\mathbb{F}_q)$ be the set of divisors of $F/\mathbb{F}_q$ of degree n and let $[D]$ be the class of a divisor D modulo $\mathcal{P}_F$. For all places P, we denote by F_P the residue class field of P, which is a finite extension of $\mathbb{F}_q$. So we define the degree of a divisor $D = \sum_P a_P P$ by $\deg D = \sum_P a_P \deg P$, where $\deg P$ is the dimension of F_P over $\mathbb{F}_q$. The order of a divisor $D = \sum_P a_P P$ in P is the number a_P, denoted by $\mathrm{ord}_P D$. We call $\mathcal{L}(D) = \{f \in F, D + (f) \geq 0\} \cup \{0\}$ the vector space over $\mathbb{F}_q$ whose dimension $l(D)$ is given by the Riemann-Roch theorem. In particular, in elliptic case, we have [11]:

Theorem 1.2. *Let $E/\mathbb{F}_q$ be an elliptic function field and let $D \in$*

$\mathrm{Div}(E/\mathbb{F}_q)$ *be a divisor. Then, we have:*

$$l(D) = \begin{cases} 0 & \textit{if } \deg D < 0, \\ 1 & \textit{if } D \in \mathcal{P}_E, \\ \deg D & \textit{otherwise.} \end{cases}$$

If E is an elliptic curve, we denote by $\mathrm{Pic}^0(E)$ the finite group of degree zero divisor classes modulo $\mathcal{P}_E$. We define the map σ (as in [14], p.66) by:

$$\sigma : \begin{cases} \mathrm{Div}^0(E) \to E \\ D^0 \qquad\quad \mapsto P \end{cases}$$

such that $D^0 \sim (P) - (O)$, where O denotes the point at infinity of E. It is clear that σ is a map (for all divisors $D^0 \in \mathrm{Div}^0(E)$, P is unique) which is surjective and if $D_1^0, D_2^0 \in \mathrm{Div}^0(E)$, then $\sigma(D_1^0) = \sigma(D_2^0)$ if and only if $D_1^0 \sim D_2^0$. Thus σ induces a bijection of sets (also denoted by σ) $\sigma : \mathrm{Pic}^0(E) \to E$. We can notice that the geometric group law on E and the group law induced over $\mathrm{Pic}^0(E)$ by using σ are the same. If we denote by:

$$\eta : \begin{cases} E \to \mathrm{Pic}^0(E) \\ P \mapsto \textit{class of } (P) - (O) \end{cases}$$

the inverse map of σ, for all $P, Q \in E$, we have:

$$\eta(P + Q) = \eta(P) + \eta(Q),$$

where the first $+$ is the addition on E, while the second is the addition of divisor classes in $\mathrm{Pic}^0(E)$.

Let R be a divisor of degree n. It is clear that:

$$\tau_{-R} : \begin{cases} \mathrm{Div}^n(E) \to \mathrm{Div}^0(E) \\ D \qquad\quad \mapsto D - R \end{cases}$$

is a bijection. Then, let us consider the map:

$$\phi = \sigma \circ \tau_{-R}.$$

For $D \in \mathrm{Div}^n(E)$ there exists a unique place $P \in E$ such that:

$$D - R \sim (P) - (O).$$

The map ϕ is surjective. If $\phi(D) = \phi(D')$, then $D - R \sim (P) - (O)$ and $D' - R \sim (P) - (O)$, hence $D - R \sim D' - R$ and $D \sim D'$. Conversely, if $D \sim D'$, there exists $f \in \mathbb{F}_q(E)$ such that $D' = D + (f)$. If we denote by $\phi(D) = P$, then we have $D - R \sim (P) - (O)$. There exists $g \in \mathbb{F}_q(E)$ such that $D - R = (P) - (O) + (g)$, and $D' - R = D - R + (f) = (P) - (O) + (fg)$. Hence, $\phi(D') = P = \phi(D)$. This implies that ϕ induces a bijection of sets from $\mathrm{Pic}^n(E) = \mathrm{Div}^n(E)/\mathcal{P}_E$ over E.

2. Improvement of a theorem of Shokrollahi

In [11], M. A. Shokrollahi has applied the algorithm of D. V. Chudnovsky and G. V. Chudnovsky to elliptic curves and he has shown that the rank of the finite field $\mathbb{F}_{q^n}$ is equal to $2n$ if n satisfies the inequality $\frac{1}{2}q + 1 < n < \frac{1}{2}(q + 1 + \epsilon(q))$. In this paper, we enlarge the range of n with rank of $\mathbb{F}_{q^n}$ equal to $2n$. First, by using the abelian group structure of the set of rational points on elliptic curve, we obtain the following proposition:

Proposition 2.1. *Let q be a prime power. Let n be an integer satisfying the inequality $3 \leq n \leq \frac{1}{2}(q + 1 + \epsilon(q))$ where $\epsilon(q)$ is defined by:*

$$\epsilon(q) = \begin{cases} 2\sqrt{q} \text{ if } q \text{ is a perfect square,} \\ \text{the greatest integer } \leq 2\sqrt{q} \text{ and prime to } q \text{ otherwise.} \end{cases}$$

There exists an elliptic function field $E/\mathbb{F}_q$ containing k rational places $P_0, \ldots, P_{k-1}$ over $\mathbb{F}_q$ with $k \geq 2n$, a place Q of degree n and a divisor $\mathcal{D}$ of degree n such that:

(1) $[\mathcal{D}] \neq [Q]$,
(2) $2\mathcal{D} - (P_{i_0} + \cdots + P_{i_{2n-1}})$ *is not principal,*
(3) $\mathrm{ord}_{P_i}(\mathcal{D}) = 0$ *for* $i = 0, \ldots, k-1$.

This proposition is justified by the two following lemmas:

Lemma 2.1. *Let $q \geq 4$ be a prime power. Let $n \geq 3$ be an integer. Then all elliptic function field over $\mathbb{F}_q$ has at least one place of degree n.*

Proof. In [8], D. V. Chudnovsky and G. V. Chudnovsky have shown that the number N_n of places of degree n on algebraic function field of genus g over $\mathbb{F}_q$ satisfy the inequality:

$$N_n \geq \frac{1}{n}(q^n - 2gq^{n/2} - q^{n/2+1} - 2gq^{n/4+1/2}).$$

As $n > 2$, this inequality becomes:

$$N_n > \frac{1}{n}q^{n/2}(q^{n/2} - q - 4g).$$

Moreover, the genus of elliptic function field being $g = 1$, it is sufficient to prove that for $q \geq 4, q^{n/2} - q \geq 4$, i.e. $q(q^{n/2-1} - 1) \geq 4$. This last inequality is true because we have $n \geq 3$ and $q \geq 4$. □

Lemma 2.2. *Let q be a prime power. Let E be an elliptic curve over $\mathbb{F}_q$. Let P_j be a $\mathbb{F}_q$-rational point of E and $+$ be the group law on E, described in [14]. Then the equation $P+P=P_j$ has at most four solutions in the set of rational points $E(\mathbb{F}_q)$.*

Proof. First, we can notice that $E(\mathbb{F}_q)$ is a subgroup of E. The multiplication map by 2 from $E(\mathbb{F}_q)$ into $E(\mathbb{F}_q)$ dfined by : $P \mapsto 2P$ is a group morphism which kernel has at most four elements [17], [16]. And we get the result. □

Then we can prove Proposition 2.1.

Proof. For $q=2$ the condition on n is not satisfied and there is nothing to prove. First, let us show the existence of an elliptic function field containing at least $2n$ rational places and a place Q of degree n. For $q=3$, $n=3$ the elliptic function field defined by the equation $y^2=x^3+x^2+x+1$ contains six rational places and four places of degree 3. For $q \geq 4$, from the results of Waterhouse [17] and Vlăduţ [16], there exists an elliptic function field over $\mathbb{F}_q$ containing $k=q+1+\epsilon(q)$ rational places and from Lemma 2.1 there is at least one place of degree n. Now, let us show the existence of a divisor $\mathcal{D}$ of degree n satisfying the assumptions of Proposition 2.1. Let us calculate the class of $P_{i_0}+\cdots+P_{i_{2n-1}}$:

$$[P_{i_0}+\cdots+P_{i_{2n-1}}] = \eta(P_{i_0}+\cdots+P_{i_{2n-1}}) = \eta(P_j) = [P_j].$$

Moreover, let Q be a place of degree n. From Lemma 2.2, the equation $P+P=P_j$ has at most four solutions. Then there exists at least one place P_i such that:

$$[P_i] \neq [Q] \text{ and } P_i + P_i \neq P_j$$

because $n \geq 3$ and $|E(\mathbb{F}_q)| = k \geq 6$. As ϕ is a bijection, there exists a divisor D of degree n such that $[D]=[P_i]$. From [10], Lecture 14, Lemma 1, $[D]$ contains a divisor $\mathcal{D}$ such that $\mathrm{ord}_P(\mathcal{D})=0$ for all places P of degree 1. And the proof is complete. □

From the existence of an elliptic function field satisfying the assumptions of Proposition 2.1, we obtain the following theorem:

Theorem 2.1. *Let q be a prime power and n be an integer such that:*

$$\frac{1}{2}q + 1 < n \leq \frac{1}{2}(q + 1 + \epsilon(q)),$$

where $\epsilon(q)$ is defined by:

$$\epsilon(q) = \begin{cases} 2\sqrt{q} \text{ if } q \text{ is a perfect square,} \\ \text{the greatest integer } \leq 2\sqrt{q} \text{ and prime to } q \text{ otherwise.} \end{cases}$$

Then the bilinear complexity of the multiplication in $\mathbb{F}_{q^n}$ satisfies:

$$\mu_q(n) = 2n.$$

Proof. In [11], M. A. Shokrollahi has shown that for all integer n satisfying $\frac{1}{2}q + 1 < n < \frac{1}{2}(q + 1 + \epsilon(q))$, $\mu_q(n) = 2n$. Then let us consider the case $2n = q + 1 + \epsilon(q)$. From Proposition 2.1, there exists an elliptic function field over $\mathbb{F}_q$ containing $k \geq 2n$ rational places $P_0, \ldots, P_{k-1}$, a place Q of degree n and a divisor $\mathcal{D}$ of degree n such that:

(1) $[\mathcal{D}] \neq [Q]$,
(2) $2\mathcal{D} - (P_{i_0} + \cdots + P_{i_{2n-1}})$ is not principal,
(3) $\mathrm{ord}_{P_i}(\mathcal{D}) = 0$ for $i = 0, \ldots, k-1$.

Let us prove that the two evaluation maps ev_Q and ev_G, in the algorithm of Chudnovsky-Chudnovsky [8] and defined above, are isomorphisms:

$$ev_Q : \begin{cases} \mathcal{L}(\mathcal{D}) \to F_Q \simeq \mathbb{F}_{q^n} \\ f \quad\quad \mapsto f(Q) \end{cases}$$

$$ev_G : \begin{cases} \mathcal{L}(2\mathcal{D}) \to \mathbb{F}_q^{2n} \\ f \quad\quad\quad \mapsto (f(P_{i_0}), \ldots, f(P_{i_{2n-1}})) \end{cases}.$$

For $q = 2$, there is nothing to prove. Then, let $q \geq 3$ be a prime power. As Q is a place of degree n, the residue class field F_Q of Q is isomorphic to $\mathbb{F}_{q^n}$. Furthermore, from [10], Lecture 14, Lemma 1, we can suppose that $\mathrm{ord}_Q(\mathcal{D}) = 0$ showing that $\mathcal{L}(\mathcal{D})$ is contained in the valuation ring of Q. Thus, ev_Q is the restriction of the residue class map over $\mathcal{L}(\mathcal{D})$. Then ev_Q defines a vector space homomorphism which kernel is $\mathcal{L}(\mathcal{D}-Q)$. As $\mathcal{D}$ and Q belong to different classes, $\mathcal{D}-Q$ is not principal of degree 0 and the kernel of ev_Q is trivial showing that ev_Q is onto. Moreover, $l(\mathcal{D}) = \deg Q = n$, hence ev_Q is an isomorphism. On the other hand, the evaluation map ev_G is well defined because $\mathrm{ord}_P(2\mathcal{D}) = 0$ for all places P of degree 1. The vector space $\mathcal{L}(2\mathcal{D})$ is contained in the valuation ring of each place of degree 1. Its kernel

is $\mathcal{L}(2\mathcal{D}-(P_{i_0}+\cdots+P_{i_{2n-1}}))$ which is trivial because $2\mathcal{D}-(P_{i_0}+\cdots+P_{i_{2n-1}})$ is not principal. We conclude that ev_G is onto. Moreover, $l(2\mathcal{D}) = 2n$, hence ev_G is also an isomorphism. By combining with the results of Winograd and De Groote, we conclude that $\mu_q(n) = 2n$ and the proof is complete. □

Remark 2.1. For example, if we want to multiply in $\mathbb{F}_{8^7}$, the result of Shokrollahi doesn't allow us to apply the algorithm of Chudnovsky-Chudnovsky to elliptic curve. We can at best use a hyperelliptic curve of genus 2 and we get a bilinear complexity $\mu_8(7) \leq 15$, from [1], [12]. The result obtained in this paper shows the existence of an elliptic function field with bilinear complexity of the multiplication in $\mathbb{F}_{8^7}$ equal to 14. An other example is $q = 32$ and $n = 22$. The rank of the tensor of the multiplication in $\mathbb{F}_{32^{22}}$ is equal to 44.

.

References

1. S. Ballet, Curves with Many Points and Multiplication Complexity in Any Extension of $\mathbb{F}_q$, Finite Fields and Their Applications 5 (1999) 364-377.
2. S. Ballet, Low increasing tower of algebraic function fields and bilinear complexity of multiplication in any extension of $\mathbb{F}_q$, Finite Fields and Their Applications 9 (2003) 472-478.
3. S. Ballet, R. Rolland, Multiplication algorithm in a finite field and tensor rank of the multiplication, Journal of Algebra 272 (1) (2004) 173-185.
4. S. Ballet, J. Chaumine, On the Bounds of the Bilinear Complexity of Multiplication in Some Finite Fields, AAECC 15 (3-4) (2004) 205-211.
5. S. Ballet, D. Le Brigand, On the existence of non-special divisors of degree g and $g-1$ in algebraic function fields over F_q, Journal of Number Theory 116 (2006) 293-310.
6. S. Ballet, D. Le Brigand, R. Rolland, On an application of the definition field descent of a tower of function fields, Colloque International Arithmetic, Geometry and Coding Theory (2005) (AGCT 10), to appear.
7. J. Chaumine, Corps de Fonctions Algébriques et Algorithme de D. V. Chudnovsky et G. V. Chudnovsky pour la Multiplication dans les Corps Finis, PhD, University of French Polynesia (2005).
8. D. V. Chudnovsky, G. V. Chudnovsky, Algebraic Complexities and Algebraic Curves over Finite Fields, Journal of Complexity 4 (1988) 285-316.
9. H. F. De Groote, Characterization of Division Algebras of Minimal Rank and the Structure of their Algorithm Varieties, SIAM J. of Comput. 12 (1) (1983) 101-117.
10. M. Deuring, Lectures on the Theory of Algebraic Functions of One Variable, Lecture Notes in Math. 314, Springer-Verlag, Heidelberg/New York/Tokyo (1973).

11. M. A. Shokrollahi, Optimal Algorithms for Multiplication in certain Finite Fields using Elliptic Curves, SIAM J. of Comput. 21 (6) (1992), 1193-1198.
12. M. A. Shokrollahi, On the Rank of Certain Finite Fields, Comput. Complexity 1 (1991) 157-181.
13. I. E. Shparlinski, M. A. Tsfasman, S. G. Vlăduţ, Curves with Many Points and Multiplication in Finite Fields, Lecture Notes in Mathematics 1518, Springer-Verlag, Berlin (1992) 145-169.
14. J. H. Silverman, The Arithmetic of Elliptic Curves, Graduate Texts in Mathematics 106, Springer-Verlag, New York (1986).
15. H. Stichtenoth, Algebraic Function Fields and Codes, Lecture Notes in Mathematics 314, Springer-Verlag, Berlin/Heidelberg/New York (1993).
16. S. G. Vlăduţ, Cyclicity Statistics for Elliptic Curves over Finite Fields, Finite Fields and Their Applications 5 (1999) 13-25.
17. W. C. Waterhouse, Abelian Varieties over Finite Fields, Ann. Scient. Ec. Norm. Sup., 4^e série, t.2 (1969) 521-560.
18. S. Winograd, On Multiplication in Algebraic Extension Fields, Theoretical Computer Science 8 (1979) 359-377.

An optimal unramified tower of function fields

Kristian Brander

Department of Mathematics
Technical University of Denmark
Kgs. Lyngby, DK 2800, Denmark
E-mail: K.Brander@mat.dtu.dk

Efficient construction of long algebraic geometric–codes resulting from optimal towers of function fields is known to be difficult. In the following a tower which is both optimal and unramified after its third level, is investigated in the hope that its simple ramification structure can be exploited in the construction of algebraic geometric–codes. Results are mostly negative, but help clarifying the difficulties in computing bases of Riemann–Roch spaces.

1. Introduction

The problem of constructing algebraic geometric–codes from optimal towers of function fields has been a topic of great interest since the existence of such towers was first discovered in [7]. There are two major challenges in this. Firstly, one needs to find an optimal tower, and secondly this tower must be explicit and simple enough that one can compute bases of the (one–point) Riemann–Roch spaces involved in the construction of algebraic geometric–codes. In [3] Garcia and Stichtenoth found an optimal tower which can be recursively defined by a single equation. This makes the tower relatively easy to work with, but nonetheless it is still difficult to determine bases of one–point spaces in the tower. In [4,5] algorithms for computing such bases are presented, but these are to resource–intensive to be practical except at the first few levels of the tower. Thus the problem of constructing algebraic geometric–codes from an optimal tower is still not fully resolved.

In [2] Elkies gives defining equations for a family of optimal unramified towers, and this construction coincides with a family of towers considered in [8], in exactly one tower. This tower is the subject of this paper. In the following the properties of the tower are investigated, and an algorithm for computing Riemann–Roch spaces using pole cancellation is developed. The main motivation for considering the tower is the hope that its simple ram-

ification structure might be exploited for efficient construction of algebraic geometric-codes from the tower.

2. An optimal unramified tower

We first define the tower which will occupy the rest of this paper, and as we shall see, it is both optimal and unramified after its third level.

Definition 2.1. Let $K = \mathbb{F}_{2^6}$ and $F_0 = K(x_0)$. We let $F_n = F_{n-1}(x_n)$ denote the n–th extension in the sequence $\mathcal{T} = (F_n)_{n=0}^{\infty}$ of function fields recursively defined by the equations

$$x_n^3 = 1 + \left(\frac{x_{n-1}}{x_{n-1}+1}\right)^3 = \frac{x_{n-1}^2 + x_{n-1} + 1}{(x_{n-1}+1)^3} \quad n \geq 1. \tag{1}$$

Note that $K = \mathbb{F}_{2^6}$ contains the third roots of unity and thus F_n/F_{n-1} is a Kummer extension. Furthermore, all extensions are tame, and this allows us to determine most of the tower's structural properties by standard techniques. But before we do this, we introduce some notation for the rational places in $\mathbb{P}_{F_n}$.

Definition 2.2 (Coordinates and coordinate sequence). *Let P be a place of F_n. We say that $x_i(P)$ is the i–th coordinate of P, and writing*

$$P = P_{a_0 a_1 \ldots a_n}$$

means that P has coordinates $x_i(P) = a_i$, with the convention that $x_i(P) = \infty$ when $v_P(x_i) < 0$. The sequence $a_0, a_1, \ldots, a_n$ is called P's coordinate sequence.

A place is not uniquely determined by its coordinates. Nonetheless we will sometimes abuse the coordinate notation and refer to $P_{a_0,\cdots,a_n}$ as an actual place. When this happens it will be understood that the properties under consideration will hold for *any* place P with coordinates $P_{a_0,\cdots,a_n}$. For an easy way of referring to the place of $K(x_i, x_{i+1}, \cdots, x_j)$ lying below $P_{a_0,a_1,\cdots,a_n}$ we will use the notation

$$P \cap K(x_i, x_{i+1}, \cdots, x_j) = P_{\underbrace{\bullet,\cdots,\bullet}_{i \text{ times}},a_i,a_{i+1},\cdots,a_j,\underbrace{\bullet,\cdots,\bullet}_{n-j \text{ times}}}$$

In the following α will denote a primitive element of $\mathbb{F}_4$. The places $P = P_{a_0,a_1,\cdots,a_n}$ having $a_0 \in \mathbb{F}_4 \cup \{\infty\}$ turn out to be of special importance. By plugging a_i into equation (1) one can determine the possible values of the coordinates that can succeed it, and this information can be compiled into

a directed graph as in Figure 1. One sees that the succession–structure of the coordinates is a closed system in the sense that if a_0 is an element of $\mathbb{F}_4 \cup \{\infty\}$ so are all the a_i's. To get a way of referring to these special places we define

$$\begin{aligned}\mathbb{P}_{F_n|\mathbb{F}_4} &= \{P \in \mathbb{P}_{F_n} \mid x_i(P) \in \mathbb{F}_4 \cup \{\infty\} \text{ for all } i \in \{0,1,\ldots,n\}\} \\ &= \{P \in \mathbb{P}_{F_n} \mid x_i(P) \in \mathbb{F}_4 \cup \{\infty\} \text{ for some } i \in \{0,1,\ldots,n\}\}\,.\end{aligned}$$

It is not immediately clear from equation (1) that all extensions F_n/F_{n-1}

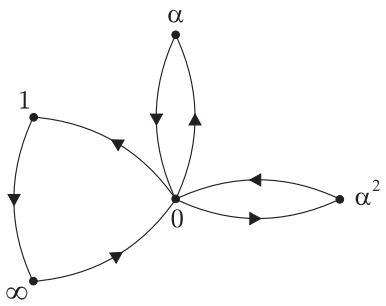


Fig. 1. Possible coordinate sequences of places with coordinates in $\mathbb{F}_4 \cup \{\infty\}$

are non–trivial. However, using the theorem by Kummer [6](III.3.7) one can show that for each $n \geq 1$ there exists a place in $\mathbb{P}_{F_n}$ for $n \geq 1$ with coordinates

$$P_{\underbrace{\beta,\cdots,\beta}_{n \text{ times}},\alpha\beta} \quad ,$$

where $\beta \in K$ is a root of the polynomial $T^3 + T + 1$. Furthermore, one can show that

$$f(P_{\beta,\cdots,\beta,\alpha\beta} \mid P_{\beta,\cdots,\beta,\bullet}) = 3,$$

and hence $[F_n : F_{n-1}] = 3$. Using the theorem by Kummer again, one also sees that $P_{\beta,\cdots,\beta,\bullet}$ splits completely in the extension F/F_{n-1}, and one can use this fact to show that K is the full constant field of all the F_n's, by an argument similar to the one in [8]. Thus we can now rightfully call $\mathcal{T}$ a tower. Next the ramification structure of $\mathcal{T}$ is addressed.

As it will be shown below, the ramification in the entire tower is controlled by three pyramids of places, which are shown in Figure 2. For the

purpose of working with these places they are best represented in the pyramids of function fields containing them. We call these pyramids *fundamental* due to the fact that the pyramids correspond to the three cycles in the graph on Figure 1, and thus in some sense every place in $\mathbb{P}_{F_n|\mathbb{F}_4}$ is made up from the fundamental ones. But before we can justify the name completely we must determine the ramification indices in the pyramids.

Proposition 2.1 (Fundamental pyramids). *The ramification indices in the fundamental pyramids are as shown in Figure 2.*

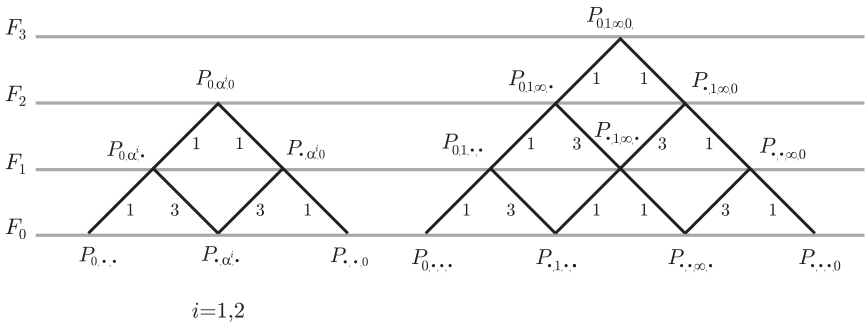


Fig. 2. Ramification in the fundamental pyramids.

Proof. Follows by the theorem [6](III.7.3) on Kummer extensions and Abhyankar's Lemma [6](III.8.9). □

By [6](III.7.3) it follows that the places that ramify in F_n/F_{n-1} must lie in $\mathbb{P}_{F_{n-1}|\mathbb{F}_4}$. Now, every such place has a coordinate sequence whose termination is either a full, or cut–off version of a fundamental pyramid, see Figure 3. Hence, since all outer edges facing north–west in Figure 2 have ramification index 1, it follows from Abhyankar's Lemma that for $n \geq 2$ any place $P \in \mathbb{P}_{F_n}$ is unramified. Furthermore, using the theorem by Kummer one sees that for $P \in \mathbb{P}_{F_n|\mathbb{F}_4}$ the relative degree of any extension of P is 1, and thus we arrive at the following:

Theorem 2.1. *The extension F_n/F_2 is unramified.*

Hence the only ramification in the tower takes place in the fundamental pyramids as described in Proposition 2.1, and since all places in $\mathbb{P}_{F_n|\mathbb{F}_4}$ have

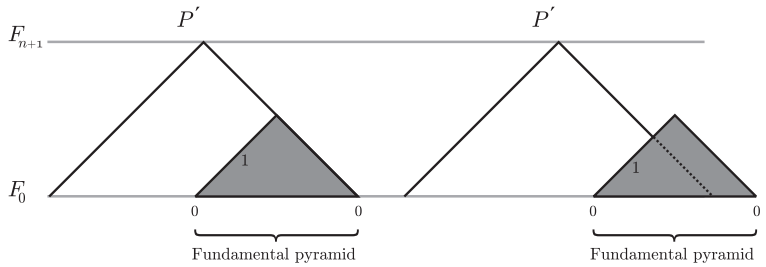


Fig. 3. The termination of a coordinate sequence must be either a full or cut–off fundamental pyramid.

relative degree one, it follows that the number of rational places $N(F_n)$ at the n–th level of the tower is at least

$$N(F_n) \geq 7 \cdot 3^{n-1} \text{ for } n \geq 2,$$

Furthermore, using the ramification structure determined above together with Hurwitz Genus Formula [6](III.4.12), one sees that the genus $g(F_n)$ of the n–th level of the tower is

$$g(F_n) = 3^{n-1} + 1, \text{ for } n \geq 2.$$

Putting the above equations together we get

$$\lim_{n\to\infty} \frac{N(F_n)}{g(F_n)} \geq \lim_{n\to\infty} \frac{7 \cdot 3^{n-1}}{3^{n-1}+1} = 7,$$

and since the constant field of the tower is $\mathbb{F}_{2^6}$ we conclude:

Theorem 2.2. *The tower $\mathcal{T}$ is optimal.*

Having proved the two claimed properties of the tower, we proceed to show how Riemann–Roch spaces in the tower can be computed.

3. An algorithm for computing Riemann–Roch spaces

In [5] an algorithm for computing bases of Riemann–Roch spaces in the Garcia–Stichtenoth tower was proposed. In the following it is shown how this algorithm can be adapted to work for the tower $\mathcal{T}$ for spaces $\mathcal{L}(D)$ with the support of D lying entirely above P_∞. In the following D is assumed to be of this type.

The core idea of the pole cancellation algorithm is the following observation: Let t be a local parameter for a place P. If two functions f, g have

the same pole order at P, with expansions say

$$f = a_{-m}t^{-m} + a_{-m+1}t^{-m+1} + \cdots$$
$$g = b_{-m}t^{-m} + b_{-m+1}t^{-m+1} + \cdots$$

then $v_P\left(b_{-m}f - a_{-m}g\right) > -m$. Thus the linear combination of f and g results in a function with strictly larger valuation. If we have yet another function with the same pole order as $b_{-m}f - a_{-m}g$, we can use the same procedure again to get a new function with even larger valuation, and if we have enough functions with the right pole orders, we can continue in this fashion to ultimately *cancel* the pole at P completely. Furthermore if we wish to cancel poles at more than one place coherently, we can use the above method again, but this time with several coupled linear equations in play. Using linear algebra we can exploit this to get a quite general and powerful technique for pole cancellation.

Now, say we are given the divisor D and one wishes to compute functions in $\mathcal{L}(D)$. The basic idea is first to find a set of functions E such that $\mathcal{L}(D)$ is contained in the K–span of E, and then use pole cancellation to remove undesired poles from elements in the span, i.e. finding linear combinations of the elements in E which only have poles in the support of D of the allowed order. In this way we trim down the span of E to eventually obtain $\mathcal{L}(D)$. To find such an E we introduce

$$\mathcal{L}(\infty P_\infty) = \bigcup_{m\geq 0} \mathcal{L}(mP_\infty),$$

and let R_n denote the integral closure of this ring in F_n. By [6](III.3.5) this ring has the alternative description

$$R_n = \bigcap_{P\nmid P_\infty} \mathcal{O}_P,$$

where P runs over all places of F_n not lying above P_∞. Now since we assumed that the support of D consisted of places above P_∞ we further get that

$$\mathcal{L}(D) \subseteq R_n.$$

Now, assume that at each level of the tower we have an *integral* basis E_n of F_n/F_{n-1}, i.e. a basis contained in R_n, then it follows from [6](III.3.4) that the dual basis E_n^* of E_n satisfies

$$R_n \subseteq \operatorname{span}_{R_{n-1}}\{E_n^*\}. \tag{2}$$

By repeatedly using this inclusion it follows that

$$\mathcal{L}(D) \subseteq R_n \subseteq \operatorname{span}_{R_0} \left\{ \prod_{i=1}^{n} e_i^* \mid e_i^* \in E_i^* \right\}, \tag{3}$$

and thus if we can determine integral bases and their duals, we will be able to span $\mathcal{L}(D)$ which in turn allow us to use the pole cancelling technique as described above. This is done next. Using the defining equation of the tower (1) one proves by induction:

Proposition 3.1. *Let $e_0 = x_0$ and $e_n = x_n(x_{n-1}+1)(x_{n-2}+1)\cdots(x_0+1)$ for $n \geq 1$. Then*

$$E_n = \{1, e_n, e_n^2\},$$

is a basis of the extension F_n/F_{n-1} and $E_n \subseteq R_n$, i.e. E_n is an integral basis. Furthermore, the dual basis is

$$E_n^* = \{1, e_n^{-1}, e_n^{-2}\}.$$

By using the ramification structure of the tower together with Figure 1 one can determine the divisor $(e_k)^{F_n}$ completely. We record this in:

Proposition 3.2. *Let P be a place of F_n where $n \geq 2$ and let k be an integer with $0 \leq k \leq n$. If P is not in $\mathbb{P}_{F_n|\mathbb{F}_4}$, then $e_k(P) \neq 0$ and $e_k(P) \neq \infty$. On the other hand if $P = P_{a_0 \ldots a_{n-1} a_n}$ then we have*

- *$e_k(P) = \infty$ if and only if $a_0 = \infty$. If $a_k = 0$ then $v_P(e_k) = -2$ and if $a_k \neq 0$ then $v_P(e_k) = -3$.*
- *$e_k(P) = 0$ if and only if $a_0 \neq \infty$ and $a_k = 0$. In this case $v_P(e_k) = 1$.*

The span of the functions in equation (3) can be simplified slightly by using the identity

$$\left(\frac{1}{e_n e_{n-1}}\right)^3 = \frac{1}{(e_0^3+1)e_0^3}\left(\sum_{i=0}^{n-1} e_i + 1\right).$$

to "reduce" denominators in E_n^*. Doing this, we get that the functions defined by

$$f_{m,\varepsilon}^{(n)} = \begin{cases} e_0^m \dfrac{\prod_{i=1}^{n} e_i^{\varepsilon_i}}{((e_0^3+1)e_0^3)^{\frac{n}{2}}} & n \text{ even} \\ e_0^m \dfrac{\prod_{i=1}^{n} e_i^{\varepsilon_i}}{(e_0^2+e_0+1)\,((e_0^3+1)e_0^3)^{\frac{n-1}{2}}} & n \text{ odd} \end{cases} \tag{4}$$

will also span $\mathcal{L}(D)$, that is

$$\mathcal{L}(D) \subseteq \operatorname{span}_K \left\{ f_{m,\varepsilon}^{(n)} \mid \varepsilon \in \{0,1,2\}^n, m \in \mathbb{N}_0 \right\}. \tag{5}$$

For notational convenience the superscript (n) on $f_{m,\varepsilon}^{(n)}$ will be left out in the following. Now, since by the Riemann–Roch theorem [6](I.5.15) the leftmost vector space in (5) is finite dimensional there exists some *maximal power* M, such that

$$\mathcal{L}(D) \subseteq \operatorname{span}_K \{ f_{m,\varepsilon} \mid \varepsilon \in \{0,1,2\}^n, 0 \leq m \leq M \}, \tag{6}$$

which gives us the desired finite set of functions whose K–span contains $\mathcal{L}(D)$.

Remark 3.1. One can also obtain a set of function spanning $\mathcal{L}(D)$ by an alternative method. By [6](III.2.10) the ring R_n is Dedekind, and thus every ideal in this ring can be generated by just two elements, see for instance [1](p. 192). By computing such generators of the ideals

$$I_{n,j} = \left\{ a \in R_{n-1} \mid a \cdot \frac{1}{e_n^j} \in R_n \right\},$$

one can prove that for $n \geq 3$ it holds

$$R_n = R_{n-1} \oplus \left(e_{n-2} e_{n-3} R_{n-1} + e_n^3 R_{n-1} \right) \cdot \frac{1}{e_n} \oplus$$

$$\left((e_{n-2} e_{n-3})^2 R_{n-1} + e_n^3 R_{n-1} \right) \cdot \frac{1}{e_n^2}.$$

Using this recursively one obtains a set of functions whose span is *equal* to R_n, whereas the span in (2) is a (strictly) larger set. The details of the involved computations are left out here.

To compute $\mathcal{L}(D)$ we must cancel undesired poles of the functions $f_{m,\varepsilon}$. Using Figure 1 one can prove that these functions potentially have poles at all the places in $\mathbb{P}_{F_n|\mathbb{F}_4}$, and thus in order to use the pole cancellation technique described above, each function must be expanded at each of these places. To find such expansion we determine local parameters of the places $\mathbb{P}_{F_n|\mathbb{F}_4}$, and since F_n/F_2 is unramified it is enough to determine local parameters for the places in $\mathbb{P}_{F_2|\mathbb{F}_4}$. This can be done directly using (1), and we get:

Proposition 3.3. *Let $P = P_{a_0,a_1,\ldots,a_n}$ be a place in $\mathbb{P}_{F_n|\mathbb{F}_4}$ with $n \geq 2$, then the following holds*

$$\begin{aligned} v_P(x_0) &= 1 \quad \textit{for } a_0 = 0, \\ v_P(x_2) &= 1 \quad \textit{for } a_0 = 1, \\ v_P(x_1) &= 1 \quad \textit{for } a_0 \in \{\alpha, \alpha^2, \infty\}. \end{aligned}$$

In order to use the pole cancellation technique sketched at the beginning of the section, we need a bound on how long the negative tails of the expansions of the $f_{m,\varepsilon}$'s at a place in $\mathbb{P}_{F_n|\mathbb{F}_4}$ can be. For P in $\mathbb{P}_{F_n|\mathbb{F}_4}$ let

$$\mu(P) = \max_{m,\varepsilon} \; -v_P(f_{m,\varepsilon}),$$

be such a bound. Using the results in Proposition 3.2 one can determine $\mu(P)$ in terms of n and M:

Proposition 3.4. *Let $P = P_{a_0,a_1,\cdots,a_n}$ be a place in $\mathbb{P}_{F_n|\mathbb{F}_4}$, and let*

$$u = \#\{a_i \mid a_i = 0\},$$

be the number of coordinates equal to 0. If $n \geq 2$ is even, then

$$\mu(P) = \begin{cases} 3(M-n) - 2u & a_0 = \infty \\ \frac{3n}{2} & a_0 \in \mathbb{F}_4 \end{cases}$$

Similarly when $n \geq 3$ is odd, it holds that

$$\mu(P) = \begin{cases} 3(M-n+1) - 2u & a_0 = \infty \\ \frac{3(n-1)}{2} & a_0 = 0 \\ \frac{3(n+1)}{2} & a_0 \in \mathbb{F}_4^*. \end{cases}$$

With Proposition 3.3 and 3.4, we can translate the problem of cancelling poles at undesired places into a linear algebra problem, and to this end a *pole cancellation matrix* A is introduced. For each of the functions $f_{m,\varepsilon}$ spanning $\mathcal{L}(D)$ the pole cancellation matrix has an attached row. Furthermore for each place P in $\mathbb{P}_{F_n|\mathbb{F}_4}$ and for each of the $\mu(P) - v_P(D)$ terms in the expansions of the $f_{m,\varepsilon}$'s beyond the "allowed" pole order, a column is attached. Let $f_{m,\varepsilon}$ be the function attached to the i–th row. Consider the $\mu(P) - v_P(D)$ columns attached to P, and let the expansion of $f_{m,\varepsilon}$ at P be

$$f_{m,\varepsilon} = a_{-\mu(P)}t^{-\mu(P)} + a_{-\mu(P)+1}t^{-\mu(P)+1} + \cdots +$$

$$a_{-v_P(D)-1}t^{-v_P(D)-1} + a_{-v_P(D)}t^{-v_P(D)} + \cdots,$$

$$f_{m,\varepsilon} \longrightarrow \left[\begin{array}{l|ccc|l} 1\,0\cdots & 0 & \cdots & 1 & \cdots 1\,1 \\ 1\,1\cdots & 1 & \cdots & 0 & \cdots 0\,1 \\ \quad\ddots & & \vdots & & \\ 0\,1\cdots & a_{-\mu(P)} & \cdots & a_{-v_P(D)-1} & \cdots 0\,1 \\ & & \vdots & & \ddots \\ 1\,0\cdots & 0 & \cdots & 1 & \cdots 1\,1 \\ 1\,1\cdots & 1 & \cdots & 0 & \cdots 0\,1 \end{array}\right] \underbrace{\qquad\qquad\qquad}_{P}$$

Fig. 4. Setup of the matrix for pole cancellation

then i–th row of these columns is filled with the coefficients $a_{-\mu(P)}, \ldots, a_{-v_P(D)-1}$, as indicated in Figure 4. With this definition of A, the linear combinations of the $f_{m,\varepsilon}$'s in $\mathcal{L}(D)$ are exactly the vectors v in the kernel of A, i.e. the vectors satisfying

$$vA = 0,$$

and these can be found by gaussian elimination.

Remark 3.2. By definition A has $(M+1)\cdot 3^n$ rows, and using Proposition 3.4 one can show that the number of columns is $(2n + M) \cdot 3^n$ when n is even, and $(2n + M + 1) \cdot 3^n$ when n is odd. Thus the dimension of the pole cancellation matrix is exponential in n and hence one must expect the amount of memory needed to store this, as well as the running time of the algorithm to grow rapidly with n.

We now cut the somewhat branched description of the pole cancellation algorithm above, to fit a pseudo–code formulation. First we introduce two auxiliary functions shown in Figure 5. The function `FillMatrix` handles the setup of the pole cancellation matrix, given the power series expansions of the x_i's. The function `Expand` recursively computes the power series expansion of x_{i+1} from the one of x_i. If x_L is a local parameter for the place $P = P_{a_0,\ldots,a_i}$ on which `Expand` is called, then the function assumes that $i \geq L$, which by the ramification structure found in Proposition 2.1

```
FillMatrix({x_0, x_1, ..., x_n}, P = P_{a_0,...,a_n}):
    For all m ∈ {0, 1, ..., M} and ε ∈ {0, 1, 2}^n do
       Compute the expansion of f_{m,ε}  ↩
         from the ones of x_0, x_1, ..., x_n
       Fill the row corresponding to f_{m,ε}  ↩
          in the columns attached to P.
    end for

Expand({x_0, x_1, ..., x_i}, P = P_{a_0,...,a_i}):
    If x_n has been computed then
       Call FillMatrix(x_0, x_1, ..., x_n, P_{a_0,...,a_n})
    else
       Compute [x_{i+1}^(1), x_{i+1}^(2), x_{i+1}^(3)] := ∛((x_i^2+x_i+1)/(x_i+1)^3)
       Call Expand(x_0, x_1, ..., x_i, x_{i+1}^(1), P_{a_0,...,a_i,x_{i+1}^(1)(P)})
       Call Expand(x_0, x_1, ..., x_i, x_{i+1}^(2), P_{a_0,...,a_i,x_{i+1}^(2)(P)})
       Call Expand(x_0, x_1, ..., x_i, x_{i+1}^(3), P_{a_0,...,a_i,x_{i+1}^(3)(P)})
    end if
```

Fig. 5. Auxiliary functions for filling the pole cancellation matrix and for computing power series expansions.

implies that P splits completely in the extension F_{i+1}/F_i, i.e. that

$$\frac{x_i^2 + x_i + 1}{(x_i + 1)^3}$$

always has three distinct third roots, cf. (1).

With these two auxiliary functions at hand we can now describe the main algorithm for pole cancellation, shown in Figure 6. In Proposition 3.3 local parameters for all places in $\mathbb{P}_{F_n|\mathbb{F}_4}$ were computed and it was seen that they could be chosen to be of the type x_i for some $i \in \{0, 1, 2\}$. For each of the different values of a_0 in $\mathbb{F}_4 \cup \{\infty\}$, the `Main`–function uses this to call the `Expand`–function with parameters

$$\texttt{Expand}(\{x_0, \cdots, x_L\}, P_{a_0,\cdots,a_L}),$$

where x_L is a local parameter for places with 0–th coordinate equal to a_0. After that, the recursive call in `Expand` makes the function traverse all "nodes" in the ramification tree of $\mathcal{T}$. At a node in the i–th level of the tower the expansion of x_i is computed from the expansion of x_{i-1}. The recursion is stopped once the expansions of all x_i's for $i \in \{1, \cdots, n\}$ are

found, i.e. when the leaves of the ramification tree are reached, and when this happens the `FillMatrix`–function is called. Another way to say all this, is that the call–structure of `Main` and `Expand` is chosen exactly to match the ramification structure of $\mathcal{T}$, so that the expansions of the x_i's are found at all places in $\mathbb{P}_{F_n|\mathbb{F}_4}$ and, through the `FillMatrix`–function, inserted into the pole cancellation matrix. Finally, at the end of the `Main`–function a basis of the kernel of this matrix is computed and thus, as described above, completing the pole cancellation algorithm.

It is somewhat unsatisfactory that the maximal power M is a parameter of the algorithm, since it is not directly relevant for the output and hence should preferably have been determined and kept internally in the algorithm. Unfortunately it is not possible to estimate the value of M needed to make the algorithm output functions in $\mathcal{L}(D)$. But by (6) we know that for *some* M this will happen, and thus using the algorithm in a trial–and–error manner by repeatedly increasing M, one is guarantied that the algorithm will output functions at some point.

4. Results

While the pole cancellation method for computing bases of the spaces $\mathcal{L}(D)$ is conceptually simple, the amount of computation needed to run the algorithm is extensive. In the following we fix a place

$$P_\infty^{(n)} = P_{\infty,0,1,\infty,0,1,\ldots} \in \mathbb{P}_{F_n|\mathbb{F}_4},$$

above infinity at each level of the tower. We wish to compute bases of the spaces $\mathcal{L}(mP_\infty^{(n)})$, and for illustrating the general nature of these it is not important which of the places with the above coordinate sequence is chosen to be $P_\infty^{(n)}$, as it turns out that the bases of the one–point spaces of all these places are similar.

The algorithm outlined in the previous section has been implemented in Magma. For the first few levels of the tower the algorithm rapidly arrives at moderately sized expressions for the basis elements, for instance it finds the function

$$f = (e_1^2 + 1) \cdot \left(\frac{e_2}{e_0}\right)^2 + (e_1^2 + e_0 + 1) \cdot \frac{e_2}{e_0} + e_1^2 + e_1,$$

in $\mathcal{L}(\infty P_\infty^{(2)})$ with pole order 4. But already a few levels up in the tower, the running time of the algorithm gets long, and the computed functions highly complex. As an illustration, a general element in $\mathcal{L}(\infty P_\infty^{(5)})$ involves roughly thousand terms with no apparent structure.

Input:
1. The level of the tower n, must be at least 2.
2. A divisor D with support above P_∞.
3. A maximal power M bounding the power m of e_0 in the spanning functions $f_{m,\varepsilon}$.

Output:
If M is large enough, the algorithm outputs K-independent functions in $\mathcal{L}(D)$.

Main(n, D, M):
Allocate a matrix A with $3^n(M+1)$ rows and $3^{n-1}(2n+M)$ columns if n is even, and $3^{n-1}(2n+M+1)$ columns if n is odd
For $a_0 = 0$:
- A local parameter for places with first coordinate 0 is x_0.
  Call Expand($\{x_0\}$, P_0)

For $a_0 \in \{\alpha, \alpha^2, \infty\}$:
- A local parameter for places with first coordinate a_0 is x_1.
- Given a_0, there is only one possible expansion of x_0.
  Compute $x_0 := \frac{\sqrt[3]{x_1^3+1}}{1+\sqrt[3]{x_1^3+1}}$
  Call Expand($\{x_0, x_1\}$, $P_{a_0,x_1(P)}$)

For $a_0 = 1$:
- A local parameter for places with first coordinate 1 is x_2
- There is only one possible expansion of x_1.
- Given the expansion of x_1, there are three possible expansions of x_0
  Compute $x_1 := \frac{\sqrt[3]{x_2^3+1}}{1+\sqrt[3]{x_2^3+1}}$
  Compute $[x_0^{(1)}, x_0^{(2)}, x_0^{(3)}] := \frac{\sqrt[3]{x_0^3+1}}{1+\sqrt[3]{x_0^3+1}}$
  For $i = 1, 2, 3$ do
    Call Expand($\{x_0^{(i)}, x_1, x_2\}$,$P_{1,\infty,0}$)

Compute a basis $\mathcal{B}$ for the kernel of A
If $\mathcal{B}$ is empty then output: FAILURE else output: $\mathcal{B}$

Fig. 6. Main function in the pole cancellation algorithm.

Let $\mathcal{W}(P)$ denote the Weierstraß semigroup of the place P. Using the pole cancellation algorithm, the Weierstraß semigroups of the places $P_\infty^{(n)}$ for $n \in \{1,2,3,4,5\}$ has been computed, and are shown in Figure 7. The framed numbers are the conductors. It has not been possible to compute

$$\begin{aligned}
\mathcal{W}\left(P_\infty^{(1)}\right) &= \{0, \boxed{2}, 3, 4, \ldots\} \\
\mathcal{W}\left(P_\infty^{(2)}\right) &= \{0, 4, \boxed{6}, 7, 8, \ldots\} \\
\mathcal{W}\left(P_\infty^{(3)}\right) &= \{0, 10, \boxed{12}, 13, 14, \ldots\} \\
\mathcal{W}\left(P_\infty^{(4)}\right) &= \{0, \boxed{29}, 30, 31, \ldots\} \\
\mathcal{W}\left(P_\infty^{(5)}\right) &= \{0, 68, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, \\
&\qquad 92, 94, 95, 96, 97, 98, \boxed{100}, 101, 102, \ldots\}
\end{aligned}$$

Fig. 7. Weierstraß semigroups of the places $P_\infty^{(n)}$.

$\mathcal{W}(P_\infty^{(6)})$ using the pole cancellation algorithm, since in this case the amount of memory needed to store the pole cancellation matrix gets tremendous. This supports the observation in Remark 3.2. There is seemingly no pattern in or among the semigroups, and this might indicate that the complexity of the bases of $\mathcal{L}(mP_\infty^{(n)})$ is an intrinsic property of the one–point spaces. Another fact supporting this, is that a Magma–implementation of the pole cancellation algorithm using the representation induced by the spanning functions in Remark 3.1, outputs equally complex basis functions

5. Conclusion

The pole cancellation algorithm allows us to compute bases of one–point spaces a few levels up in the tower. The computed basis functions are highly complex, and already at the sixth level they involve roughly a thousand terms. Expressions of this size overburden the algorithm and limit the practical relevance of the results. Furthermore, the voluminous output means that any hope of spotting general structures in the bases by inspection of the computed functions, must be abandoned.

There are however, some things that can be learnt from this work. Firstly, it shows that the pole cancellation algorithm can be applied in towers other than the Garcia–Stichtenoth tower, which was not clear previously. Secondly, it is seen that the unramifiedness can be used to explic-

itly determine functions spanning the integral closures R_n, but also that this does not directly help in determining bases of $\mathcal{L}(D)$. This contrasts the situation in the Garcia–Stichtenoth tower in which functions spanning the integral closure would immediately give bases for the Riemann–Roch spaces, but the determination of such functions is hard. Thus although $\mathcal{T}$ differs from the Garcia–Stichtenoth tower in many respects, the two towers are similar when it comes to complexity of one–point spaces.

References

1. Henri Cohen. *A Course in Computational Algebraic Number Theory*. Graduate Texts in Mathematics. Springer-Verlag, 2003.
2. Noam D. Elkies. Explicit modular towers. In *Proceedings of the Thirty-Fifth Annual Allerton Conference on Communication, Control and Computing*, September 1997.
3. Arnaldo Garcia and Henning Stichtenoth. A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vladut bound. *Inventiones Mathematicae*, 121:211–222, 1995.
4. Douglas A. Leonard. Finding the defining function for one-point algebraic-geometry codes. *IEEE Transactions On Information Theory*, 47(6):2566–2573, September 2001.
5. Kenneth W. Shum, Ilia Aleshnikov, P. Vijay Kumar, Henning Stichtenoth, and Vinay Deolalikar. A low-complexity algorithm for the construction of algebraic-geometric codes better than the Gilbert-Varshamov bound. *IEEE Transactions On Information Theory*, 47(6):2225–2241, September 2001.
6. Henning Stichtenoth. *Algebraic Function Fields and Codes*. Universitext. Springer, 1993.
7. M. A. Tsfasman, S. G. Vladut, and Th. Zink. Modular curves, Shimura curves, and Goppa codes, better than the Varshamov–Gilbert bound. *Math. Nachrichtentech*, 109:21–28, 1982.
8. Jörg Wulftange. *Zahme Türme algebraischer Funktionenkörper*. PhD thesis, Universität Essen, 2003.

Partial covering sequences: a method for designing classes of cryptographic functions

Claude Carlet

University of Paris 8
Department of Mathematics (MAATICAH),
2 rue de la liberté,
93526 Saint-Denis, Cedex, France;
E-mail: claude.carlet@inria.fr.

Few general constructions of cryptographic Boolean functions have been obtained so far. They have been found in empirical ways, and no general method for designing such constructions is known. Weakening a notion, called covering sequence and originally introduced for studying resilient functions, we show that, astonishingly enough, the knowledge of such "partial covering sequence" can give more information on the function. We deduce a method for designing constructions of Boolean functions with efficiently computable weights and Walsh spectra. The nonlinearity is then easier to handle for these functions. We illustrate this method with examples.

1. Introduction

Boolean functions (i.e. F_2-valued functions defined on the set F_2^n of all binary vectors of length n, where n is a positive integer) play a central role in the security of stream ciphers and of block ciphers. Much is known on the cryptographic criteria these functions must satisfy (cf. [4,5]) and a crucial problem in the design of fast stream ciphers resisting all known attacks is to determine sufficiently numerous fastly computable functions, say in 20 variables, satisfying these criteria in optimal ways (or, better, in nearly optimal ways, which is in fact more difficult). Even for small values of n, searching for the best cryptographic functions by visiting all Boolean functions in n variables is computationally impossible since their number 2^{2^n} is too large for $n \geq 6$. And their classification under the action of the group of affine automorphisms of F_2^n is unknown* for $n \geq 7$. Thus, we need constructions of

*Even if such classification was known, the number of classes would most probably be too large.

Boolean functions whose cryptographic parameters can be efficiently computed. But little is known in this matter; only one (class of) general direct construction(s) is known: the Maiorana-McFarland's construction and its generalizations (see [4]). This general method is based on the idea of concatenating the truth tables of several simple functions to obtain the truth table of a more complex one (in more variables) and its limits have been pointed out in [2,3]. Other direct or recursive constructions exist but they lead to few functions achieving good characteristics.

We give in the present paper a general method for designing new constructions (including the Maiorana-McFarland's one) and we illustrate it with examples. We begin by recalling the necessary background.

2. Background on Boolean functions

Notation: We shall have to distinguish in the whole paper between the additions of bits considered as integers, denoted by $+$, and the additions mod 2 (i.e. in F_2), denoted by $\oplus$. So all the multiple sums computed in characteristic 0 will be denoted by $\sum_i$ and all the sums computed modulo 2 will be denoted by $\bigoplus_i$. For simplicity and because there will be no ambiguity, we shall denote by $+$ the additions of the vectors of F_2^n (words) and of the field F_{2^n}.

Any Boolean function f on n variables admits a unique algebraic normal form (A.N.F.):

$$f(x_1, \ldots, x_n) = \bigoplus_{I \subseteq \{1,\ldots,n\}} a_I \prod_{i \in I} x_i,$$

where the a_I's are in F_2. We call *algebraic degree* of a Boolean function f and we denote by $d^\circ f$ the degree of its algebraic normal form. The *affine functions* are those functions of degrees at most 1. They are the simplest functions, from cryptographic viewpoint. On the contrary, *cryptographic functions must have high degrees* (cf. [1,9,10]).

The *Hamming weight* $w_H(f)$ of a Boolean function f on n variables is the size of its support $supp(f) = \{x \in F_2^n;\ f(x) = 1\}$. The *Hamming distance* $d_H(f, g)$ between two Boolean functions f and g is the Hamming weight of their difference, i.e. of $f \oplus g$. The *nonlinearity* N_f of f is its minimum distance to all affine functions. *Functions used in stream or block ciphers must have high nonlinearities* to resist the known attacks on these ciphers (correlation and linear attacks) [1,11]. A Boolean function f is called *bent* if its nonlinearity equals $2^{n-1} - 2^{n/2-1}$, which is the maximum possible value

(obviously, n must be even). Then, its distance to every affine function equals $2^{n-1} \pm 2^{n/2-1}$. This property can also be stated in terms of the discrete Fourier (or Hadamard) transform of f defined on F_2^n as $\widehat{f}(u) = \sum_{x \in F_2^n} f(x)\,(-1)^{x \cdot u}$ (where $x \cdot u$ denotes the usual inner product $x \cdot u = \bigoplus_{i=1}^{n} x_i\, u_i$). But it is more easily stated in terms of the *Walsh transform* of f, i.e. the discrete Fourier transform of the the "sign" function $\mathrm{f}(x) = (-1)^{f(x)}$, equal to $\widehat{\mathrm{f}}(u) = \sum_{x \in F_2^n} (-1)^{f(x) \oplus x \cdot u}$: f is bent if and only if $\widehat{\mathrm{f}}(u)$ has constant magnitude $2^{n/2}$ (cf. [12,13]). Indeed, the weight of f and the value $\widehat{\mathrm{f}}(0) = \sum_{x \in F_2^n} (-1)^{f(x)}$ being related through the relation

$$\widehat{\mathrm{f}}(0) = 2^n - 2w_H(f), \tag{1}$$

the Hamming distances between f and the affine functions $u \cdot x$ and $u \cdot x \oplus 1$ are equal to $2^{n-1} - \frac{1}{2}\widehat{\mathrm{f}}(u)$ and $2^{n-1} + \frac{1}{2}\widehat{\mathrm{f}}(u)$. Thus:

$$N_f = 2^{n-1} - \frac{1}{2} \max_{u \in F_2^n} |\widehat{\mathrm{f}}(u)|. \tag{2}$$

The upper bound on N_f is then a direct consequence of the Parseval relation on $\widehat{\mathrm{f}}$. Bent functions have degrees upper bounded by $n/2$. They are characterized by the fact that their *derivatives* $D_a f(x) = f(x) \oplus f(x + a)$, $a \neq 0$, are all *balanced*, i.e. have weight 2^{n-1}. But *cryptographic functions themselves must be balanced*, so that the systems using them resist statistical attacks [15]. Bent functions are not balanced but can be used to design highly nonlinear balanced functions.

Another criterion considered in this paper is resiliency. It plays a central role in some stream ciphers: in the standard original model of these ciphers, the outputs to n linear feedback shift registers are the inputs to a Boolean function, called combining function. The output to the function produces the keystream, which is then bitwise xored with the message to produce the cipher. Some devide-and-conquer attacks exist on this method of encryption (cf. [1,16]). To resist these attacks, the system must use a combining function whose output distribution probability is unaltered when any m of the inputs are fixed [16], with m as large as possible. This property, called *m-th order correlation-immunity*, is characterized by the set of zero values in the Walsh spectrum [17] (see also [4]): f is m-th order correlation-immune if and only if $\widehat{\mathrm{f}}(u) = 0$, i.e. $\widehat{f}(u) = 0$, for all $u \in F_2^n$ such that $1 \leq w_H(u) \leq m$, where $w_H(u)$ denotes the Hamming weight of the n-bit vector u, (the number of its nonzero components). Balanced m-th order correlation-immune functions are called *m-resilient* functions. They are characterized by the fact that $\widehat{\mathrm{f}}(u) = 0$ for all $u \in F_2^n$ such that $0 \leq w_H(u) \leq m$.

Any m-th order correlation immune function on n variables has degree at most $n-m$, any m-resilient function ($0 \leq m < n-1$) has algebraic degree smaller than or equal to $n-m-1$ and any $(n-1)$-resilient function has algebraic degree 1. The nonlinearity of any m-resilient function ($m \leq n-2$) is divisible by 2^{m+1} and is therefore upper bounded by $2^{n-1} - 2^{m+1}$.

High degree and high nonlinearity are needed for all applications of Boolean functions in stream ciphers. High resiliency is also necessary in the combiner model. Designing constructions of Boolean functions meeting these cryptographic criteria is still a crucial challenge nowadays. Moreover, the recent algebraic attacks [7] add further important criteria. The most basic one is that the functions f and $f \oplus 1$ must not admit annihilators (that is, functions g whose product with f or $f \oplus 1$ is null) of low degrees, see [14]. It has been proven in [8] that random functions have such property, but this criterion, and the others related to algebraic attacks, are further reasons why we need to be able to construct numerous balanced functions achieving high nonlinearity and possibly high resiliency order.

We recall now the definitions of the main known constructions cited above:

Maiorana-McFarland: The principle for designing the truth tables of Maiorana-McFarland's functions is concatenating the truth tables of affine functions. Up to permutation of the variables, when we fix the (say) s last variables in the ANF of the function, we obtain affine functions in $r = n-s$ variables. Denoting by x the vector of the r first variables of the function and by y the vector of the s last variables, there exists then a Boolean function g on F_2^s and a mapping ϕ from F_2^s to F_2^r such that the global function has the expression:

$$f_{\phi,g}(x,y) = x \cdot \phi(y) \oplus g(y) = \bigoplus_{i=1}^{r} x_i \phi_i(y) \oplus g(y),\ x \in F_2^r,\ y \in F_2^s \qquad (3)$$

where we denote by $\phi_i(y)$ the ith coordinate of $\phi(y)$. If every element in $\phi(F_2^s)$ has Hamming weight strictly greater than k, then $f_{\phi,g}$ is m-resilient with $m \geq k$. Indeed, for every $a \in F_2^r$ and every $b \in F_2^s$, we have

$$\widehat{\mathrm{f}_{\phi,\mathrm{g}}}(a,b) = 2^r \sum_{y \in \phi^{-1}(a)} (-1)^{g(y) \oplus b \cdot y}, \qquad (4)$$

Remark: For $r = 1$, we obtain all Boolean functions (but Relation (4) gives little information). The Boolean functions that we will design in this paper will be contained in a Maiorana-McFarland's class for some r but we shall

have more information on the Walsh transforms of these functions than with Relation (4).

Super-classes of Maiorana-McFarland [2,3]: The functions in these super-classes are the concatenations of particular quadratic functions instead of affine ones. Since these particular quadratic functions are themselves concatenations of affine functions, the functions in these super-classes can be viewed as particular Maiorana-McFarland's functions, but we can say more on these functions than when viewing them just as Maiorana-McFarland's functions.

1. Let n and r be positive integers such that $r < n$. Denote the integer part $\left\lfloor \frac{r}{2} \right\rfloor$ by t and $n - r$ by s. Let ψ be a mapping from F_2^s to F_2^t and let $\psi_1, \cdots, \psi_t$ be its coordinate functions. Let ϕ be a mapping from F_2^s to F_2^r and let $\phi_1, \cdots, \phi_r$ be its coordinate functions. Let g be a Boolean function on F_2^s. The function $f_{\psi,\phi,g}$ is defined on $F_2^n = F_2^r \times F_2^s$ as

$$f_{\psi,\phi,g}(x,y) = \bigoplus_{i=1}^{t} x_{2i-1}x_{2i}\psi_i(y) \oplus x \cdot \phi(y) \oplus g(y) =$$

$$\bigoplus_{i=1}^{t} x_{2i-1}x_{2i}\psi_i(y) \bigoplus_{j=1}^{r} x_i\, \phi_i(y) \oplus g(y);\ x \in F_2^r,\ y \in F_2^s. \tag{5}$$

Then for every $a \in F_2^r$ and every $b \in F_2^s$ we have

$$\widehat{\mathrm{f}_{\psi,\phi,\mathrm{g}}}(a,b) = \sum_{y \in E_a} 2^{r - w_H(\psi(y))} (-1)^{\bigoplus_{i=1}^{t} (\phi_{2i-1}(y) \oplus a_{2i-1})(\phi_{2i}(y) \oplus a_{2i}) \oplus g(y) \oplus y \cdot b}, \tag{6}$$

where E_a is the superset of $\phi^{-1}(a)$ equal if r is even to

$$\{y \in F_2^s / \ \forall i \leq t,\ \psi_i(y) = 0 \Rightarrow (\phi_{2i-1}(y) = a_{2i-1} \text{ and } \phi_{2i}(y) = a_{2i})\},$$

and if r is odd to

$$\left\{ y \in F_2^s / \ \begin{cases} \forall i \leq t,\ \psi_i(y) = 0 \Rightarrow (\phi_{2i-1}(y) = a_{2i-1} \text{ and } \phi_{2i}(y) = a_{2i}) \\ \phi_r(y) = a_r \end{cases} \right\}.$$

2. Let ϕ_1, ϕ_2 and ϕ_3 be three functions from F_2^s into F_2^r and let g be any Boolean function on F_2^s. The functions of the second class are defined by[†]:

$$f_{\phi_1,\phi_2,\phi_3,g}(x,y) = (x \cdot \phi_1(y))\,(x \cdot \phi_2(y)) \oplus x \cdot \phi_3(y) \oplus g(y), \tag{7}$$

where $x \in F_2^r$, $y \in F_2^s$. Denoting by E the set of all $y \in F_2^s$ such that the vectors $\phi_1(y)$ and $\phi_2(y)$ are linearly independent, if the vector $\phi_2(y)$

[†]There exists a more general version, see [3].

is nonzero for every $y \in F_2^s$, then for every $a \in F_2^r$ and every $b \in F_2^s$, $\widehat{\mathrm{f}_{\phi_1,\phi_2,\phi_3,\mathrm{g}}}(a,b)$ equals

$$2^{r-1} \sum_{\substack{y\in E;\\ \phi_3(y)+a\in\{0,\phi_1(y),\phi_2(y)\}}} (-1)^{g(y)\oplus b\cdot y} - 2^{r-1} \sum_{\substack{y\in E;\\ \phi_3(y)+a=\phi_1(y)+\phi_2(y)}} (-1)^{g(y)\oplus b\cdot y}$$

$$+2^r \sum_{\substack{y\in F_2^s\setminus E;\\ \phi_3(y)+a=\phi_1(y)}} (-1)^{g(y)\oplus b\cdot y}. \tag{8}$$

A few other constructions are known (see [4] for a review) but some lead to functions whose nonlinearities are unknown and the others (e.g. the secondary constructions) lead to un-numerous functions. So, to summarize, only one general idea of construction of cryptographic Boolean functions whose nonlinearities and resiliency orders can be handled exists (the Maiorana-McFarland's one and its generalizations), and no general method to design such constructions is known. The purpose of this paper is to give such a method, which leads to further constructions (including the Maiorana-McFarland's one).

3. Partial covering sequences

As recalled in the introduction, the resiliency and the nonlinearity can be expressed by means of the Walsh transform. Thus the crucial point for computing these cryptographic parameters of a Boolean function is to get information on its Walsh spectrum. Computing the Walsh spectrum of f is equivalent to computing the sum $\sum_{x\in F_2^n}(-1)^{f(x)}$ (which is equivalent to computing the weight of f) as well as all the similar sums in which $f(x)\oplus a\cdot x$ ($a\in F_2^n$) replaces $f(x)$. A natural way for this is to relate the sum $\sum_{x\in F_2^n}(-1)^{f(x)\oplus a\cdot x}$ to a sum $\sum_{x\in A}(-1)^{f(x)\oplus a\cdot x}$, where A is a subset of F_2^n such that the properties of this partial sum (divisibility by powers of 2, magnitude, ...) appear more clearly than for the global sum. This kind of *reduction* is what we have for instance with the Maiorana-McFarland functions and their generalizations, according to Relations (4), (6) and (8) and to the other properties recalled in [4]. This is why the cryptographic parameters of these functions could be more easily computed than for general functions!

Proposition 3.3 and Theorem 3.2, below, describe a general situation in which this occurs, by generalizing a property pointed out in [6] for balanced functions that we first recall.

3.1. *Covering sequences*

Covering sequences have been introduced in relationship with the resiliency criterion.

Definition 3.1. A covering sequence for a Boolean function f is an integer-valued sequence $(\lambda_a)_{a\in F_2^n}$ (or a real-valued one, but taking real-valued sequences instead of integer-valued ones has no practical interest) such that the integer-valued function $\sum_{a\in F_2^n} \lambda_a D_a f(x)$, where $D_a f(x) = f(x) \oplus f(x+a)$, takes a constant value ρ, called the level of the sequence. If ρ is non-zero, we say that the covering sequence is non-trivial.

The term of "covering sequence" had been chosen in [6] to express the fact that the corresponding weighted sum of derivatives "covers" ρ times the whole vector space (the indicator of). The following results were shown:

Proposition 3.1. *[6] Any function admitting a non-trivial covering sequence is balanced. Conversely, any balanced function admits the constant sequence 1 as covering sequence, with level* 2^{n-1}.

Theorem 3.1. *[6] Let f be any Boolean function on ${F_2}^n$ and $\lambda = (\lambda_a)_{a\in {F_2}^n}$ any numerical sequence. Let us denote by $\mathcal{S}_f$ the support $\{b \in {F_2}^n \mid \widehat{\mathrm{f}}(b) \neq 0\}$ of $\widehat{\mathrm{f}}$ and let $\widehat{\lambda}$ be the discrete Fourier transform of the sequence:*

$$\forall b \in {F_2}^n,\ \widehat{\lambda}(b) = \sum_{a\in F_2^n} \lambda_a (-1)^{a\cdot b}.$$

Then f admits λ as covering sequence if and only if $\widehat{\lambda}$ takes constant value on $\mathcal{S}_f$. Let r be this constant value, then the level of the covering sequence is the number $\frac{1}{2}[\widehat{\lambda}(0) - r]$.

Hence, knowing a covering sequence (trivial or not) of a Boolean function f gives information on the support of its Walsh transform (equivalently, on those values of a such that $f(x) \oplus a \cdot x$ is balanced), and therefore on the resiliency of f. More precisely:

Corollary 3.1. *[6] 1. If f admits the covering sequence $\lambda = (\lambda_a)_{a\in F_2^n}$ with level $\rho \neq 0$, then f is k-resilient where*

$$k + 1 = \min\{w_H(b);\ b \in F_2^n;\ \widehat{\lambda}(b) = \widehat{\lambda}(0) - 2\rho\}.$$

2. Conversely, if f is k-resilient and if it is not $(k+1)$-resilient, then there exists at least one covering sequence $\lambda = (\lambda_a)_{a\in F_2^n}$ with level $\rho \neq 0$ such that $k + 1 = \min\{w_H(b);\ b \in F_2^n;\ \widehat{\lambda}(b) = \widehat{\lambda}(0) - 2\rho\}$.

The interest of the notion is that complex Boolean functions can admit simple covering sequences[‡]. But the knowledge of a covering sequence gives no information on the non-zero values of the Walsh transform of the function and therefore on its nonlinearity. In the next definition, we weaken the notion of covering sequence. The resulting notion is then also relevant to non-balanced functions. This allows obtaining information on the non-zero values of the Walsh transform.

3.2. *Generalization to partial covering sequences*

Definition 3.2. Let f be a Boolean function on F_2^n. A partial covering sequence for f is a sequence $(\lambda_a)_{a\in F_2^n}$ such that the numerical function on F_2^n equal to $\sum_{a\in F_2^n} \lambda_a D_a f(x)$ takes two values ρ and ρ' (distinct or not) called the *levels* of the sequence. The partial covering sequence is called *non-trivial* if one of the constants is nonzero.

Proposition 3.2. *For every Boolean function f, the all-one sequence is a partial covering sequence whose levels are $\rho = w_H(f)$ and $\rho' = 2^n - w_H(f)$.*

Proof:
We have:

$$\sum_{a\in F_2^n} D_a f(x) = 2^{n-1} - \frac{1}{2}\sum_{a\in F_2^n} (-1)^{D_a f(x)} = 2^{n-1} - \frac{(-1)^{f(x)}}{2}\sum_{y\in F_2^n} (-1)^{f(y)}.$$

Hence, the function $\sum_{a\in F_2^n} D_a f(x)$ takes two values $2^{n-1} - \frac{1}{2}\sum_{y\in F_2^n}(-1)^{f(y)} = w_H(f)$ and $2^{n-1} + \frac{1}{2}\sum_{y\in F_2^n}(-1)^{f(y)} = 2^n - w_H(f)$. □

3.3. *The reduction principle*

Proposition 3.3. *Let $(\lambda_a)_{a\in F_2^n}$ be a non-trivial partial covering sequence of a Boolean function f, of levels $\rho \neq 0$ and ρ'.*
Let $A = \{x \in F_2^n \,/\, \sum_{a\in F_2^n} \lambda_a D_a f(x) = \rho'\}$.
Then:

$$\widehat{\mathrm{f}}(0) = \left(1 - \frac{\rho'}{\rho}\right)\sum_{x\in A}(-1)^{f(x)}.$$

[‡]e.g. the constant ones, but these ones give too little information since their Fourier transforms are non-zero at one point only.

Proof:
For every $a \in F_2^n$, we have

$$\sum_{x \in D_a f^{-1}(1)} (-1)^{f(x)} = 0.$$

Indeed the set $(D_a f)^{-1}(1)$ is stable under translation by a and we have for every x in this set: $f(x+a) = f(x) \oplus 1$ and thus $(-1)^{f(x+a)} + (-1)^{f(x)} = 0$. We deduce that $\sum_{x \in F_2^n} D_a f(x)(-1)^{f(x)} = 0$ and by summing up λ_a times this relation when a ranges over F_2^n, we obtain that the sum

$$\sum_{a \in F_2^n} \lambda_a \left(\sum_{x \in F_2^n} D_a f(x)(-1)^{f(x)} \right)$$

is null. This sum equals $\rho' \sum_{x \in A}(-1)^{f(x)} + \rho \sum_{x \notin A}(-1)^{f(x)}$.
We deduce:

$$\sum_{x \in F_2^n} (-1)^{f(x)} = \sum_{x \in A}(-1)^{f(x)} + \sum_{x \notin A}(-1)^{f(x)} = \left(1 - \frac{\rho'}{\rho}\right) \sum_{x \in A}(-1)^{f(x)}.$$

□

3.4. *More on the Walsh Transform*

We can derive in fact much more information on the Walsh transform of a function f, given a partial covering sequence of f:

Theorem 3.2. *Let $(\lambda_a)_{a \in F_2^n}$ be a partial covering sequence of a Boolean function f, of levels ρ and ρ'.*
Let $A = \{x \in F_2^n \,/\, \sum_{a \in F_2^n} \lambda_a D_a f(x) = \rho'\}$ (assuming that $\rho' \neq \rho$; otherwise, when λ is in fact a covering sequence of level ρ, we set $A = \emptyset$).
Then, for every vector $b \in F_2^n$, we have:

$$\left(\widehat{\lambda}(b) - \widehat{\lambda}(0) + 2\,\rho\right)\, \widehat{\mathrm{f}}(b) = 2\,(\rho - \rho') \sum_{x \in A} (-1)^{f(x) \oplus b \cdot x}.$$

Proof: By definition, we have, for every $x \in F_2^n$:

$$\sum_{a \in F_2^n} \lambda_a D_a f(x) = \rho'\, 1_A(x) + \rho\, 1_{A^c}(x)$$

and therefore:

$$\sum_{a \in F_2^n} \lambda_a (-1)^{D_a f(x)} = \sum_{a \in F_2^n} \lambda_a (1 - 2\, D_a f(x))$$

$$= \sum_{a \in F_2^n} \lambda_a - 2\,\rho'\,1_A(x) - 2\,\rho\,1_{A^c}(x).$$

We deduce:

$$\sum_{a \in F_2^n} \lambda_a (-1)^{f(x+a)} = (-1)^{f(x)} \left(\sum_{a \in F_2^n} \lambda_a - 2\,\rho'\,1_A(x) - 2\,\rho\,1_{A^c}(x) \right). \quad (9)$$

The Fourier transform of the function $(-1)^{f(x+a)}$ maps every vector $b \in F_2^n$ to the value $\sum_{x \in F_2^n} (-1)^{f(x+a) \oplus x \cdot b} = \sum_{x \in F_2^n} (-1)^{f(x) \oplus (x+a) \cdot b} = (-1)^{a \cdot b}\, \widehat{\mathrm{f}}(b)$. Hence, taking the Fourier transform of both terms of equality (9), we get:

$$\left(\sum_{a \in F_2^n} \lambda_a (-1)^{a \cdot b} \right) \widehat{\mathrm{f}}(b) =$$

$$\left(\sum_{a \in F_2^n} \lambda_a \right) \widehat{\mathrm{f}}(b) - 2\,\rho' \sum_{x \in A} (-1)^{f(x) \oplus b \cdot x} - 2\,\rho \sum_{x \in A^c} (-1)^{f(x) \oplus b \cdot x},$$

that is

$$\widehat{\lambda}(b)\, \widehat{\mathrm{f}}(b) = \widehat{\lambda}(0)\, \widehat{\mathrm{f}}(b) - 2\,\rho \widehat{\mathrm{f}}(b) + 2\,(\rho - \rho') \sum_{x \in A} (-1)^{f(x) \oplus b \cdot x}.$$

Hence:

$$\left(\widehat{\lambda}(b) - \widehat{\lambda}(0) + 2\,\rho \right)\, \widehat{\mathrm{f}}(b) = 2\,(\rho - \rho') \sum_{x \in A} (-1)^{f(x) \oplus b \cdot x}. \qquad \square$$

Corollary 3.2. *If we have* $\widehat{\lambda}(b) - \widehat{\lambda}(0) + 2\,\rho \neq 0$ *for every* $b \in F_2^n$ *then*

$$\max_{b \in F_2^n} |\widehat{\mathrm{f}}(b)| \leq \frac{2\,|\rho - \rho'|}{\min_{b \in F_2^n} \left| \widehat{\lambda}(b) - \widehat{\lambda}(0) + 2\,\rho \right|}\, card(A)$$

and

$$N_f \geq 2^{n-1} - \frac{|\rho - \rho'|}{\min_{b \in F_2^n} \left| \widehat{\lambda}(b) - \widehat{\lambda}(0) + 2\,\rho \right|}\, card(A).$$

Hence, the knowledge of a partial covering sequence of f gives not only information on the support of its Walsh transform but also on the values of its Walsh transform and, in some cases, on its nonlinearity.

3.5. *Partially 0-regular functions*

A function f is called ρ-regular (see [6]) if the indicator λ of all the vectors of Hamming weight 1 is a covering sequence of level ρ of f. It was observed in [6] that ρ-regular functions are $(\rho-1)$-resilient. This is in fact a direct consequence of Theorem 3.2, since $\widehat{\lambda}(b)$ being then equal to $n-2w_H(b)$, we have $w_H(b) \neq \rho \Rightarrow \widehat{\mathrm{f}}(b) = 0$ (since A is empty). Iterative constructions of ρ-regular functions were derived, among which some had optimal degrees and nonlinearities.

More generally, if the indicator λ of all the vectors of Hamming weight 1 is a partial covering sequence of levels ρ and ρ' of f, we call f a *partially ρ-regular function*. According to Theorem 3.2, for every vector b whose weight is different from ρ, we have then $\widehat{\mathrm{f}}(b) = \frac{\rho-\rho'}{\rho-w_H(b)} \sum_{x\in A}(-1)^{f(x)\oplus b\cdot x}$. In particular, if $\rho = 0$, we have $\widehat{\mathrm{f}}(b) = \frac{\rho'}{w_H(b)} \sum_{x\in A}(-1)^{f(x)\oplus b\cdot x}$ for every nonzero vector b. This implies that $N_f \geq \min\left(w_H(f), 2^{n-1} - \frac{\rho'}{2}\, card(A)\right)$.

4. Linear sets of derivatives

We have seen that all functions admit the all-one sequence as partial covering sequence. But this precise sequence does not give information on the function (this seems natural, since this sequence is valid for every Boolean function). Indeed, let us assume that f is not balanced so that $\rho = w_H(f) \neq \rho' = 2^n - w_H(f)$. Then $A = \{x \in F_2^n / \sum_{a\in F_2^n} D_a f(x) = 2^n - w_H(f)\}$ equals the support $\{x \in F_2^n / f(x) = 1\}$ of f, since we have $\sum_{a\in F_2^n} D_a f(x) = w_H(f)$ if $f(x) = 0$ and $\sum_{a\in F_2^n} D_a f(x) = 2^n - w_H(f)$ if $f(x) = 1$. Proposition 3.3 gives then the trivial equality $2^n - 2w_H(f) = \left(1 - \frac{2^n - w_H(f)}{w_H(f)}\right)(-card(A)) = 2^n - 2w_H(f)$. Theorem 3.2 does not give more insight: denoting by δ_0 the indicator of $\{0\}$, it gives the equality:

$$(2^n\delta_0(b) - 2^n + 2\,w_H(f))\ \widehat{\mathrm{f}}(b) = (2^{n+1} - 4\,w_H(f)) \sum_{x\in supp(f)} (-1)^{b\cdot x}$$

that is, $w_H(f)\ \widehat{\mathrm{f}}(0) = (2^n - 2\,w_H(f))\,w_H(f)$ if $b = 0$, which is a trivial equality, and $(2\,w_H(f) - 2^n)\ \widehat{\mathrm{f}}(b) = (2^{n+1} - 4\,w_H(f))\ \sum_{x\in supp(f)}(-1)^{b\cdot x}$ if $b \neq 0$. It is a simple matter to see that this last equality is trivial too.

Another (obvious) case of partial covering sequence, valid for every function, is the indicator of a singleton: $\lambda_a = 1$ if $a = a_0$; $\lambda_a = 0$ otherwise. Then $\rho = 1$, $\rho' = 0$, $A = F_2^n \setminus supp(D_{a_0} f)$ and Theorem 3.2 states that

$$\left((-1)^{a_0\cdot b} + 1\right)\ \widehat{\mathrm{f}}(b) = 2 \sum_{x\in A} (-1)^{f(x)\oplus b\cdot x}.$$

Here again, this equality valid for every function is straightforward.
So, we still need to find a way of determining usable (and specialized) partial covering sequences of Boolean functions.

Remarks: 1. If some derivatives of f, say $D_{a_1}f, \cdots, D_{a_k}f$, have disjoint supports, then the indicator of $\{a_1, \cdots, a_k\}$ is also clearly a partial covering sequence of f. Then $\rho = 1$, $\rho' = 0$, $A = \{x \in F_2^n \,/\, D_{a_1}f(x) = \cdots = D_{a_k}f(x) = 0\}$ and Theorem 3.2 states that

$$\left(\sum_{i=1}^{k}(-1)^{a_i\cdot b} - k + 2\right)\,\widehat{\mathrm{f}}(b) = 2\,\sum_{x\in A}(-1)^{f(x)\oplus b\cdot x}.$$

But only a constant function can have two derivatives with disjoint supports.
2. If the supports of two derivatives $D_{a_1}f$, $D_{a_2}f$ of f cover the whole space F_2^n, then the indicator of $\{a_1, a_2\}$ is a partial covering sequence, $\rho = 1$, $\rho' = 2$, $A = \{x \in F_2^n \,/\, D_{a_1}f(x) = D_{a_2}f(x) = 1\}$ and Theorem 3.2 states that

$$\left(\sum_{i=1}^{2}(-1)^{a_i\cdot b}\right)\,\widehat{\mathrm{f}}(b) = 2\,\sum_{x\in A}(-1)^{f(x)\oplus b\cdot x}.$$

□

We introduce now a simple general case of partial covering sequence.

Definition 4.1. We call *linear set of derivatives* of f any set $\mathcal{E} = \{D_af;\, a \in E\}$ of derivatives which is not reduced to the null function and which is stable under addition (in other words, which is a non-trival F_2-vector space).

Several sets E can correspond to a same linear set $\mathcal{E}$, when some functions D_af, $D_{a'}f$, $a \neq a' \in E$, are identical. We shall then allow repetitions, but we shall assume that every element of $\mathcal{E}$ appears the same number of times when a ranges over E.

Let $\mathcal{E} = \{D_af;\, a \in E\}$ be a linear set of derivatives of f. For every $x \in F_2^n$, the function $g \in \mathcal{E} \mapsto g(x)$ is an F_2-linear form over the vector-space $\mathcal{E}$. The sum $\sum_{a\in E} D_af(x)$ is then the Hamming weight of this linear function, counted once if we do not allow repetitions and several times if we do allow them. It equals therefore 0 or $\frac{card(E)}{2}$. We deduce, also using Theorem 3.2 and the facts that, if λ is the indicator of a set E, then $\widehat{\lambda}(0)$ equals the size of E and if λ is the indicator of a vector-space, then $\widehat{\lambda}$ equals $\widehat{\lambda}(0)$ times the indicator of its orthogonal:

Corollary 4.1. *Let $\mathcal{E} = \{D_a f;\ a \in E\}$ be a linear set of derivatives of f. Let $(\lambda_a)_{a\in F_2^n}$ be the sequence equal to the indicator of E (defined by $\lambda_a = 1$ if $a \in E$ and $\lambda_a = 0$ otherwise). Then $(\lambda_a)_{a\in F_2^n}$ is a partial covering sequence for f with levels $\rho = \frac{card(E)}{2}$ and $\rho' = 0$. Denoting $A = \{x \in F_2^n \,/\, D_a f(x) = 0;\ \forall a \in E\}$, we have then:*

$$\widehat{\lambda}(b)\,\widehat{\mathrm{f}}(b) = \widehat{\lambda}(0)\sum_{x\in A}(-1)^{f(x)\oplus b\cdot x}.$$

If E is a vector-space, this is equivalent to:

$$\widehat{\mathrm{f}}(b) = \sum_{x\in A}(-1)^{f(x)\oplus b\cdot x},\ \forall b \in E^{\perp}$$

and $\sum_{x\in A}(-1)^{f(x)\oplus b\cdot x} = 0,\ \forall b \notin E^{\perp}$.

When $\widehat{\lambda}(b)$ is nonzero for every b, this corollary gives an information on the value of $\widehat{\mathrm{f}}(b)$, similar to what gives Relation (4) in the case of Maiorana McFarland's functions. It implies then the bound:

$$N_f \geq 2^{n-1} - \frac{|\widehat{\lambda}(0)|}{2\ \min_{a\in F_2^n}|\widehat{\lambda}(b)|}\ card(A).$$

4.1. *Revisiting known classes*

4.1.1. *Quadratic functions*

- Let f be any quadratic function. Denote by $\varphi_f(a,x)$ the associated symplectic form (see [12], ch. 15): $\varphi_f(a,x) = f(0)\oplus f(x)\oplus f(a)\oplus f(x+a)$. Denote its kernel $\{a \in F_2^n;\ \forall x \in F_2^n,\ \varphi_f(a,x) = 0\}$ by E_f. We have $D_a f(x) = \varphi_f(a,x)\oplus f(0)\oplus f(a)$ for every $a \in F_2^n$ and thus $D_a f(x) = f(0)\oplus f(a)$ for every $a \in E_f$. The set $\mathcal{E} = \{D_a f;\ a \in E_f\}$ is a linear set of (constant) derivatives, since the function $a \mapsto f(0)\oplus f(a)$ is linear on the vector space E_f. Taking for partial covering sequence the indicator of E_f, we have $\rho = \frac{card(E_f)}{2}$, and the set A of this partial covering sequence is empty if there exists $a \in E_f$ such that $f(a) \neq f(0)$, in which case Corollary 4.1 gives:

$$\widehat{\lambda}(b)\,\widehat{\mathrm{f}}(b) = 0,$$

and since $\widehat{\lambda}(b)$ equals $card(E_f)$ if $b \in E_f^{\perp}$ and is null otherwise, we deduce that $\widehat{\mathrm{f}}(b) = 0$ for every $b \in E_f^{\perp}$ (in particular, f is balanced). The set A equals F_2^n otherwise, in which case Corollary 4.1 gives:

$$\widehat{\lambda}(b)\,\widehat{\mathrm{f}}(b) = \widehat{\lambda}(0)\,\widehat{\mathrm{f}}(b),$$

that is $\widehat{\mathrm{f}}(b) = 0$ for every $b \notin E_f^{\perp}$.
- We know that f is equal, up to a linear bijective transformation, to the function $f(x, y, z) = x \cdot y \oplus h(x, y, z)$; $x, y \in F_2^r$; $z \in F_2^{n-2r}$, where r is a non-negative integer smaller than or equal to $n/2$ and where h is affine on F_2^n (say $h(x, y, z) = u \cdot x \oplus v \cdot y \oplus w \cdot z \oplus \epsilon$). Notice that we have then $E_f = \{0\} \times \{0\} \times F_2^{n-2r}$. We can take for partial covering sequence the indicator of the super-set of E_f, equal to $\{0\} \times F_2^r \times F_2^{n-2r}$. For every element $a = (0, a', a'')$ in this set, we have $D_a f(x, y, z) = x \cdot a' \oplus v \cdot a' \oplus w \cdot a''$ and we obtain a linear set of derivatives. The nonzero level of this partial covering sequence equals 2^{n-r-1} and the set $A = \{(x, y, z);\ x, y \in F_2^r,\ z \in F_2^{n-2r};\ \forall a' \in F_2^r,\ \forall a'' \in F_2^{n-2r},\ x \cdot a' \oplus v \cdot a' \oplus w \cdot a'' = 0\}$ is empty if $w \neq 0$ and equals $\{v\} \times F_2^r \times F_2^{n-2r}$ if $w = 0$. Corollary 4.1 gives then:
if $w \neq 0$ then $\widehat{\mathrm{f}}(b) = 0$ if $b \in F_2^r \times \{0\} \times \{0\}$ (f is then balanced);
if $w = 0$ then $\widehat{\mathrm{f}}(b) = 2^{n-r}(-1)^{u \cdot v \oplus \epsilon \oplus b' \cdot v}$ if $b = (b', 0, 0) \in F_2^r \times \{0\} \times \{0\}$.

4.1.2. *Maiorana-McFarland's functions*

An example of linear set of derivatives is when E is a vector subspace of F_2^n and when the function $a \in E \to D_a f(x)$ is linear on this vector space, that is, when $D_a D_b f(x)$ is null for every two elements a, b of E (indeed, we have $D_{a+b} f(x) \oplus D_a f(x) \oplus D_b f(x) = f(x) \oplus f(x+a) \oplus f(x+b) \oplus f(x+a+b) = D_a D_b f(x)$). This is the case of Maiorana-McFarland's functions:

Let $n = r + s$ and $f_{\phi,g}(x, y) = x \cdot \phi(y) \oplus g(y),\ x \in F_2^r,\ y \in F_2^s$. We have $D_{(u,0)} f_{\phi,g}(x, y) = u \cdot \phi(y)$ for every $u \in F_2^r$. The set $\mathcal{E} = \{u \cdot \phi(y);\ u \in F_2^r\}$ is a linear set of derivatives. We have $E = F_2^r \times \{0\}$ and $A = F_2^r \times \phi^{-1}(0)$. Corollary 4.1 gives, for every $b \in F_2^s$:

$$\widehat{\mathrm{f}_{\phi,g}}(0, b) = 2^r \sum_{y \in \phi^{-1}(0)} (-1)^{g(y) \oplus b \cdot y},$$

which allows recovering Relation (4) by changing $f_{\phi,g}$ into $f_{a+\phi,g}$.
Similarly, let $f_{\psi,\phi,g} = \sum_{i=1}^{t} x_{2i-1} x_{2i} \psi_i(y) \oplus x \cdot \phi(y) \oplus g(y)$.
We have:

$$D_{(u,0)} f_{\psi,\phi,g}(x, y) =$$

$$\sum_{i=1}^{t} (u_{2i-1} x_{2i} \oplus x_{2i-1} u_{2i} \oplus u_{2i-1} u_{2i}) \psi_i(y) \oplus u \cdot \phi(y)$$

for every $u \in F_2^r$.
Thus, the set $\mathcal{E} = \{D_{u,0} f;\ u_{2i-1} = 0,\ \forall i \leq t\}$ is a linear set of derivatives

of $f_{\psi,\phi,g}$. We have then $E = \{(u,0) \in F_2^r \times \{0\} \,/\, u_{2i-1} = 0, \forall i \leq t\}$ and Theorem 3.2 implies for instance if r is even:

$$\sum_{x \in F_2^r, y \in F_2^s} (-1)^{f_{\psi,\phi,g}(x,y)} = \sum_{(x,y) \in A} (-1)^{\bigoplus_{i=1}^t x_{2i-1}\phi_{2i-1}(y) \oplus g(y)}$$

and leads to relation (6) (but not quite as simply as for Maiorana-McFarland's functions).

4.2. *Constructing functions admitting linear sets of derivatives*

There exists an obvious relationship between the sum of two derivatives $D_a f$, $D_b f$ and the derivative $D_{a+b} f$: for every a, b in F_2^n, we have

$$D_a f(x) \oplus D_b f(x) = D_{a+b} f(x+b) = D_{a+b} f(x+a). \tag{10}$$

We deduce:

Proposition 4.1. *Let f be a Boolean function on F_2^n. Let E be a subspace of F_2^n. Assume there exists a mapping $\psi : E \times E \mapsto E$ such that*

$$\forall a, b \in E; \forall x \in F_2^n, D_b f(x+a) = D_{\psi(a,b)} f(x). \tag{11}$$

Then the set $\{D_a f,\, a \in E\}$ is a linear set of derivatives of f.

Proof: According to Relation (10), the set $\mathcal{E} = \{D_a f,\ a \in E\}$ is stable under addition, since $D_{a+b} f(x+a) = D_{\psi(a,a+b)} f(x)$. □

Notation: In the sequel, we shall write $\psi_a(b)$ instead of $\psi(a,b)$.

Remarks:
1. Relation (11) shows that, for every $c \in E$ such that there exist $a, b \in E$ such that $c = \psi_a(b)$, the derivative $D_c f$ is not only invariant under translation by c but also invariant under translation by b.
2. Condition (11) on f is equivalent to $D_a D_b f(x) = f(x+b) \oplus f(x+\psi_a(b))$. Set $\phi_a(b) = b + \psi_a(b)$. This condition can then be written $D_a D_b f(x) = D_{\phi_a(b)}(x+b)$. Since $D_a D_b f(x) = D_a D_b f(x+b)$ this is equivalent to

$$D_a D_b f = D_{\phi_a(b)} f. \tag{12}$$

Note that, if f has degree k, then the second order derivative $D_a D_b f$ has degree at most $k-2$. Thus $D_{\phi_a(b)} f$ has also degree at most $k-2$ (instead of $k-1$ which is, in general, the degree of the first order derivative of a

function of degree k).

A method for designing classes of cryptographic functions is the following:

- Find a vector subspace E of F_2^n (E can be taken equal to $F_2^r \times \{(0,\ldots,0)\}$, without loss of generality, up to a linear equivalence) and choose an homomorphism $\psi : a \mapsto \psi_a$ from E to the group of those permutations of E which fix 0, such that $\psi_a(a) = a$ for every $a \in E$.
- Characterize those Boolean functions on F_2^n such that $D_b f(x+a) = D_{\psi_a(b)} f(x)$ for every $a, b \in E$ and every $x \in F_2^n$.

All these conditions on ψ (and in particular the fact that it is an homomorphism) are not absolutely necessary for finding functions f satisfying the conditions of Proposition 4.1, but they are proper to make larger the class of functions f satisfying condition (11) when ψ is chosen, since for every Boolean function f, the function $D_0 f$ is constant and $D_a f(x+a) = D_a f(x)$. Notice that these properties of ψ imply that for every $a, a' \in E$, the vector $\psi_{a'}(a)$ is a fix point for ψ_a, since $\psi_a(\psi_{a'}(a)) = \psi_{a+a'}(a) = \psi_{a'}(\psi_a(a)) = \psi_{a'}(a)$.

4.3. *An example of construction*

We shall fix $E = F_2^r \times \{(0,\ldots,0)\}$. Note that if we take $\psi_a = id$, $\forall a \in E$, the Boolean functions on F_2^n such that $D_b f(x+a) = D_{\psi_a(b)} f(x)$ for every $a, b \in E$ and every $x \in F_2^n$ are those functions whose derivatives $D_b f$, $b \in E$ are constant on every coset of E, i.e. whose restrictions to the cosets of E are affine. These functions are nothing but those of Maiorana-McFarland (3). Hence we have to take ψ_a different from the identity, at least for some a's.

Let E' be a subspace of E. Let ϕ be any symplectic (i.e. bilinear and null on the diagonal) mapping from $E \times E$ to E'. Assume that, for every $a \in E'$, $\phi(a, b)$ is null for every $b \in E$. Then the mapping $\psi(a, b) = b + \phi(a, b)$ satisfies the desired properties. Indeed, we have $\psi(a, a) = a + \phi(a, a) = a$ and $\psi(a, 0) = \phi(a, 0) = 0$ for every $a \in E$ and we have $\phi_a \circ \phi_{a'}(b) = 0$ for every $a, a', b \in E$ and hence $\psi_a \circ \psi_{a'}(b) = \psi_{a'}(b) + \phi_a(\psi_{a'}(b)) = b + \phi_{a'}(b) + \phi_a(b) + \phi_a \circ \phi_{a'}(b) = b + \phi_{a'}(b) + \phi_a(b) = b + \phi_{a+a'}(b) = \psi_{a+a'}(b)$. Notice that Relation (11) implies that $D_b f$ is constant on every coset of E', i.e. that f is affine on every such coset, but we shall have more information on the functions than for the corresponding Maiorana-McFarland's functions. Let us now specify an example. Let $E = F_{2^r} \times \{0\}$, identified to F_{2^r} which

is therefore viewed as a subset of $F_2^n \sim F_{2^r} \times F_2^{n-r}$. Let $E' = F_2$. Note that since E' has dimension 1, knowing that we are in the framework of Maiorana-McFarland's functions gives no information. Let $i \in \{1, 2, \ldots, n-1\}$ and let $\phi(a,b) = tr(a^{2^i}b + ab^{2^i})$, where tr is the trace function from F_{2^r} to F_2. Then ϕ is clearly symplectic and $\phi(a,b) = 0$ if $b \in F_2$. We shall exclude the case "r even and $i = r/2$": ϕ being then null, this would give only Maiorana-McFarland's functions.

Let us see now what are the corresponding functions. According to Relation (12), we have $D_aD_bf(x,y) = tr(a^{2^i}b + ab^{2^i})D_1f(x,y)$ (where $D_1f(x,y) = f(x,y) \oplus f(x+1,y)$). Let $k = gcd(2i, r)$. For every $b \notin F_{2^k}$, we have $b^{2^{n-i}} + b^{2^i} \neq 0$ and there exists a such that $tr(a^{2^i}b + ab^{2^i}) = 1$. Then $D_1f(x,y) = D_aD_bf(x,y)$ is invariant by the translation $x \to x+b$ and we have $D_1f(x+b,y) = D_1f(x,y)$. Since the linear space spanned by $F_{2^r} \setminus F_{2^k}$ equals F_{2^r} (because $r \neq 2i$), we deduce that D_1f is constant on every coset $F_{2^r} \times \{y\}$ of E. If, for some y this constant is 0 then $D_aD_bf(x,y) = 0$ for every $a, b \in E$ and we obtain $f(x,y) = tr(\theta x) \oplus \epsilon$, $\forall x \in F_{2^r}$ where $\epsilon \in F_2$, $\theta \in F_{2^r}$ and $tr(\theta) = 0$ (since $D_1f(x,y) = tr(\theta)$ is the constant function 0). If this constant is 1, then $D_aD_bf(x,y) = tr(a^{2^i}b + ab^{2^i})$ for every $a, b \in E$ and we obtain $f(x,y) = tr(x^{2^i+1}) \oplus tr(\theta x) \oplus \epsilon$, $\forall x \in F_{2^r}$ where $tr(\theta) = tr(1) \oplus 1$ (since $D_1f(x,y) = tr(1) \oplus tr(\theta)$ is the constant function 1), that is $tr(\theta) = 0$ if r is odd and $tr(\theta) = 1$ if r is even. Hence we get

$$f(x,y) = h(y)\, tr(x^{2^i+1}) \oplus tr(\theta(y)\, x) \oplus g(y), \tag{13}$$

where g and h are Boolean functions on F_2^{n-r}, and where θ is a function from F_2^{n-r} to F_{2^r} such that $tr(\theta(y)) = h(y)$ if r is even and $tr(\theta(y)) = 0$ if r is odd. We obtain this way functions whose restrictions obtained by fixing y are quadratic.

These functions are similar to some of the functions of the super-class of Maiorana-McFarland and, hence, not new, up to linear isomorphism. However, we have here more insight on the Fourier spectra of these functions than for the functions of the super-Maiorana-McFarland's class.

We have $D_bf(x,y) = h(y)\, tr(bx^{2^i} + b^{2^i}x + b^{2^i+1}) \oplus tr(\theta(y)\, b)$ for every $b \in E$. The mapping $b \mapsto D_bf(x,y)$ is not linear, because of the quadratic term $tr(b^{2^i+1})$. However, the set $\{D_bf(x,y),\ b \in E\}$ is linear, thanks to the property of f with respect to the mappings ψ_a (the equality $D_af(x,y) \oplus D_bf(x,y) = D_{a+b+tr(a^{2^i}b+ab^{2^i})}f(x,y)$ can be directly checked here).

Corollary 4.1 gives

$$\forall b' \in F_2^{n-r},\ \widehat{\mathrm{f}}(0,b') = \sum_{y\in F_2^{n-r}} \sum_{x\in A_y} (-1)^{f(x,y)\oplus b'\cdot y}$$

where $A_y = \{x \in E;\ \forall b \in E, h(y)\, tr(bx^{2^i} + b^{2^i}x + b^{2^i+1}) + tr(\theta(y)\, b) = 0\}$ is equal to E if $\theta(y) = h(y) = 0$ and to $\emptyset$ otherwise since we know that if $i \neq r/2$ then the function $tr(b^{2^i+1} + \lambda b)$ is not null, whatever is λ. We deduce that

$$\widehat{\mathrm{f}}(0,b') = 2^r \sum_{y\in F_2^{n-r};\ \theta(y)=h(y)=0} (-1)^{g(y)\oplus b'\cdot y}.$$

We can deduce that, if $b \in E$ is such that $tr(b) = 0$, then

$$\forall b' \in F_2^{n-r},\ \widehat{\mathrm{f}}(b,b') = 2^r \sum_{y\in F_2^{n-r};\ \theta(y)=b \text{ and } h(y)=0} (-1)^{g(y)\oplus v\cdot y}.$$

Indeed, the condition that $tr(\theta(y)+b) = h(y)$ if r is even and $tr(\theta(y)+b) = 0$ if r is odd permits to apply the observations above to the function $f(x,y) \oplus tr(bx) = h(y)\, tr(x^{2^i+1}) \oplus tr((\theta(y) + b)\, x) \oplus g(y)$.
We give in Appendix the calculation of the complete Walsh transform of these functions.

Remarks:
1. We can use the same idea in a more general way. Assume that r is a composite number, say $r = r_1 r_2$ and take $\phi(a,b) = tr_{F_{2^r}/F_{2^{r_1}}}(a^{2^{ir_1}}b + ab^{2^{ir_1}})$ where $tr_{F_{2^r}/F_{2^{r_1}}}$ is the trace function from F_{2^r} to $F_{2^{r_1}}$ and where i is an integer, $1 \leq i \leq r_2 - 1$. Then $\phi(a,b)$ has the desired properties: it is symplectic and it is null for every $b \in F_{2^{r_1}}$ and every $a \in F_{2^r}$.
In the case $r = 2r_1$ ($r_2 = 2$ and hence $i = 1$), let w be an element of F_{2^r} such that $tr_{F_{2^r}/F_{2^{r_1}}}(w) = 1$. The ordered pair $(1, w)$ is an $F_{2^{r_1}}$-basis of F_{2^r} and we have $tr_{F_{2^r}/F_{2^{r_1}}}(a + a'w) = a'$, for every $a, a' \in F_{2^{r_1}}$. Assume $w^{2^{r_1}} = \beta w + \alpha$, then $\phi(a + a'w, b + b'w)) = tr_{F_{2^r}/F_{2^{r_1}}}((a + a'w)^{2^{r_1}}(b + b'w) + (a + a'w)(b + b'w)^{2^{r_1}}) = tr_{F_{2^r}/F_{2^{r_1}}}[(a + \alpha a' + \beta a'w)(b + b'w) + (a + a'w)(b + \alpha b' + \beta b'w)] = (a + \alpha a')b' + \beta a'b + \beta a'b'tr(w^2) + \beta ab' + a'(b + \alpha b') + \beta a'b'tr(w^2) = (a'b + ab')(1 + \beta)$. Notice that another function has also the desired properties $\phi(a + a'w, b + b'w)) = a'^{2^i}b' + a'b'^{2^i}$ where i is an integer, $1 \leq i \leq r_1 - 1$.
2. Denote by $(e_1, \ldots, e_r)$ the canonical basis of $E = F_2^r$. Let k be such that $1 \leq k \leq r$. For every $i = 1, \ldots, k$, let $\varphi_i(x,y)$ be a Boolean symplectic form on E, such that $\varphi_i(x, e_i)$ is null for every $x \in E$ and every $i \leq k$ (i.e.

$\varphi_i(x,y)$ has the form $\bigoplus_{k+1\leq j<l\leq r} a_{j,l}(x_jy_l \oplus x_ly_j)$ where $a_{j,l} \in F_2$). Define $\phi(x,y) = \bigoplus_{i=1}^{k} \varphi_i(x,y)\, e_i$. Then ϕ is a symplectic mapping from $E \times E$ to $E' = F_2^k$ (viewed as a subspace of E) and, for every $a \in E'$, $\phi(a,b)$ is null for every $b \in E$.

5. Another approach for generating functions admitting partial covering sequences

We suggest now, without going into details, another method for the design of functions for which Theorem 3.2 can be applied. In this method, the derivatives of the function are designed prior to the function itself. For applying this method, we choose a set $\mathcal{E}$ of functions whose sum $\sum_{g\in\mathcal{E}} g$ takes two values only and which satisfy a condition ensuring that they are the derivatives of a same function. This sufficient condition is given in the next Proposition.

Proposition 5.1. *Let $\mathcal{E}$ be a set of Boolean functions on F_2^n. Assume that for every $g \in \mathcal{E}$, there exists a vector $a_g \in F_2^n$ such that the function $D_{a_g}g$ is the null function. Assume that the vectors $(a_g)_{g\in\mathcal{E}}$ are linearly independent and that for every $g, h \in \mathcal{E}$, $D_{a_h}g = D_{a_g}h$.*
Then there exists a Boolean function f on F_2^n such that, for every $g \in \mathcal{E}$, we have $g = D_{a_g}f$.

Proof: Let $\mathcal{E} = \{g_1, \ldots, g_k\}$. There exists a linear isomorphism ϕ such that $(\phi(a_{g_1}), \ldots, \phi(a_{g_k})) = ((1,0,\ldots,0),(0,1,0,\ldots,0),\ldots)$. For every Boolean function g and every $a \in F_2^n$, we have $D_{\phi(a)}(g\circ\phi^{-1}) = (D_ag)\circ\phi^{-1}$. Hence, by replacing every g_i by $g_i \circ \phi^{-1}$, we may without loss of generality assume that $\{a_{g_1}, \ldots, a_{g_k}\} = \{(1,0,\ldots,0),(0,1,0,\ldots,0),\ldots\}$. We specify now f by choosing the coefficients in its ANF. Let $\prod_{i\in I} x_i$ be any monomial such that at least one index of I belongs to $\{1,\ldots,k\}$. We give to the coefficient of this monomial in the ANF of f the value of the coefficient of the monomial $\prod_{i\in I\setminus\{j\}} x_i$ in the ANF of g_j where j is any index in $I\cap\{1,\ldots,k\}$ (this value is the same for any such j, thanks to the property $D_{a_h}g = D_{a_g}h$). Then, by construction, for every $j \in \{1,\ldots,k\}$, the ANF of $D_{a_{g_j}}f$ is the part of the ANF of g_j which does not include any monomial including x_j, and since $D_{a_{g_j}}g_j = 0$, then g_j has no monomial including x_j in its ANF, and we have $D_{a_{g_j}}f = g_j$, whatever is the choice of the other coefficients in the ANF of f. □

Conclusion.

We have introduced a general principle for constructing cryptographic functions with information on their Walsh spectra. We have deduced several methods of constructions. Further work needs to be done to exploit better all these results, by specifying the constructions, in order to have more precise information on the Walsh transforms of the functions. Such work is beyond the limits of a single paper.

Partially 0-regular functions should be further studied. It would also be nice to find a general class of Boolean functions f admitting a linear set of derivatives $\mathcal{E} = \{D_a f; a \in E\}$ such that the Fourier transform of the indicator of E never vanishes. In both cases, this would allow having a complete knowledge on the Walsh spectra of the constructed functions, and therefore on the two important cryptographic parameters of the functions that are their nonlinearities and their resiliency orders.

References

1. A. Canteaut and M. Trabbia. "Improved fast correlation attacks using parity-check equations of weight 4 and 5", *Advanced in Cryptology-EUROCRYPT 2000. Lecture notes in computer science* 1807 (2000), pp. 573-588.
2. C. Carlet. A larger Class of Cryptographic Boolean Functions via a Study of the Maiorana-McFarland Construction. *Advances in Cryptology - CRYPTO'02, Lecture Notes in Computer Science* 2442, pp. 549-564 (2002).
3. C. Carlet. Concatenating indicators of flats for designing cryptographic functions. *Design, Codes and Cryptography* volume 36, Number 2, pp.89 - 202, 2005
4. C. Carlet. Boolean Functions for Cryptography and Error Correcting Codes. Chapter of the monography *Boolean Methods and Models*, Y. Crama and P. Hammer eds, Cambridge University Press, to appear. Preliminary version available at http://www-rocq.inria.fr/codes/Claude.Carlet/pubs.html
5. C. Carlet. Vectorial (multi-output) Boolean Functions for Cryptography. Chapter of the monography *Boolean Methods and Models*, Y. Crama and P. Hammer eds, Cambridge University Press, to appearever. Preliminary version available at http://www-rocq.inria.fr/codes/Claude.Carlet/pubs.html
6. C. Carlet and Y. Tarannikov. "Covering sequences of Boolean functions and their cryptographic significance". Designs, Codes and Cryptography, 25, pp. 263-279 (2002).
7. N. Courtois and W. Meier. Algebraic Attacks on Stream Ciphers with Linear Feedback. *Advances in cryptology–EUROCRYPT 2003, Lecture Notes in Computer Science* 2656, pp. 346-359, Springer, 2002.
8. F. Didier. A new upper bound on the block error probability after decoding over the erasure channel. *IEEE Transactions on Information Theory* 52, pp. 4496- 4503, 2006.

9. L.R. Knudsen. *Truncated and higher order differentials*. Fast Software Encryption, Second International Workshop, Lecture Notes in Computer Science, n 1008. pp. 196–211. – Springer-Verlag, 1995.
10. X. Lai. *Higher order derivatives and differential cryptanalysis*. Proc. "Symposium on Communication, Coding and Cryptography", in honor of J. L. Massey on the occasion of his 60'th birthday. 1994.
11. M. Matsui, "Linear cryptanalysis method for DES cipher," *Advances in Cryptology - EUROCRYPT'93, number 765 in Lecture Notes in Computer Science. Springer-Verlag*, pp. 386–397 1994.
12. Mac Williams, F. J. and N. J. Sloane (1977). *The theory of error-correcting codes*, Amsterdam, North Holland.
13. O. S. Rothaus (1976). " On bent functions", *J. Comb. Theory*, 20A, 300-305.
14. W. Meier, E. Pasalic and C. Carlet. Algebraic attacks and decomposition of Boolean functions. *Advances in Cryptology, EUROCRYPT 2004, Lecture Notes in Computer Science, Springer Verlag* 3027, pp. 474-491, 2004.
15. R. A. Rueppel *Analysis and design of stream ciphers* Com. and Contr. Eng. Series, Berlin, Heidelberg, NY, London, Paris, Tokyo 1986
16. Siegenthaler, T. "Decrypting a Class of Stream Ciphers Using Ciphertext Only". *IEEE Transactions on Computer, V. C-34*, No 1 (1985), pp. 81-85.
17. Xiao Guo-Zhen and J. L. Massey. "A Spectral Characterization of Correlation-Immune Combining Functions". *IEEE Trans. Inf. Theory*, Vol IT 34, n° 3 (1988), pp. 569-571.

6. Appendix

We study now the Walsh spectra of the functions (13) investigated at subsection 4.3:

$$f(x,y) = h(y)\, tr(x^{2^i+1}) \oplus tr(\theta(y)\, x) \oplus g(y).$$

For every $a \in F_{2^r}$ and $b \in F_2^{n-r}$, the sum $\sum_{x\in F_{2^r}, y\in F_2^{n-r}} (-1)^{f(x,y)\oplus tr(ax)\oplus b\cdot y}$ equals $\sum_{y\in F_2^{n-r}} (-1)^{g(y)\oplus b\cdot y} \left(\sum_{x\in F_{2^r}} (-1)^{h(y)\, tr(x^{2^i+1})\oplus tr(\theta(y)\, x)\oplus tr(ax)}\right)$.

If $h(y) = 0$ then $\sum_{x\in F_{2^r}} (-1)^{h(y)\, tr(x^{2^i+1})\oplus tr(\theta(y)\, x)\oplus tr(ax)}$ equals 2^r if $\theta(y) = a$ and 0 otherwise. We still have to compute the Walsh spectra of the functions $tr(x^{2^i+1}) \oplus tr(\theta(y)\, x)$. This has partially been done at subsection 4.3: when $tr(a) = 0$ and $tr(\theta(y)) = 1$ if r is even, $tr(\theta(y)) = 0$ if r is odd, we know that $\sum_{x\in F_{2^r}} (-1)^{tr(x^{2^i+1})\oplus tr(\theta(y)\, x)\oplus tr(ax)}$ is null. Keeping the same condition on $tr(\theta(y))$, we still have to compute this sum when $tr(a) = 1$. We shall do it for r odd and under the condition that r and i are co-prime. Let us denote by λ the sum $\sum_{x\in F_{2^r}} (-1)^{tr(x^{2^i+1})\oplus tr(x)}$. For every $u \in F_{2^r}$ we have $\sum_{x\in F_{2^r}} (-1)^{tr((x+u)^{2^i+1})\oplus tr(x+u)} = \lambda$. Thus,

$\sum_{x\in F_{2^r}}(-1)^{tr(x^{2^i+1}+u^{2^i}x+ux^{2^i}+u^{2^i+1}+x+u)}=\lambda$ and hence

$$\sum_{x\in F_{2^r}}(-1)^{tr(x^{2^i+1})\oplus tr((u^{2^i}+u^{2^{n-i}}+1)x)}=\lambda(-1)^{tr(u^{2^i+1}+u)}.$$

Since r and $2i$ are co-prime, the set $\{u^{2^i}+u^{2^{n-i}};\ u\in F_{2^r}\}$ equals $\{b\in F_{2^r};\ tr(b)=0\}$. Moreover, the mapping $u\mapsto u^{2^i}+u^{2^{n-i}}$ is one to one from $\{u\in F_{2^r};\ tr(u)=0$ to itself. Thus we know the value of the sum $\sum_{x\in F_{2^r}}(-1)^{tr(x^{2^i+1})\oplus tr(ax)}$ for every $a\in F_{2^r}$ such that $tr(a)=1$. And we aleready know that if $tr(a)=0$ then $\sum_{x\in F_{2^r}}(-1)^{tr(x^{2^i+1})\oplus tr(ax)}$ is null. We deduce, thanks to Parseval's relation, that $2^{n-1}\lambda^2=2^{2n}$ and thus that $\lambda=\pm 2^{(n+1)/2}$.

Let us denote $\theta(y)=\theta_1^{2^i}(y)+\theta_1^{2^{n-i}}(y)$, where $\theta_1(y)\in F_{2^r}$. We deduce from the computations above that $\sum_{x\in F_{2^r}}(-1)^{tr(x^{2^i+1})\oplus tr(\theta(y)\,x)\oplus tr(ax)}$ equals $(-1)^{tr(u^{2^i+1}+u)}\lambda$ if $a=u^{2^i}+u^{2^{n-i}}+1$ and equals 0 if $tr(a)=0$.

Thus, if $tr(a)=0$, then the sum: $\sum_{x\in F_{2^r},y\in F_2^{n-r}}(-1)^{f(x,y)\oplus tr(ax)\oplus b\cdot y}$ equals $2^r\sum_{y\in F_2^{n-r};\,h(y)=0;\,\theta_1(y)=0}(-1)^{g(y)\oplus b\cdot y}$ and, if $tr(a)=1$, $a=u^{2^i}+u^{2^{n-i}}+1$, then it equals $\lambda\sum_{y\in F_2^{n-r};h(y)=1}(-1)^{g(y)\oplus tr(\theta_1^{2^i+1}(y)+\theta_1(y)+b\cdot y}$.

Obviously this result can be generalized to cases where $\theta(y)$ has not always null trace. Write $\theta(y)=\theta_1^{2^i}(y)+\theta_1^{2^{n-i}}(y)+\eta(y)$, where $\theta_1(y)\in F_{2^r}$, $\eta(y)\in F_2$. We have

$$\sum_{x\in F_{2^r},y\in F_2^{n-r}}(-1)^{f(x,y)\oplus tr(ax)\oplus b\cdot y}=$$

$$2^r\sum_{y\in F_2^{n-r};\,h(y)=0;\,\eta(y)=tr(a);\,\theta_1(y)=u}(-1)^{g(y)\oplus b\cdot y}+$$

$$\lambda\sum_{y\in F_2^{n-r};h(y)=1\,\eta(y)=tr(a)\oplus 1}(-1)^{g(y)\oplus tr(\theta_1^{2^i+1}(y)+\theta_1(y)+u^{2^i+1}+u)+b\cdot y}$$

where $a=u^{2^i}+u^{2^{n-i}}+tr(a)$, $tr(u)=0$.

Non linéarité des fonctions booléennes données par des traces de polynômes de degré binaire 3

Eric Férard

Université de Polynésie française, Tahiti
E-mail: ferard@upf.pf

François Rodier

Institut de Mathématiques de Luminy,
C.N.R.S. 163 avenue de Luminy,
Case 907, Marseille Cedex 9, France
E-mail: rodier@iml.univ-mrs.fr

Nous étudions la non linéarité des fonctions définies sur $\mathbb{F}_{2^m}$ où m est un entier impair, associées aux polynômes de degré 7 ou à des polynômes plus généraux. Nous en déduisons un critère pour que des fonctions vectorielles ne soient pas APN.

English extended abstract - Boolean functions are an important tool in computer sciences. They are especially useful in private key cryptography for designing stream ciphers. For security reasons, and also because Boolean functions need also to have other properties than nonlinearity such as balancedness or high algebraic degree, it is important to have the possibility of choosing among many Boolean functions, not only bent functions, that is functions with the highest possible non linearity, but also functions which are close to be bent in the sense that their nonlinearity is close to the nonlinearity of bent functions. For m odd, it would be particularly interesting to find functions with nonlinearity larger than the one of quadratic Boolean functions (called *almost optimal* in [2]). This has been done for instance in the work of Patterson and Wiedemann [19] and also of Langevin-Zanotti [13] and more recently by Kavut, Maitra and Yücel [14].

Let $q = 2^m$ and $\mathbb{F}_{2^m}$ assimilated as a vector space to $\mathbb{F}_2^m$. In this article, we want to study functions of the form $\mathrm{Tr}\, G(x)$, where G is a polynomial on $\mathbb{F}_{2^m}$ and Tr the trace of $\mathbb{F}_{2^m}$ over $\mathbb{F}_2$.

For m even, many people got interested in finding bent functions of this form. To only mention the case of monomials, one can get the known cases (Gold , Dillon/Dobbertin, Niho exponents) in the paper of Leander [12].

For m odd, one might have expected that among the functions $f : x \longrightarrow \mathrm{Tr}\, G(x)$ where G is a polynomial of degree 7, there are some functions which

are close to being bent in the previous sense. This happens not to be the case, but we will show that for m odd such functions have rather good nonlinearity or autocorrelation properties. We use for that recent results of Maisner and Nart [16] about zeta functions of supersingular curves of genus 2.

On the other hand, vectorial Boolean functions are used in cryptography to construct block ciphers. An important criterion on these functions is a high resistance to the differential cryptanalysis. Nyberg [18] has introduced the notion of almost perfect nonlinearity (APN) to study differential attacks. We relate this notion to the notion above, and we will give some criterion for a function not to be almost perfect nonlinear.

Mots clé : fonction booléenne, non linéarité, indice de somme des carrés, courbe supersingulière, fonction APN de genre 2.

1. Introduction

La non-linéarité d'une fonction booléenne $f : \mathbb{F}_2^m \longrightarrow \mathbb{F}_2$ est la distance de f à l'ensemble des fonctions affines à m variables (voir le § 2.2). C'est un concept important qui intervient en cryptographie (cf. [3,5,6,8]) pour construire des cryptosystèmes performants (chiffrements symétriques), et dans la théorie de codage avec le vieux problème du rayon de recouvrement des codes de Reed-Muller d'ordre 1.

La non-linéarité est inférieure à $2^{m-1} - 2^{m/2-1}$. Cette limite est atteinte par les fonctions courbes (cf. le livre de MacWillams et de Sloane [17]) qui existent seulement si le nombre de variables m des fonctions booléennes est pair. Pour des raisons de sécurité en cryptographie, et aussi parce que les fonctions booléennes doivent avoir d'autres propriétés telles que l'équilibre ou le degré algébrique élevé, il est important d'avoir la possibilité de choix parmi beaucoup de fonctions booléennes, non seulement des fonctions courbes, mais également des fonctions presque courbes dans le sens que leur non-linéarité est voisine de la non-linéarité des fonctions courbes.

Pour m impair, il serait particulièrement intéressant de trouver des fonctions avec une non-linéarité plus grande que celle de fonctions booléennes quadratiques (appelées *presque optimales* dans [2]). Ceci a été fait dans le travail de Patterson et de Wiedemann [19] et également de Langevin et Zanotti [13] et plus récemment par Kavut, Maitra et Yücel [14].

Soit $q = 2^m$ et $k = \mathbb{F}_{2^m}$ assimilé comme espace vectoriel sur $\mathbb{F}_2$ à $\mathbb{F}_2^m$. Si G est un polynôme sur k, cela nous permet de construire une fonction booléenne $\operatorname{Tr} G(x)$, où Tr est la trace de $\mathbb{F}_{2^m}$ sur $\mathbb{F}_2$, ou plutôt la fonction $\chi(G(x))$, avec des valeurs dans ± 1, où nous dénotons par χ_0 le caractère non trivial unique de $\mathbb{F}_2$ dans les nombres complexes différents de zéro :

$$\chi_0(0) = 1 \quad , \quad \chi_0(1) = -1$$

et nous notons $\chi = \chi_0 \circ \mathrm{Tr}$.

Pour m pair, on a cherché à trouver des fonctions courbes de cette forme. Pour mentionner seulement le cas des monômes, on peut considérer les cas connus (de Gold, de Dillon, des exposants de Niho) dans l'article de Leander [12]. Ce sont des fonctions $f : x \longrightarrow \chi(ax^r)$ où $r = 3$ ou 5 (ou plus généralement $r = 2^i + 1$, où i est un nombre entier) et $a \in k$ n'est pas de la forme x^r.

On aurait pu espérer que pour $r = 7$, ou parmi les fonctions

$$f : x \longrightarrow \chi\left(G(x)\right)$$

quand G est un polynôme du degré 7, il y a quelques fonctions qui sont presque courbes au sens précédent. Cela s'avère ne pas être le cas, mais nous prouverons que pour m impair de telles fonctions ont les propriétés de non-linéarité plutôt bonnes (cf. section 4). Nous employons pour cela des résultats récents de Maisner et de Nart au sujet des fonctions de zêta des courbes supersingulières de genre 2 que nous avons regroupés dans les sections 6, 7, 8.

D'autre part, les fonctions booléennes vectorielles sont utilisées en cryptographie pour construire des algorithmes de chiffrements par bloc. Un critère important sur ces fonctions est une résistance élevée à la cryptanalyse différentielle. Nyberg [18] a défini la notion de non-linéarité presque parfaite (APN) pour étudier les attaques différentielles. Nous ramenons cette notion à la notion ci-dessus, et nous donnons un critère pour qu'une fonction ne soit pas presque parfaitement non-linéaire.

2. Préliminaires

2.1. *Fonctions booléennes*

Soit m un entier positif et $q = 2^m$.

Définition 2.1. Une fonction booléenne à m variables est une application de l'espace $V_m = (\mathbb{F}_2)^m$ dans $\mathbb{F}_2$.

Une fonction booléenne est *linéaire* si c'est une forme linéaire sur l'espace vectoriel $(\mathbb{F}_2)^m$. Elle est dite *affine* si elle est égale à une fonction linéaire à une constante près.

2.2. *Non-linéarité*

Définition 2.2. On appelle non-linéarité d'une fonction booléenne f à m variables et on la note $\mathrm{nl}(f)$ la distance qui la sépare de l'ensemble des

fonctions affines à m variables :

$$\mathrm{nl}(f) = \min_{h \text{ affine}} d(f, h)$$

où d est la distance de Hamming.

On peut prouver que la non-linéarité est égale à

$$\mathrm{nl}(f) = 2^{m-1} - \frac{1}{2}\|\widehat{f}\|_\infty$$

où

$$\|\widehat{f}\|_\infty = \sup_{v \in V_m} \Big| \sum_{x \in V_m} \chi_0(f(x) + v \cdot x) \Big|$$

et $v \cdot x$ dénote le produit scalaire usuel de V_m. C'est le maximum de la transformée de Fourier de $\chi_0\,(f)$ (ou la transformée de Walsh de f):

$$\widehat{f}(v) = \sum_{x \in V_m} \chi_0(f(x) + v \cdot x).$$

On appellera $\|\widehat{f}\|_\infty$ l'amplitude spectrale de la fonction booléenne f. La formule d'inversion est donnée par

$$\chi_0(f(x)) = \frac{1}{q} \sum_{v \in V_m} \widehat{f}(v)\chi_0\,(v.x)$$

où l'on remarque que le dual de V_m est isomorphe à V_m, avec la mesure $\frac{1}{q}$ sur chaque point. L'identité de Parseval peut s'écrire

$$\|\widehat{f}\|_2^2 = \frac{1}{q} \sum_{v \in V_m} \widehat{f}(v)^2 = q$$

et, si f est une fonction booléenne sur $\mathbb{F}_2^m$:

$$\sqrt{q} \leq \|\widehat{f}\|_\infty \leq q.$$

2.3. *L'indice de somme des carrés*

Soit f une fonction booléenne sur V_m. Zhang et Zheng ont introduit l'indice de somme des carrés [28]:

$$\sigma_f = \frac{1}{q} \sum_{x \in V_m} \widehat{f}(x)^4 = \|\widehat{f}\|_4^4.$$

Nous remarquons que

$$\|\widehat{f}\|_2 \leq \|\widehat{f}\|_4 \leq \|\widehat{f}\|_\infty. \tag{1}$$

La relation de cette fonction avec la non linéarité a été étudiée par A. Canteaut et al. [2].

3. Les fonctions $f : x \longrightarrow \mathrm{Tr}\,(G(x))$ où G est un polynôme

3.1. *Divisibilité de* $\|\widehat{f}\|_\infty$

Soit $G(x)$ le polynôme $\sum_{i=0}^{s} a_i x^i$ à coefficients dans $\mathbb{F}_q$ et f la fonction booléenne $\mathrm{Tr} \circ G$.

Définition 3.1. Le degré binaire de G est la valeur maximum des $\sigma(i)$ pour $0 \leq i \leq s$, où $\sigma(i)$ est la somme des chiffres de i écrit en chiffre binaire.

On a la proposition suivante, due à C. Moreno et O. Moreno [15], généralisant le théorème d'Ax.

Proposition 3.2. *Soit G un polynôme à coefficients dans $\mathbb{F}_q$, de degré binaire d. Alors $\|\widehat{f}\|_\infty$ est divisible par $2^{\lceil \frac{m}{d} \rceil}$.*

3.2. *Cas où G est un polynôme de degré binaire 2*

Les $\|\widehat{f}\|_\infty$ sont multiples de $2^{\lceil \frac{m}{2} \rceil}$. Donc, si m est pair $\|\widehat{f}\|_\infty$ est un multiple de $q^{1/2}$, et si m est impair, de $\sqrt{2q}$. En particulier, si m est impair, l'amplitude spectrale est supérieure ou égale à $\sqrt{2q}$ qui est égale à celle des fonctions booléennes quadratiques de rang maximal.

4. Les fonctions $f : x \longrightarrow \mathrm{Tr}\,(G(x))$ où G est un polynôme de degré binaire 3

On va simplement étudier le cas où G est un polynôme de degré binaire 2 auquel on a rajouté un monôme non nul de degré 7, c'est-à-dire un polynôme de la forme

$$G(x) = a_7 x^7 + \sum_{0}^{s} b_i x^{2^i+1}$$

où $a_7 \neq 0$ un polynôme de degré 7 à coefficients dans k. Nous voudrions évaluer $\|\widehat{f}\|_4$ sur $\mathbb{F}_{2^m}$, pour $f(x) = \mathrm{Tr}\,(G(x))$ où Tr dénote la fonction trace de $\mathbb{F}_q$ vers $\mathbb{F}_2$:

$$\mathrm{Tr}(x) = \sum_{i=0}^{m-1} x^{2^i}.$$

4.1. *Evaluation de* $\|\widehat{f}\|_4^4$

Proposition 4.1. *La valeur de* $\|\widehat{f}\|_4^4$ *sur* $\mathbb{F}_{2^m}$ *quand* m *est impair et* $f(x) = \chi(G(x))$ *est telle que*

$$|\|\widehat{f}\|_4^4 - 3q^2| \leq 185.2^{s-1}q^{3/2}.$$

Démonstration –
La démonstration sera donné dans la section 7.

Remarque 4.2. Ce résultat est à comparer avec la proposition 5.6 de [20] où on a montré que la distribution de $\|\widehat{f}\|_4^4$ pour toutes les fonction booléennes est concentrée autour de $3q^2$.

4.2. *Bornes de* $\|\widehat{f}\|_\infty$

La démonstration de ces bornes seront données dans la section 8.

4.2.1. *Borne inférieure*

Proposition 4.3. *Pour les fonctions* $f : x \longrightarrow \chi(G(x))$ *sur* $\mathbb{F}_{2^m}$ *où* G *est un polynôme donné au début de la section 4 et* m *est impair, on a, pour* $m \leq 13 + 2s$

$$\sqrt{2q} \leq \|\widehat{f}\|_\infty.$$

Pour $m \geq 15 + 2s$*, on a de plus*

$$\sqrt{2q} + 2^{\lceil \frac{m}{3} \rceil} \leq \|\widehat{f}\|_\infty.$$

Remarque 4.4. Il est connu que pour m impair et plus petit que 7, on a $\sqrt{2q} \leq \|\widehat{f}\|_\infty$ pour toutes les fonctions booléennes [11].

Remarque 4.5. Hou [10] a montré que pour m impair et plus petit que 14, on a $\sqrt{2q} \leq S(f)$ pour les fonctions booléennes de degré binaire 3.

4.2.2. *Borne supérieure*

Proposition 4.6. *Pour les fonctions* $f : x \longrightarrow \chi(G(x))$ *sur* $\mathbb{F}_{2^m}$ *où* G *est un polynôme donné au début de la section 4, on a pour* $s \leq 2$

$$\|\widehat{f}\|_\infty \leq 6\sqrt{q}$$

et pour $s \geq 3$

$$\|\widehat{f}\|_\infty \leq 2^s\sqrt{q}.$$

5. Fonctions presque parfaitement non-linéaires

Considérons une fonction $F : \mathbb{F}_q \longrightarrow \mathbb{F}_q$.

Définition 5.1. La fonction F est dite APN (presque parfaitement non-linéaire) si pour tout $a \in \mathbb{F}_q^*$ et $b \in \mathbb{F}_q$, il existe au plus 2 éléments de $\mathbb{F}_q$ tels que $F(z+a) + F(z) = b$.

Les fonction APN sont celles qui résistent le mieux aux attaques différentielles.

Proposition 5.2. *La fonction*

$$G : \mathbb{F}_q \longrightarrow \mathbb{F}_q$$
$$x \longmapsto a_7 x^7 + \sum_0^s b_i x^{2^i+1}$$

n'est pas APN pour $m \geq 13 + 2s$.

Démonstration –

Pour $\gamma \in \mathbb{F}_q$, considérons la fonction $f_\gamma(x) = \mathrm{Tr}(G(\gamma x))$. La proposition se déduit de la proposition 4.1 et du résultat suivant de Chabaud-Vaudenay [9].

Proposition 5.3. *On a* $\sum_{\gamma \in k^*} \sigma(f_\gamma) \geq 2q^2(q-1)$. *De plus la fonction* G *est APN si et seulement si l'égalité est vérifiée.*

Remarque 5.4. Pour $s \leq 2$, on peut dire plus. La fonction G n'est pas APN pour $m \geq 11$, d'après le théorème 4.1 de [22] qui donne des exemples de fonctions qui ne sont pas APN.

6. Etude de courbes hyperelliptiques

Pour démontrer les résultats précédents, on va étudier des courbes liées au polynôme G.

On obtient d'abord l'expression simple de $\|\widehat{f}\|_4$ (cf. [20,21]):

$$\|\widehat{f}\|_4^4 = \sum_{x_1+x_2+x_3+x_4=0} \chi\left(f(x_1)+f(x_2)+f(x_3)+f(x_4)\right) = q^2 + \sum_{\substack{\alpha \neq 0 \\ \alpha \in V_m}} X_\alpha$$

avec

$$X_\alpha = \Big(\sum_{x \in \mathbb{F}_q} \chi\left(G(x)+G(x+\alpha)\right) \Big)^2.$$

On note maintenant α un élément de $\mathbb{F}_q^*$. On peut vérifier que

$$G(x+\alpha)+G(x) = G(\alpha) + a_7\alpha^6 x + a_7\alpha^5 x^2 + a_7\alpha^4 x^3 + a_7\alpha^3 x^4 + a_7\alpha^2 x^5 + a_7\alpha x^6 + \sum_0^s b_i(\alpha x^{2^i} + \alpha^{2^i} x).$$

Pour calculer X_α, on peut remarquer que la courbe d'équation $y^2+y = G(x+\alpha)+G(x)$ est isomorphe à

$$\begin{aligned} y^2+y &= G(\alpha)+ \\ &+\Big(a_7\alpha^6 + a_7^{1/4}\alpha^{3/4} + a_7^{1/2}\alpha^{5/2} + \sum_0^s (b_i\alpha)^{2^{-i}} + \sum_0^s b_i\alpha^{2^i}\Big)x + \\ &+(a_7\alpha^4 + a_7^{1/2}\alpha^{1/2})x^3 + a_7\alpha^2 x^5 \end{aligned}$$

qui est une équation de la courbe C_1 de genre 2 pour $\alpha \neq 0$.

On a

$$X_\alpha = (\#C_1 - q - 1)^2.$$

6.1. *La théorie de van der Geer et van der Vlugt*

Soit C_1 la courbe d'équation affine:

$$C_1 : y^2 + y = ax^5 + bx^3 + cx + d$$

avec $a \neq 0$. Soit R le polynôme linéaire $ax^4 + bx^2 + c^2x$. L'application

$$\begin{aligned} Q : \mathbb{F}_q &\longrightarrow \mathbb{F}_2 \\ x &\longmapsto \mathrm{Tr}(xR(x)) \end{aligned}$$

est la forme quadratique associé à la forme symplectique

$$\begin{aligned} \mathbb{F}_q \times \mathbb{F}_q &\longrightarrow \mathbb{F}_2 \\ (x,y) &\longmapsto <x,y> = \mathrm{Tr}(xR(y) + yR(x)). \end{aligned}$$

Le nombre de zéros de Q détermine le nombre de points de C_1:

$$\#C_1(\mathbb{F}_q) = 1 + 2\#Q^{-1}(0).$$

Le radical W de la forme symplectique $<,>$ concide avec l'ensemble des zéros dans $\mathbb{F}_q$ du polynôme $\mathbb{F}_2$-linéaire et séparable

$$E_{a,b} = a^4x^{16} + b^4x^8 + b^2x^2 + ax.$$

On a : $0 \leq w = dim_{\mathbb{F}_2} W \leq 4$ et $w \equiv m \pmod 2$. La codimension du noyau V de Q dans W est égale à 0 ou 1. De plus, le polynôme $E_{a,b}$ se factorise dans $\mathbb{F}_q[x]$ ([26], Theorem 3.4):

$$E_{a,b}(x) = xP(x)(1 + x^5 P(x))$$

avec $P(x) = a^2x^5 + b^2x + a$.

Théorème 6.1. *(van der Geer - van der Vlugt [26])*
Si $V \subset W$, alors $\#C_1(\mathbb{F}_q) = 1 + q$.
Si $V = W$, alors $\#C_1(\mathbb{F}_q) = 1 + q \pm \sqrt{2^w q}$.

6.2. *Les travaux de Maisner et Nart*

Supposons que $a = b$ et que le polynôme P ait au moins une racine z. Alors, comme m est impair, il existe un unique $\ell \in \mathbb{F}_q$ tel que $\ell^3 = 1 + z^{-4}$.

Proposition 6.2. *Si $\operatorname{Tr} \ell = 0$ alors le polynôme P a exactement trois racines dans $\mathbb{F}_q$ et on a $w = 3$. Si $\operatorname{Tr} \ell \neq 0$ alors le polynôme P n'a qu'une racine dans $\mathbb{F}_q$, la composante restante est irréductible et on a $w = 1$.*

Démonstration –
Voir Maisner et Nart [16] propositions 2.3 et 2.6.

6.3. *Réduction de la courbe $y^2 + y = G(x + \alpha) + G(x)$*

Soit $\lambda = \alpha + a_7^{-1/4}\alpha^{-3/4}$.

6.3.1. *Cas où $\lambda = 0$*

Alors on a $\alpha^7 = a_7^{-1}$, donc l'équation de la courbe devient

$$\begin{aligned} y^2 + y = G(\alpha) + (a_7\alpha^6 + a_7^{1/4}\alpha^{3/4} + a_7^{1/2}\alpha^{5/2} + \sum_0^s b_i(\alpha^{2^{-i}} + \alpha^{2^i}))x + \\ + a_7\alpha^2 x^5 \\ = d + cx + ax^5 \end{aligned}$$

pour $a = \alpha^{-5}$.

Le polynôme P s'écrit $P(x) = a^2x^5 + a$. Si m est impair il a une unique racine $z = a^{-1/5} = \alpha$. D'après Maisner et Nart ([16], Propositions 2.5 et 2.3) on est dans le cas où $w = 1$ donc $W = \{0, z\}$. Soit c le coefficient de x. On a

$$\operatorname{Tr}(cz) = \operatorname{Tr}\left((1/\alpha + a_7^{1/4}\alpha^{3/4} + a_7^{1/2}\alpha^{5/2} + \sum_0^s (b_i\alpha)^{2^{-i}} + \sum_0^s b_i\alpha^{2^i})\alpha\right)$$

$$= \mathrm{Tr}(1 + \sum_0^s b_i^{2^{-i}} \alpha^{\frac{2^i+1}{2^i}} + \sum_0^s b_i \alpha^{1+2^i})$$
$$= \mathrm{Tr}\, 1 = 1.$$

On vérifie alors que

$$Q(z) = \mathrm{Tr}\,(az^5 + cz) = \mathrm{Tr}\,(1 + cz) = 0.$$

D'où $V = W$ et donc $X_\alpha = 2q$ par le théorème 6.1.

6.3.2. *Cas où* $\lambda \neq 0$

Cette courbe est isomorphe à

$$y^2 + y = ax^5 + ax^3 + cx + d$$

avec

$$a = \lambda^5 a_7 \alpha^2 = \lambda^3 (a_7 \alpha^4 + a_7^{1/2} \alpha^{1/2})$$

et $\lambda = \alpha + a_7^{-1/4} \alpha^{-3/4}$. On a

$$a = 1 + a_7^{-1/4} \alpha^{-7/4} + a_7^{3/4} \alpha^{21/4} + a_7 \alpha^7 \tag{2}$$

et

$$c = 1 + \Big(\sum_0^s (b_i \alpha)^{2^{-i}} + \sum_0^s b_i \alpha^{2^i} \Big) \lambda + {a_7}^{1/2} \alpha^{7/2} + a_7^{3/4} \alpha^{21/4} + a_7 \alpha^7. \tag{3}$$

6.4. *Valeurs de* X_α

Proposition 6.3. *Supposons que m soit impair. Alors* $X_\alpha = 0$, $2q$ *ou* $8q$. *ßSoit* $\ell = a_7^{-1/3} \alpha^{-7/3}$. *Alors*

$X_\alpha = 8q$ *si et seulement si*

$$\mathrm{Tr}\,\ell = 0 \quad , \quad \ell = v + v^4 \quad ,$$
$$\mathrm{Tr}\,\big(\eta v^3\big) = 1 \quad , \quad \mathrm{Tr}\,\big(\eta(v + v^2)\big) = 1 \quad ;$$
$$\textit{avec} \quad \eta = 1 + \sum_0^s (b_i \alpha^{1+2^i})^{2^{-i}} + \sum_0^s b_i \alpha^{1+2^i} +$$
$$+ {a_7}^{1/2} \alpha^{7/2} + a_7^{1/4} \alpha^{7/4}; \tag{4}$$

$X_\alpha = 2q$ *si et seulement si* $\mathrm{Tr}\,\ell = 1$;

$X_\alpha = 0$ *dans les cas restant.*

Démonstration –

Si $\lambda = 0$, alors $\ell = 1$ d'où $\operatorname{Tr}\ell = 1$. On a bien $X_\alpha = 2q$ d'après 6.3.1.

Si $\lambda \neq 0$, on étudie le polynôme $P = a^2x^5 + a^2x + a$. Remarquons que $z = \lambda^{-1}\alpha$ est racine de P. Donc

$$\begin{aligned} P &= (x+z)(a^2x^4 + a^2x^3z + a^2x^2z^2 + a^2xz^3 + a^2z^4 + a^2) \\ &= a^2z^{-4}(x+z)(x^4z^{-4} + x^3z^{-3} + x^2z^{-2} + xz^{-1} + z^{-4} + 1). \end{aligned}$$

La décomposition de P en composante irréductibles dépend de $e = 1+z^{-4}$. On a

$$e = 1+z^{-4} = 1+\lambda^4\alpha^{-4} = 1+(\alpha^4+a_7^{-1}\alpha^{-3})\alpha^{-4} = 1+(1+a_7^{-1}\alpha^{-7}) = a_7^{-1}\alpha^{-7}.$$

Comme m est impair, on a $\mathbb{F}_q^3 = \mathbb{F}_q$. Soit $\ell = e^{1/3}$. Alors, d'après la proposition 6.2, on a

$$\begin{cases} w = 1 \text{ si } \operatorname{Tr}\ell = 1; \\ w = 3 \text{ si } \operatorname{Tr}\ell = 0. \end{cases}$$

D'après le théorème 6.1, on a

dans le premier cas, $X_\alpha = 0 \quad$ ou $\quad 2q$;

dans le deuxième cas, $X_\alpha = 0 \quad$ ou $\quad 8q$.

Premier cas, $\mathbf{Tr\,\ell = 1}$. On a $W = \{0, z\}$ et

$$Q(z) = \operatorname{Tr}(az^5 + az^3 + cz) = \operatorname{Tr}(az + cz + 1)$$

car $\operatorname{Tr}(az^3) = 0$. Pour que $X_\alpha = 0$ il faut et il suffit que $\operatorname{Tr}(a+c)z = 0$. Des équations (2) et (3) on déduit

$$\begin{aligned} (a+c)z &= 1 + a_7^{1/4}\alpha^{7/4} + a_7^{1/2}\alpha^{7/2} + \Big(\sum_0^s b_i\alpha^{2^{-i}} + \sum_0^s b_i\alpha^{2^i}\Big)\alpha \\ &= 1 + a_7^{1/4}\alpha^{7/4} + a_7^{1/2}\alpha^{7/2} + \Big(\sum_0^s (b_i\alpha^{1+2^i})^{2^{-i}} + \sum_0^s b_i\alpha^{1+2^i}\Big). \end{aligned}$$

Donc $\operatorname{Tr}((a+c)z) = \operatorname{Tr}1 = 1$ et $X_\alpha = 2q$.

Deuxième cas, $\mathbf{Tr\,\ell = 0}$. On a $W =< z, z_1, z_2 >$.

Pour que $X_\alpha = 0$ il faut et il suffit que $\operatorname{Tr}(a+c)z_i = 0$ pour l'un des $i = 1, 2$ ou que $\operatorname{Tr}(a+c)z = 0$.

Les nombres z_i sont racines de $x^4z^{-4}+x^3z^{-3}+x^2z^{-2}+xz^{-1}+z^{-4}+1 = 0$. On a $e = 1+z^{-4} = \ell^3$ et $\ell = u+u^2$. D'où, d'après Maisner et Nart [16] (démonstration du lemme 2.4):

$x^4z^{-4} + x^3z^{-3} + x^2z^{-2} + xz^{-1} + z^{-4} + 1 =$

$$(x^2z^{-2} + uxz^{-1} + (1+u)^3)(x^2z^{-2} + (u+1)xz^{-1} + u^3).$$

On peut supposer $\operatorname{Tr} u = 0$ (car $\operatorname{Tr} 1 = 1$, donc u ou $1+u$ a une trace nulle). Soit donc $u = v + v^2$. On a par conséquent $\ell = v + v^4$. Alors le polynôme $x^2z^{-2} + (u+1)xz^{-1} + u^3$ est réductible: ses racines sont: $z(v(1+u)+1) = z(v(1+v+v^2)+1) = z(v^3+v+v^2+1)$ et $z(v(1+u)+u) = zv^3$.

6.5. *Calcul du nombre des α donnés par la proposition 6.3*

On peut évaluer le nombre des α qui donnent chaque cas de la proposition précédente.

6.5.1. *Le nombre des α tels que $X_\alpha = 2q$*

D'abord, on évalue le nombre des α tels que $\operatorname{Tr} \ell = 1$ dans la proposition 6.3.

Proposition 6.4. *Le nombre N_{2q} des valeurs de α telles que $X_\alpha = 2q$ vérifie*

$$\left| N_{2q} - \frac{q}{2} \right| < 3q^{1/2}.$$

Démonstration –

On a $\operatorname{Tr} \ell = \operatorname{Tr}(a_7^{-1/3}\alpha^{-7/3})$. Le nombre de α dans $\mathbb{F}_q^*$ tels que $\operatorname{Tr}(a_7^{-1/3}\alpha^{-7/3}) = 1$ est égal au nombre N_{2q} de x dans $\mathbb{F}_q^*$ tels que $\operatorname{Tr}(a_7^{-1/3}x^7) = 1$. Définissons

$$S_{2q} = \sum_{x \in \mathbb{F}_q} \chi(a_7^{-1/3}x^7) = (q - N_{2q}) - N_{2q} = q - 2N_{2q}.$$

On a $|S_{2q}| < 6\sqrt{q}$ d'où

$$\frac{q - 6\sqrt{q}}{2} \leq N_{2q} = \frac{q - S_{2q}}{2} \leq \frac{q + 6\sqrt{q}}{2}.$$

6.5.2. *Une courbe auxiliaire*

On a besoin d'évaluer le nombre des (α, v) vérifiant certaine conditions, avec v tel que $v + v^4 = \ell = a_7^{-1/3}\alpha^{-7/3}$. Soit $x^{-3} = \alpha$ et $a_7^{-1/3} = \gamma$.

Proposition 6.5. *On considère la courbe C donnée par l'équation*

$$v + v^4 = \gamma x^7$$

avec les coordonnées x et v et le modèle non singulier $\tilde{C}$. Alors le morphisme $\tilde{C} \longrightarrow C$ est bijectif. La courbe a un unique point à l'infini. Elle est de

genre 9. Les valuations au point $(0,0)$ *sont* $v_{(0,0)}(x) = 1$ *et* $v_{(0,0)}(v) = 7$. *Les valuations au point à l'infini sont* $v_\infty(x) = -4$ *et* $v_\infty(v) = -7$.

Démonstration –
Voir le livre de Stichtenoth [24] p. 200.

6.5.3. *Bornes pour les sommes exponentielles*

Sur la courbe $\tilde{C}$, on considère une fonction rationnelle f, qui n'est pas de la forme $\phi^2 + \phi$, avec ϕ un fonction rationnelle sur $\tilde{C}$. Soit

$$S = \sum_{z \in \tilde{C}_0(\mathbb{F}_q)} {}'\chi\left(f(z)\right)$$

où la somme est définie sur les points rationnels sur $\mathbb{F}_q$ de $\tilde{C}$, qui ne sont pas des pôles de f. Soit $(f)_\infty$ le diviseur des pôles de f et t le nombre de pôles de f, sans multiplicité. La proposition suivante donne une borne pour les sommes exponentielles S.

Proposition 6.6. *On a*

$$|S| \leq (2g - 2 + t + \deg(f)_\infty)\sqrt{q}.$$

Démonstration –
Voir l'article de Bombieri, [4].

6.5.4. *Le nombre des* (α, v) *tels que* $\mathrm{Tr}\left(\eta v^3\right) = 1$

On évalue le nombre des (α, v) tels que $\mathrm{Tr}\left(\eta v^3\right) = 1$, où η est donné par (4).

On a

$$\begin{aligned}
&\mathrm{Tr}\left(\eta v^3\right)\\
&= \mathrm{Tr}\left(v^3 + \textstyle\sum_0^s v^3 (b_i \alpha^{1+2^i})^{2^{-i}} + \sum_0^s v^3 b_i \alpha^{1+2^i} + v^3 \sqrt{a_7}\alpha^{7/2} + v^3 a_7^{1/4}\alpha^{7/4}\right)\\
&= \mathrm{Tr}\left(v^3 + \textstyle\sum_0^s (v^{3.2^i} + v^3) b_i x^{-3-3.2^i} + (v^6 + v^{12}) a_7 x^{-21}\right).
\end{aligned}$$

Sur la courbe C, on considère la fonction

$$f(x) = v^3 + \sum_0^s (v^{3.2^i} + v^3) b_i x^{-3-3.2^i} + (v^6 + v^{12}) a_7 x^{-21}.$$

Pour vérifier que f n'est pas de la forme $\phi^2 + \phi$, on considère

$$\psi = \gamma^{1/4} \frac{x}{v} (v^3 x^{-3})^{2^{i-2}} = \gamma^{1/4} (v x^{-1})^{3.2^{i-2}-1}.$$

Si $i \geq 2$, on a

$$(v^{3.2^i}+v^3)x^{-3-3.2^i}+\psi^4+\psi = \Big(\frac{x^{-3}(\gamma x^7+v)+\gamma x^4}{v^4}\Big)(v^3x^{-3})^{2^i}+v^3x^{-3-3.2^i}+\psi$$
$$= (x^{-3}v^{-3})(v^3x^{-3})^{2^i}+v^3x^{-3-3.2^i}+\psi.$$

Et sa valuation à l'infini est donnée par

$$v_\infty\Big((v^{3.2^i}+v^3)x^{-3-3.2^i}+\psi^4+\psi\Big) = v_\infty\Big((x^{-3}v^{-3})(v^3x^{-3})^{2^i}+v^3x^{-3-3.2^i}+\psi\Big)$$
$$= 33-9.2^i.$$

si $i \geq 3$. C'est un entier négatif impair.

En faisant de même pour chaque entier i dans l'expression de f, on trouve une fonction ψ telle que la valuation au point à l'infini de $f+\psi^2+\psi$ soit un entier impair négatif.

On peut vérifier que la fonction f est définie sur chaque point fini de C sauf peut-être en les points tels que $x=0$.

On considère la somme

$$S_1 = \sum_{(x,v)\in C(\mathbb{F}_q)-C_\infty} \chi(f)$$

où $C_\infty = \{(0,0),(0,1),(0,\beta),(0,\beta^2),\infty\}$ et β est une racine primitive $3^{\text{ème}}$ de l'unité. Les pôles de f ne peuvent être que parmi les points dans C_∞. La valuation de f à l'infini est

$$v_\infty(f) \geq \inf(v_\infty(v^3), v_\infty(b_s x^{-3(1+2^s)}v^{3.2^s})) \geq -9.2^s+12$$

si $s \geq 2$. La valuation de f en $(0,0)$ est

$$v_{(0,0)}(f) = v_{(0,0)}(v^3x^{-3-3.2^s}) = 21-3(1+2^s) = 18-3.2^s.$$

La valuation de $v^{3.2^i}+v^3$ en $(0,1)$ est

$$v_{(0,1)}(v^{3.2^i}+v^3) = v_{(0,1)}\Big((v^3+1)\prod_{\delta\in\mathbb{F}_{2^i}-\{1\}}(v^3-\delta)\Big) = v_{(0,1)}\Big(\frac{x^7}{v}\Big) = 7.$$

La valuation de $(v^{3.2^i}+v^3)x^{-3-3.2^i}$ en $(0,1)$ est donc

$$v_{(0,1)}(v^{3.2^i}+v^3)x^{-3-3.2^i} = 7-3-3.2^i = 4-3.2^i.$$

La valuation de v^6+v^{12} en $(0,1)$ est

$$v_{(0,1)}(v^6+v^{12}) = 2v_{(0,1)}(1+v^3) = 2v_{(0,1)}(x^7) = 14.$$

La valuation de $(v^6+v^{12})x^{-21}$ en $(0,1)$ est

$$v_{(0,1)}((v^6+v^{12})x^{-21}) = 14-21 = -7.$$

La valuation de f en $(0,1)$ est finalement

$$v_{(0,1)}(f) = \inf(4 - 3.2^s, -7) = 4 - 3.2^s$$

si $4 - 3.2^s < -7$ c'est-à-dire si $s \geq 2$.

Le même calcul vaut pour la valuation de f en $(0,\beta)$. On a donc, pour $s \geq 2$:

$$\deg(f)_\infty = -3(4 - 3.2^s) - (18 - 3.2^s) - 12 + 9.2^s = -42 + 21.2^s$$

et

$$|S_1| \leq (18 - 2 + 5 - 42 + 21.2^s)q^{1/2} = (21.2^s - 21)q^{1/2}.$$

Considérons sur la courbe C le nombre N_1 des couples (α, v) tels que $\mathrm{Tr}\left(\eta v^3\right) = 1$. Alors

$$S_1 = \sum_{(x,v)\in C - C_\infty} \chi\left(f\right) = \sum_{\mathrm{Tr}\, f = 0} 1 - N_1 = \#C - 2N_1 - 5$$

où $\#C$ est le nombre des points de la courbe C. Donc

$$\left|N_1 - \frac{\#C}{2}\right| = \frac{|S_1 + 5|}{2} \leq \frac{21.2^s - 21}{2} q^{1/2} + 5/2.$$

6.5.5. *Le nombre des* (α, v) *tels que* $\mathrm{Tr}\left(\eta(v^2 + v)\right) = 1$

Ensuite, nous évaluons le nombre des (α, v) tels que $\mathrm{Tr}\left(\eta(v^2 + v)\right) = 1$, où η est donné par (4).

$$\begin{aligned}
&\mathrm{Tr}\left(\eta(v^2 + v)\right) \\
&= \mathrm{Tr}\left((a_7^{1/4}\alpha^{7/4} + a_7^{1/2}\alpha^{7/2} + (\textstyle\sum_0^s (b_i\alpha^{1+2^i})^{2^{-i}} + \sum_0^s b_i\alpha^{1+2^i}))(v^2 + v)\right) \\
&= \mathrm{Tr}\left(a_7\gamma^2 x^{-7} + \textstyle\sum_0^s (b_i x^{-3(1+2^i)})(v^{2^{i+1}} + v^{2^i} + v^2 + v)\right).
\end{aligned}$$

On définit la fonction $g(x) = a_7\gamma^2 x^{-7} + \sum_0^s (b_i x^{-3(1+2^i)})(v^{2^{i+1}} + v^{2^i} + v^2 + v)$. Elle n'est pas de la forme $\phi^2 + \phi$ parce que avec $\psi = \gamma^{1/2} x^{2-3.2^{s-1}}$, on a

$$\begin{aligned}
v_{0,0}(x^{-3(1+2^s)}v + \psi^2 + \psi) &= v_{0,0}(x^{-3(1+2^s)}v + \gamma(x^{4-3.2^s}) + \gamma^{1/2}x^{2-3.2^{s-1}}) \\
&= v_{0,0}(x^{-3.2^s}(x^{-3}(v^4 + \gamma x^7) + \gamma x^4) + \gamma^{1/2}x^{2-3.2^{s-1}}) \\
&\geq \inf(-3.2^s + 25, 2 - 3.2^{s-1})
\end{aligned}$$

d'où

$$v_{0,0}(x^{-3(1+2^s)}(v^2 + v) + \psi^2 + \psi) = -3.2^s + 11$$

car $-3.2^s + 11 < -3.2^s + 25$ et $-3.2^s + 11 < 2 - 3.2^{s-1}$ si $3 < 2^{s-1}$ c'est-à-dire si $s \geq 3$. Si $s = 2$, on obtient le même résultat. En tout état de cause, $v_{0,0}(x^{-3(1+2^s)}v + \psi^2 + \psi)$ est un entier impair négatif.

La valuation de $x^{-21}(v^8+v^2)=x^{-7}$ en ∞ est

$$v_\infty(x^{-21}(v^8+v^2))=84-7.8=28.$$

La valuation de $(b_ix^{-3(1+2^i)})(v^{2^{i+1}}+v^{2^i}+v^2+v)$ en ∞ est

$$v_\infty\Big(x^{-3(1+2^i)}(v^{2^{i+1}}+v^{2^i}+v^2+v)\Big)=12(1+2^i)-7.2^{i+1}=-2.2^i+12.$$

Donc la fonction g a pour valuation à l'infini

$$v_\infty(g)=-2.2^i+12.$$

La valuation de $x^{-21}(v^8+v^2)=x^{-7}$ en $(0,0)$, ..., $(0,\beta^2)$ est

$$v_{(0,0)}(x^{-21}(v^8+v^2))=-21+7.2=-7.$$

La valuation de $(b_ix^{-3(1+2^i)})(v^{2^{i+1}}+v^{2^i}+v^2+v))$ en $(0,0)$ est

$$-3(1+2^i)+7=4-3.2^i.$$

La valuation de g en $(0,0)$ est

$$v_{(0,0)}(g)=4-3.2^s$$

si $4-3.2^s<-7$, c'est-à-dire si $\frac{11}{3}<2^s$ c'est-à-dire si $s\geq 2$.

La valuation de $(v^2+v))$ en $(0,1)$ est

$$v(v^2+v)=v\Big(\frac{v^4+v}{1+v+v^2}\Big)=v\Big(\frac{x^7}{1+v+v^2}\Big)=7.$$

La valuation de $(b_ix^{-3(1+2^i)})(v^{2^{i+1}}+v^{2^i}+v^2+v))$ en $(0,1)$ est

$$v_{(0,0)}(x^{-3(1+2^i)})(v^{2^{i+1}}+v^{2^i}+v^2+v))=-3(1+2^i)+7=4-3.2^i.$$

La valuation de $v^{2^{i+1}}+v^{2^i}+v^2+v$ en $(0,\beta)$ est

$$\begin{aligned}v_{(0,\beta)}(v^{2^{i+1}}+v^{2^i}+v^2+v)&=v_{(0,\beta)}((v^2+v)^{2^i}+v^2+v)\\&=v_{(0,\beta)}(v^2+v+1)\\&=7.\end{aligned}$$

La valuation de $(b_ix^{-3(1+2^i)})(v^{2^{i+1}}+v^{2^i}+v^2+v))$ en $(0,\beta)$ est

$$v_{(0,\beta)}\Big((b_ix^{-3(1+2^i)})(v^{2^{i+1}}+v^{2^i}+v^2+v)\Big)=-3(1+2^i)+7=4-3.2^i.$$

Donc la valuation de g en $(0,1)$, $(0,\beta)$, $(0,\beta^2)$ est

$$v_{(0,v)}(g)=4-3.2^s$$

si $4-3.2^s<-7$, c'est-à-dire si $\frac{11}{3}<2^s$ c'est-à-dire si $s\geq 2$.

Calculons maintenant

$$S_2 = \sum_{(x,v)\in C(\mathbb{F}_q)-C_\infty} \chi(g).$$

La valuation de g en chacun de ces points finis est supérieure à la plus faible des valuations de $(v^2-v^8)/x^{21}$ et $(b_i x^{-3(1+2^i)})(v^{2^{i+1}}+v^{2^i}+v^2+v)$, elle est donc plus grande que $4-3.2^s$. La valuation de g à l'infini est supérieure à la plus faible des valuations de $(v^2-v^8)/x^{21}$ et $(b_i x^{-3(1+2^i)})(v^{2^{i+1}}+v^{2^i}+v^2+v)$, elle est donc plus grande que $12-2^{s+1}$. Donc

$$\deg(g)_\infty \leq 4(-4+3.2^s)-12+2^{s+1} = 14.2^s - 28.$$

Par conséquent, on a

$$|S_2| \leq (18-2+5+14.2^s-28)q^{1/2} = 7(2^{s+1}-1)q^{1/2}.$$

Soit N_2 le nombre des couples (α, v) tels que $\mathrm{Tr}\left(\eta(v^2+v)\right)=1$. Alors

$$S_2 = \sum_{(x,v)\in C-C_\infty} \chi(g) = \sum_{\mathrm{Tr}\, g=0} 1 - N_2 = \#C - 2N_2 - 5$$

car $\#C_\infty = 5$. Donc

$$\left|N_2 - \frac{\#C}{2}\right| = \frac{|S_2+5|}{2} \leq \frac{7}{2}(2^{s+1}-1)q^{1/2} + \frac{5}{2}.$$

6.5.6. *Le nombre des* (α, v) *tels que* $\mathrm{Tr}\left(\eta(v^2+v)\right) = \mathrm{Tr}\left(\eta v^3\right)$

Ensuite, nous évaluons le nombre des (α, v) tels que $\mathrm{Tr}\left(\eta(v^2+v)\right) = \mathrm{Tr}\left(\eta v^3\right)$ c'est-à-dire $\mathrm{Tr}\left(\eta(v^3+v^2+v)\right)=0$. Nous avons à calculer le nombre des (x, v) tels que

$$\mathrm{Tr}(g(x)+f(x)) = 0.$$

On considère la somme

$$S_3 = \sum_{(x,v)\in C(\mathbb{F}_q)-C_\infty} \chi(f+g).$$

Pour vérifier que $f+g$ n'est pas de la forme $\phi^2+\phi$, il suffit de calculer valuation en $(0,0)$ de $f+g+b_s\phi^2+b_s^{1/2}\phi$ comme dans la sous-section précédente (6.5.5). On a

$$v_{(0,0)}f = 18-3.2^s \quad \text{et} \quad v_{(0,0)}(g+b_s\phi^2+b_s^{1/2}\phi) = 11-3.2^s.$$

On obtient dans tous les cas une valuation impaire négative.

Par l'analyse précédente, on a

$$\deg(f+g)_\infty = 21.2^s - 63 + 14.2^s - 28 = 35.2^s - 91.$$

Donc on a

$$|S_3| \leq (18 - 2 + 5 + 35.2^s - 91)q^{1/2} = (35.2^s - 70)q^{1/2}.$$

Soit N_3 le nombre des couples (α, v) tels que $\mathrm{Tr}\left(\eta(v^3 + v^2 + v)\right) = 0$. Alors

$$S_3 = \sum_{(x,v)\in C - C_\infty} \chi\left(f + g\right) = N_3 - \sum_{\mathrm{Tr}\, f+g=1} 1 = 2N_3 - \#C + 5$$

car $\#C_\infty = 5$. Donc

$$\left|N_3 - \frac{\#C}{2}\right| = \frac{|S_3 + 5|}{2} \leq \frac{1}{2}(35.2^s - 70)q^{1/2} + \frac{5}{2}.$$

6.5.7. *Le nombre des α tels que $X_\alpha = 8q$*

Nous avons besoin d'un lemme.

Lemme 6.7. *Soient deux fonctions ϕ et ψ définies sur un ensemble fini X à valeurs dans $\mathbb{F}_2$. Supposons que*

$$\begin{aligned} \#\{x : \phi(x) = 0\} &= N_1; \\ \#\{x : \psi(x) = 0\} &= N_2; \\ \#\{x : \phi(x) = \psi(x)\} &= N_3. \end{aligned}$$

Alors

$$\#\{x : \phi(x) = \psi(x) = 0\} = \frac{1}{2}(N_1 + N_2 + N_3 - N)$$

où N est le nombre d'éléments de X.

Démonstration –

Posons

$$\begin{aligned} \{x : \phi(x) = \psi(x) = 0\} = N_{0,0} \quad &, \quad \{x : \phi(x) = 0, \psi(x) = 1\} = N_{0,1} \quad , \\ \{x : \phi(x) = 1, \psi(x) = 0\} = N_{1,0} \quad &, \quad \{x : \phi(x) = \psi(x) = 1\} = N_{1,1} \quad . \end{aligned}$$

On a

$$N_{0,0} + N_{0,1} = N_1 \quad , \quad N_{0,0} + N_{1,0} = N_2 \quad , \quad N_{0,0} + N_{1,1} = N_3$$

La somme des $N_{i,j}$ étant égale à N, on a donc

$$\begin{aligned} N = \sum N_{i,j} = N_{0,0} + (N_1 - N_{0,0}) + (N_2 - N_{0,0}) + (N_3 - N_{0,0}) \\ = N_1 + N_2 + N_3 - 2N_{0,0}. \end{aligned}$$

D'où

$$N_{0,0} = \frac{N_1 + N_2 + N_3 - N}{2}.$$

Proposition 6.8. *Le nombre N_{8q} des valeurs de α telles que $X_\alpha = 8q$ vérifie*

$$\left|N_{8q} - \frac{q}{8}\right| < 23.2^{s-1}q^{1/2}$$

pour $q \geq 32$.

Démonstration –

D'après la proposition 6.3, il faut calculer le nombre N' des points (x, v) tels que $\mathrm{Tr}\left(\eta v^3\right) = 1$ et $\mathrm{Tr}\left(\eta(v^2 + v)\right) = 1$. D'après le lemme 6.7, ce nombre vérifie

$$\begin{aligned} N' &= \frac{1}{2}(N_1 + N_2 + N_3 - \#C) \\ &= \frac{1}{2}\Big(N_1 - \frac{\#C}{2} + N_2 - \frac{\#C}{2} + N_3 - \frac{\#C}{2}\Big) + \frac{\#C}{4} \end{aligned}$$

et on a

$$\begin{aligned} &\left|N' - \tfrac{\#C}{4}\right| \\ &= \left|\tfrac{1}{2}\left(N_1 - \tfrac{\#C}{2} + N_2 - \tfrac{\#C}{2} + N_3 - \tfrac{\#C}{2}\right)\right| \\ &\leq \tfrac{1}{2}\left(\tfrac{21.2^s - 21}{2}q^{1/2} + 5/2 + \tfrac{7}{2}(2^{s+1} - 1)q^{1/2} + \tfrac{5}{2} + \tfrac{1}{2}(35.2^s - 70)q^{1/2} + \tfrac{5}{2}\right) \\ &\leq (15/4 - 25q^{1/2} + 91.2^{(s-2)}q^{1/2}). \end{aligned}$$

Comme pour chaque α tel que $\mathrm{Tr}\left(\eta v^3\right) = 1$ et $\mathrm{Tr}\left(\eta(v^2 + v)\right) = 1$ il y a deux valeurs de v (soit v et $v + 1$), le nombre N_{8q} de tels α vérifie donc

$$\left|N_{8q} - \frac{\#C}{8}\right| \leq (15/8 - 25/2.q^{1/2} + 91.2^{(s-3)}q^{1/2}).$$

Comme m est impair, il existe une solution de $v + v^4 = \gamma x^7$ si et seulement si la trace $\mathrm{Tr}(\gamma x^7)$ est nulle et, dans ce cas, il y a exactement deux solutions. Donc

$$\#C(\mathbb{F}_q) = 2\#\{\mathrm{Tr}(\gamma x^7) = 0\} + 1 = S_7 + q + 1$$

où S_7 est la somme exponentielle $S_7 = \sum_{x \in \mathbb{F}_q} (-1)^{\mathrm{Tr}(\gamma x^7)}$. Donc

$$|\#C(\mathbb{F}_q) - q - 1| \leq 6\sqrt{q}.$$

On a

$$\left|N_{8q} - \frac{q}{8}\right| \leq \left|N_{8q} - \frac{\#C}{8}\right| + \left|\frac{\#C}{8} - \frac{q}{8} - \frac{1}{8}\right| + \frac{1}{8}.$$

Donc le nombre N_{8q} vérifie

$$\left|N_{8q} - \frac{q}{8}\right| \leq 15/8 - 25/2.q^{1/2} + 91.2^{(s-3)}q^{1/2} + \frac{3}{4}q^{1/2} + \frac{1}{8} \leq 23.2^{s-1}q^{1/2}.$$

7. Démonstration de l'évaluation de $\|\widehat{f}\|_4^4$ (proposition 4.1)

On déduit facilement de la proposition 6.8 le calcul de la valeur de $\|\widehat{f}\|_4^4$. Sachant que

$$\left|N_{8q} - \frac{q}{8}\right| \leq 23.2^{s-1}q^{1/2},$$
$$\left|N_{2q} - \frac{q}{2}\right| \leq 3q^{1/2} + 1,$$

calculons

$$\begin{aligned}\|\widehat{f}\|_4^4 &= q^2 + \sum_{\substack{\alpha\neq 0\\ \alpha\in V_m}} X_\alpha \\ &= 3q^2 + 8q(N_{8q} - q/8) + 2q(N_{2q} - q/2).\end{aligned}$$

D'où

$$\begin{aligned}|\|\widehat{f}\|_4^4 - 3q^2| &\leq 8q\left|N_{8q} - \frac{q}{8}\right| + 2q\left|N_{2q} - \frac{q}{2}\right| \\ &\leq 185.2^{s-1}q^{3/2}.\end{aligned}$$

8. Démonstration des bornes de $\|\widehat{f}\|_\infty$ (propositions 4.3 et 4.6)

8.1. *Borne inférieure*

L'évaluation du nombre des α tels que $\operatorname{Tr}\ell = 1$ dans la proposition 6.3 donne:

$$2q^2 - 6q^{3/2} \leq \|\widehat{f}\|_4^4.$$

On a

$$\sum_{\alpha\in\mathbb{F}_q^*} X_\alpha \geq 2qN_{8q} \geq 2q\frac{q - 6\sqrt{q}}{2} = q^2 - 6q^{3/2}$$

et

$$\|\widehat{f}\|_4^4 = q^2 + \sum_{\alpha\in\mathbb{F}_q^*} X_\alpha \geq 2q^2 - 6q^{3/2}.$$

Comme il est facile de montrer que

$$\|\widehat{f}\|_4^4 \leq q\|\widehat{f}\|_\infty^2$$

nous obtenons $2q - 6q^{1/2} \leq \|\widehat{f}\|_\infty^2$, donc $\sqrt{2q} - 3\sqrt{2} \leq \|\widehat{f}\|_\infty$, d'où le résultat si $m \geq 7$, parce que $\|\widehat{f}\|_\infty$ est un entier divisible par $2^{\lceil m/3\rceil}$. Le résultat pour $m \leq 7$ est connu (cf. remarque 4.4).

On a, de plus

$$\|\widehat{f}\|_4^4 \geq 3q^2 - 185.2^{s-1}q^{3/2}$$

par la proposition 4.1. On en déduit que pour que $\|\widehat{f}\|_4^4$ dépasse $2q^2$, il suffit que $m \geq 15+2s$. Pour des raisons de divisibilité, $\|\widehat{f}\|_\infty$ est alors plus grand que $\sqrt{2q} + 2^{\lceil \frac{m}{3} \rceil}$.

8.2. *Borne supérieure*

On a, d'après la borne de Weil

$$|\widehat{f}(v)| = \Big| \sum_{x \in V_m} \chi_0(f(x) + v \cdot x) \Big| \leq (\deg f - 1)\sqrt{q}.$$

References

1. Anne Canteaut *Differential cryptanalysis of Feistel ciphers and differentially δ-uniform mappings,* in Selected Areas on Cryptography, SAC'97, pages 172-184, Ottawa, Canada, 1997.
2. A. Canteaut, C. Carlet, P. Charpin, C. Fontaine *Propagation characteristics et correlation-immunity of highly nonlinear Boolean functions,* Advances in cryptology, EUROCRYPT 2000 (Bruges), 507–522, Lecture Notes in Comput. Sci., Vol. 1807, Springer, Berlin, 2000.
3. P. Barthélémy, R. Rolland, P. Véron *Cryptographie*, Hermès, Paris, 2005. N
4. E. Bombieri, *On exponential sums in finite fields.* Amer. J. Math., 88, 1966, pp. 71-105.
5. C. Carlet, *On cryptographic complexity of Boolean functions*, Proceedings of the Sixth Conference on Finite Fields with Applications to Coding Theory, Cryptography et Related Areas (G.L. Mullen, H. Stichtenoth et H. Tapia-Recillas Eds), Springer (2002) pp. 53-69.
6. C. Carlet, *On the algebraic thickness et non-normality of Boolean functions, with developments on symmetric functions*, submitted to IEEE Trans. Inform. Theory.
7. G. Cohen, I. Honkala, S. Litsyn, A. Lobstein, *Covering codes.* North-Holland Mathematical Library, 54, North-Holland Publishing Co., Amsterdam (1997).
8. C. Fontaine, *Contribution à la recherche de fonctions booléennes hautement non linéaires et au marquage d'images en vue de la protection des droits d'auteur*, Thèse, Université Paris VI (1998).
9. Chabaud, Florent; Vaudenay, Serge *Links between differential and linear cryptanalysis.* De Santis, Alfredo (ed.), Advances in cryptology - EUROCRYPT '94. Workshop on the theory and application of cryptographic techniques, Perugia, Italy, May 9-12, 1994. Proceedings. Berlin: Springer-Verlag. Lect. Notes Comput. Sci. 950, 356-365 (1995).

10. X. Hou, On the covering radius of $R(1,m)$ in $R(3,m)$. IEEE Trans. Inform. Theory 42 (1996), no. 3, 1035–1037.
11. X. Hou, Covering radius of the Reed-Muller code $R(1,7)$—a simpler proof. J. Combin. Theory Ser. A 74 (1996), no. 2, 337–341.
12. Leander, Nils Gregor Monomial bent functions. IEEE Trans. Inform. Theory 52 (2006), no. 2, 738–743.
13. Langevin, P.; Zanotti, J.-P. Nonlinearity of some invariant Boolean functions. Des. Codes Cryptogr. 36 (2005), no. 2, 131–146.
14. Selçuk Kavut, Subhamoy Maitra and Melek D. Yücel There exist Boolean functions on n (odd) variables having nonlinearity $> 2^{n-1} - 2^{\frac{n-1}{2}}$ if and only if $n > 7$, prépublication, `http://eprint.iacr.org/2006/181`
15. C. Moreno et O. Moreno The MacWilliams-Sloane conjecture on the tightness of the Carlitz-Uchiyama bound and the weights of duals of BCH codes. IEEE Trans. Inform. Theory 40 (1994), no. 6, 1894–1907.
16. Daniel Maisner et Enric Nart, *Zeta functions of supersingular curves of genus 2*, arXiv:math.NT/0408383
17. F.J. MacWilliams et N.J.A. Sloane, *The Theory of Error-Correcting Codes,* North-Holland, Amsterdam (1977).
18. Nyberg, Kaisa Differentially uniform mappings for cryptography. Advances in cryptology—EUROCRYPT '93 (Lofthus, 1993), 55–64, Lecture Notes in Comput. Sci., 765, Springer, Berlin, 1994.
19. N. Patterson et D. Wiedemann, *The covering radius of the* $(2^{15}, 16)$ *Reed-Muller code is at least* 16 276, IEEE Trans. Inform. Theory 29, no. 3 (1983), 354-356.
20. F. Rodier, *Sur la non-linéarité des fonctions booléennes*, Acta Arithmetica, vol 115, (2004), 1-22, prépublication: arXiv: math.NT/0306395.
21. F. Rodier, *On the nonlinearity of Boolean functions*, Proceedings of WCC2003, Workshop on coding et cryptography 2003 (D. Augot, P. Charpin, G. Kabatianski eds), INRIA (2003), pp. 397-405.
22. F. Rodier, Borne sur le degré des polynômes presque parfaitement nonlinéaires; prépublication. Disponible dans ArXiv: math.AG/0605232, 2006.
23. J.-P. Serre, *Majorations de sommes exponentielles.* Journées Arithmétiques de Caen (Univ. Caen, Caen, 1976), pp. 111-126. Astérisque No. 41-42, Soc. Math.France, Paris, 1977.
24. H. Stichtenoth, Algebraic Function Fields et Codes, Springer, 1993.
25. P. Stănică, *Nonlinearity, local et global avalanche characteristics of balanced Boolean functions*, Discrete Math. 248 (2002), no. 1-3, 181–193.
26. G. van der Geer, M. van der Vlugt, *Reed-Muller codes and supersingular curves. I*, Compositio Math. 84, (1992), 333-367.
27. G. van der Geer, M. van der Vlugt, *Supersingular Curves of Genus 2 over finite fields of Characteristic 2* , Math. Nachr. 159, (1992), 73-81.
28. Xian-Mo Zhang and Yuliang Zheng, *GAC —the Criterion for Global Avalanche Characteristics of Cryptographic Functions*, Journal of Universal Computer Science, vol. 1, no. 5 (1995), 316-333

On Exponents with highly divisible Fourier Coefficients and Conjectures of Niho and Dobbertin

Gregor Leander

Institut de Mathématiques de Toulon,
Université du Sud Toulon-Var, France
E-mail: gregor.leander@ruhr-uni-bochum.de

Philippe Langevin

Institut de Mathématiques de Toulon,
Université du Sud Toulon-Var, France
E-mail: langevin@univ-tln.fr

We describe a sieving algorithm to enumerate all the power functions having Fourier coefficients divisible by a large power of 2 up to dimension 33. This algorithm enable us to check an important conjecture of Dobbertin concerning the non-existence of almost bent monomials. Furthermore we give an update on some conjectures stated by Niho.

1. Almost bent exponent

Let L be an extension of degree m of $\mathbb{F}_2$, and let q be the order of L. The *Fourier coefficient* at $a \in L$ of the power function $f(x) = x^d$ of an exponent d is defined by the exponential sum

$$\widehat{f}_d(a) = \sum_{x \in L} \mu_L(x^d + ax)$$

where μ_L is the additive canonical character of L defined by $\mu_L(x) = (-1)^{\mathrm{Tr}_L(x)}$ with $\mathrm{Tr}_L(x) = x + x^2 + \cdots + x^{2^{m-1}}$. The *spectrum* of d is the set $\mathrm{spec}\,(d) = \{\widehat{f}_d(a) \mid a \in L\}$. Note that $\mathrm{spec}\,(d) = \mathrm{spec}\,(2d)$ because the trace is invariant under the Fröbenius automorphism of L. All along this note, we assume that the exponent d is coprime to $q-1$. In this case, it is easy to prove that $\mathrm{spec}\,(d) = \mathrm{spec}\,(d^{-1})$, and for this reason, the exponent d' is said equivalent to d if there exists an integer k such that $d' \equiv 2^k d$ or $dd' \equiv 2^k$ modulo $q-1$. By a bound of Sidelnikov [18],

$$\sup_{a \in L} |\widehat{f}_d(a)| \geq \sqrt{2q} \tag{1}$$

An exponent achieving this lower bound is called *almost bent*. Of course, such exponents exist only when m is odd. As a consequence of Parseval's identity

$$q^2 = \sum_{a \in L} \widehat{f}_d(a)^2$$

the spectrum of an almost bent exponent is

$$\{0, \pm 2^{\frac{m+1}{2}}\}.$$

The largest integer ν such that 2^ν divides all the Fourier coefficients of d is denoted by $\mathrm{val}\,(d)$. In the literature, an exponent d is said to be *almost perfect nonlinear* when all the equations :

$$(x+a)^d + x^d = b, \quad a \in L^\times, b \in L$$

have at most two solutions. These notions are connected in the sense that

Lemma 1.1. *An exponent d is almost bent if and only it is almost perfect non-linear with* $\mathrm{val}\,(d) \geq \frac{m+1}{2}$.

For a proof, see [4].

2. Dobbertin's conjecture

¿From now and on, we assume that m is odd. It is well known that the spectrum of the exponents

$$2^r + 1 \quad \text{(Gold)}, \qquad 2^{2r} - 2^r + 1 \quad \text{(Kasami)},$$

are three valued. More precisely, denoting by e the greatest common divisor of r and m, the spectrums of these exponents are

$$\{-2^{\frac{m+e}{2}}, \quad 0, \quad +2^{\frac{m+e}{2}}\}$$

In particular, there are $\varphi(m)/2$ classes of almost bent exponents of each of the above types. Remark that these classes are not distinct, namely we have $2^2 - 2^1 + 1 = 2^1 + 1$. For $m > 9$ this is the only exponent which is both Gold and Kasami. Based of numerical experiments, Niho conjectured that the following exponents are almost bent :

$$2^{\frac{m-1}{2}} + 3 \quad \text{(Welch)}, \qquad \text{and} \qquad 2^{2r} + 2^r - 1 \quad \text{(Niho)}$$

where $4r \equiv -1 \pmod{m}$.

This conjecture has been proved three decades later by Dobbertin, Canteaut, Charpin, Xiang and Hollmann. According to Lemma 1.1, their proofs

split in two distinct parts to obtain the divisibility condition [2,3,10], and the APN-property [7,8] of these two exponents.

Table 1. Known almost bent exponents m odd.

type	s	condition	nb. classes
Gold	2^r+1	$(r,m)=1)$	$\frac{1}{2}\varphi(m)$
Kasami	$2^{2r}-2^r+1$	$(r,m)=1)$	$\frac{1}{2}\varphi(m)$
Welch	$2^{(m-1)/2}+3$		1
Niho	$2^{2r}+2^r-1$	$4r\equiv -1 \mod m$	1

Dobbertin conjectures claims Table 1 is already complete.

Conjecture 2.1 (Dobbertin). *Up to equivalence, if $m>9$ then the number of almost bent exponents is equal to $\varphi(m)+1$.*

The main purpose of this paper is to check Dobbertin's conjecture for m as big as possible. For this we developed an algorithm, described in the next section, that is able to compute a list of exponents with valuation greater than $\frac{m+1}{2}$. As a side effect this algorithm allows us to give an update on some conjectures stated by Niho. Finally we prove that the exponent $d=13/3$ has valuation $\frac{m+1}{2}$ for all odd m.

3. Sieving algorithm

The Fourier coefficient in a of $f(x)=x^d$ can be developed in term of Gauss sums (see [16])

$$\widehat{f}(a)=\frac{q}{q-1}+\frac{1}{q-1}\sum_{1\neq\chi\in\widehat{L^\times}} G_L(\chi)G_L(\bar{\chi}^d)\chi^d(a) \tag{2}$$

It is a good idea to identify the finite field L with the residual field of the unramified extention of degree m of the the field of dyadic numbers. Indeed, in this case, there exists a multiplicative character ω, the Teichmüller character of L, such that the Stickelberger's congruences holds i.e.

$$\forall j,\quad 0\leq j<q-1,\qquad G_L(\bar{\omega}^j,\mu_L)\equiv 2^{\mathrm{wt}\,(j)} \mod 2^{\mathrm{wt}\,(j)+1}$$

where $\mathrm{wt}\,(j)$ is the sum of the bits in the binary expansion of the smallest non negative residue of j modulo 2^m-1. Using this relation in the equality (2), we obtain

$$\mathrm{val}\,(d) \geq \nu_d = \min_{1 \leq j \leq q-2} \mathrm{wt}\,(-j) + \mathrm{wt}\,(jd) \tag{3}$$

Let X be a non empty set of multiplicative characters. The orthogonality relation of characters shows that $\sum_{\chi \in X} \chi \not\equiv 0 \pmod 2$, in particular, since we assume d coprime to $q-1$, the equality $\nu_d = \mathrm{val}\,(d)$ holds, see [15] for details. As it is illustrated by Table 2, the exponents having a valuation greater than or equal to $\frac{m+1}{2}$ seem very rare.

Table 2. Valuation distribution of all the invertible exponents ($d \not\sim 1$) in dimension $m = 25$. Most of the exponents have a valuation less than $\frac{m+1}{2}$.

ν	0	1	2	3	4	5	6	7
nb.	0	0	1	12	155	1549	11396	68348
ν	8	9	10	11	12	13	14	15
nb.	260754	287221	18228	249	8	79	0	3
ν	16	17	18	19	20	21	22	23
nb.	0	0	0	0	0	0	0	0

Our strategy to check Dobbertin's conjecture consists in enumerating the *good exponents* d such that $\mathrm{val}\,(d) \geq \frac{m+1}{2}$. It is a set containing the AB-exponents, and if this set of candidates is not too large, it will be possible to decide which are almost bent in computing the Fourier spectrums.The main idea of the algorithm is the following lemma.

Lemma 3.1. *An exponent d has valuation smaller then $\frac{m+1}{2}$ if and only if there exist integers s, r such that*

$$sd = -r \bmod (2^m - 1), \qquad \mathrm{wt}\,(r) + \mathrm{wt}\,(s) \leq \frac{m-1}{2},$$

Proof. If for a given d we find s, r such that the conditions are fulfilled then

$$\mathrm{val}\,(d) \leq \mathrm{wt}\,(s) + \mathrm{wt}\,(-sd) = \mathrm{wt}\,(s) + \mathrm{wt}\,(r) < \frac{m+1}{2}.$$

On the other hand, if d has valuation smaller then $\frac{m-1}{2}$ then, as explained above, we know that there exists an element j such that

$$\mathrm{wt}\,(j) + \mathrm{wt}\,(-jd) < \frac{m+1}{2}$$

and then the values $s = j$ and $r = -jd$ fulfill the conditions of the lemma. □

Our sieving algorithm consists in generating all the pairs (r, s) of residues such that

$$\mathrm{wt}\,(s) \leq \mathrm{wt}\,(r), \quad \mathrm{wt}\,(s) + \mathrm{wt}\,(r) \leq \frac{m-1}{2}.$$

and marking as *bad* all exponents that fulfill the equation $sd = -r \bmod (2^m - 1)$. By Lemma (3.1), the marked exponents are not AB and all the exponents which are not marked have valuation greater than $\frac{m-1}{2}$. They are good candidates to be AB-exponents. Experimentally the work factor of sieving is about $2^{1.2m}$. There where only a very few exponents with valuation greater or equal $(m+1)/2$ that are not Gold, Kasami, Niho, Welch : 69 in dimension 27, 80 in dimension 29, 93 in dimension 31 and only 141 for dimension 33. After approximately one week of computation, we get

- Dobbertin's conjecture is correct up to the dimension $m \leq 33$.
- Nearly all the invertible d of valuation greater or equal to $\frac{m+1}{2}$ are Kasami-Welch exponents, see Section 5.

The last point is very interesting. Indeed, from dimension $m = 19$ up to dimension $m = 33$ all the exponents of valuation $(m+1)/2$ are Niho, Welch, Gold or Kasami-Welch with only three exceptions. Based on these numerical results we state the following Conjecture.

Conjecture 3.1. *Outside the Gold, Niho, Welch and Kasami-Welch classes for $m > 17$, there are exactly three exponents of valuation $\frac{m+1}{2}$.*

(1) $2^{\frac{m-1}{2}} + 2^{\frac{m-3}{2}} + 1$,
(2) $\frac{13}{3}$
(3) $\frac{2^{\frac{m-1}{2}}+2^{\frac{m+1}{4}}+1}{3}$ *or* $\frac{2^{\frac{m+1}{2}}+2^{\frac{m-1}{4}}+1}{3}$ *according to the congruence of m mod 4.*

Moreover, all these exponents have the 5-valued spectrum $\{0, \pm 2^{(m+1)/2}, \pm 2^{(m+3)/2}\}$.

The spectrum of the exponents (1) and (3) was already conjectured to be at most 5-valued by Niho in Conjecture 4-6 of [17]. In passing, we inform the readers that a part of his conjecture is false since the spectrum of the exponent $2^{(m+3)/4} + 3$ is not 5-valued in dimension 21. In the next Section, we will prove that the valuation of the second exponent is indeed $\frac{m+1}{2}$.

4. Valuation of 13/3

We use the *modular add-carry* approach of [10] to determine the valuation of $\frac{13}{3}$.

Let j be a residue modulo $q-1$. If $(j_{m-1}\dots j_1 j_0)$ and $(s_{m-1}\dots s_1 s_0)$ denote the binary expansions of j and $jd \mod (q-1)$ then there exists, see [10], one and only one sequence of *carries* $(c_{m-1}\dots c_1 c_0)$ such that

$$\forall i, \quad 2c_i + s_i = \sum_{k\in \mathrm{supp}\,(d)} j_{i-k} + c_{i-1}, \quad 0 \le c_i < \mathrm{wt}\,(d). \tag{4}$$

with $\mathrm{supp}\,(d) = \{i| \mid d_i = 1\}$ in the binary expansion of $d = d_0 + d_1 2^1 + \dots + d_{m-1}2^{m-1}$, and where the operations over the indices are done modulo m. From the relation (3), the valuation of $\frac{13}{3}$ is given by

$$\nu_d = \min_{1\le j\le q-2} \mathrm{wt}\,(-j) + \mathrm{wt}\,(j\frac{13}{3}) = \min_{1\le j\le q-2} \mathrm{wt}\,(-3j) + \mathrm{wt}\,(13j)$$

Let s and s' be the binary expansions of $13j$ and $3j$. Using (4), there exists a pair of carry sequences c and c' such that for all i, $0 \le i < m$:

$$\begin{aligned} 2c_i + s_i &= j_{i-3} + j_{i-2} + j_i + c_{i-1}, && 0 \le c_i < 3. \\ 2c'_i + s'_i &= j_{i-1} + j_i + c'_{i-1}, && 0 \le c'_i < 2. \end{aligned}$$

In particular,

$$\sum_{i=0}^{m-1} c_i - \sum_{i=0}^{m-1} c'_i + \mathrm{wt}\,(13j) - \mathrm{wt}\,(3j) = \mathrm{wt}\,(j)$$

since for all residue $r \neq 0$, $\mathrm{wt}\,(r) + \mathrm{wt}\,(-r) = m$, this equality becomes

$$\mathrm{wt}\,(13j) + \mathrm{wt}\,(-3j) = m + \mathrm{wt}\,(j) + \sum_{i=0}^{m-1} (c'_i - c_i). \tag{5}$$

Let us consider the graph of order $2^4 \times 2 \times 3$ whose vertex set is $\{0,1\}^4 \times \{0,1\} \times \{0,1,2\}$ drawing an edge

$$(\delta, \gamma, \beta, \alpha, x, x') \to (\epsilon, \delta, \gamma, \beta, y, y')$$

if and only if $y = [(\alpha + \beta + \delta + x)/2]$ and $y' = [(\alpha + \gamma + y)/2]$, where $[t]$ denotes the integer part of t. Note that for all index i, the vertices $(j_i, j_{i-1}, j_{i-2}, j_{i-3}, c'_{i-1}, c_{i-1})$ and $(j_{i+1}, j_i, j_{i-1}, j_{i-2}, c'_i, c_i)$ are connected in the graph. More precisely, there is a one to one correspondence between cycles of length m and the solution of of (4). After simplification, deleting the

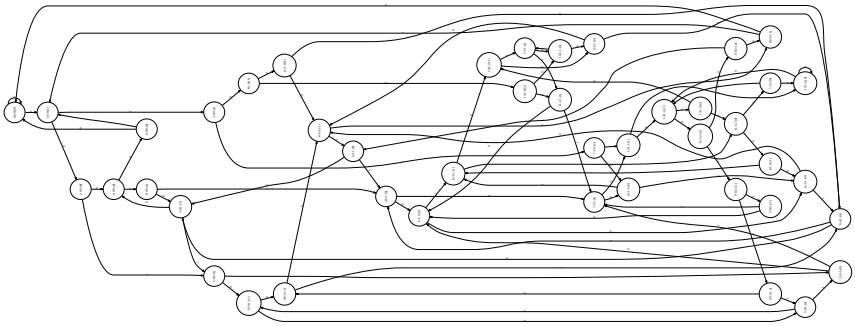


Fig. 1. Graph of 13/3.

vertices that are not on a cycle, we obtain the graph of order 44 represented by Figure (1).

The elements j such that $\mathrm{wt}\,(-j) + \mathrm{wt}\,(jd)$ is minimal correspond to the cycles of length m minimizing the cost function K defined on vertices by

$$K(\delta, \gamma, \beta, \alpha, x, x') = \delta + x' - x,$$

and extended to a cycle summing the cost of the vertices of this cycle.

Using a computer program, it is easy to enumerate all the elementary cycles computing their costs. It appears that the cost of an elementary cycle of length $2L$ or $2L+1$ is greater than $-L$. Therefor it follows that the valuation of $\frac{13}{3}$ is greater or equal to $\lfloor \frac{m+1}{2} \rfloor$.

Using the fact that $2^{m-1} = \frac{1}{2}$, a straightforward calculation gives

$$2 \times \frac{13}{3} = 9 + \sum_{i=0}^{\frac{m-3}{2}} 2^{2i+1}$$

and this shows that $\mathrm{wt}\,(\frac{13}{3}) = \frac{m+1}{2}$. It follows that $\mathrm{wt}\,(1) + \mathrm{wt}\,(-\frac{13}{3}) = \frac{m+1}{2}$. In particular the valuation $\frac{13}{3}$ is smaller or equal to $\frac{m+1}{2}$.

Summarizing we have proven

Theorem 4.1. *For m odd, the valuation of $d = \frac{13}{3}$ is equal to $\frac{m+1}{2}$.*

5. Kasami-Welch exponent

Using quadratic form theory, one can easily prove that the Fourier coefficients of the exponent

$$d = \frac{2^{tk}+1}{2^k+1} \quad \text{(Kasami-Welch)}$$

takes values in

$$\{0, \quad \pm 2^{\frac{m+e}{2}}, \quad \pm 2^{\frac{m+3e}{2}}, \quad \pm 2^{\frac{m+5e}{2}}, \quad \ldots\}$$

where $e = (m, k)$.

The case $t = 3$ corresponds to the Kasami exponent and thus the spectrum is actually 3-valued. In the case $t = 5$, Niho proved the spectrum is at most 5-valued. One can use the work of Kasami to prove that the spectrum is actually 5-valued. A simpler proof was given recently by Bracken [1] adapting the method of Dobbertin [9]. In his thesis, Niho (page 72) proposes the following conjectures on Kasami-Welch exponents :

conjecture	cond.	m		spectrum
conj. 4-2	$e > 1$		3-valued	$0, \pm 2^{\frac{m+e}{2}}$
conj. 4-3	$e = 1$	not prime	5-valued	
conj. 4-4	$e = 1$	prime	5-valued	$0, \pm 2^{\frac{m+1}{2}}, \pm 2^{\frac{m+3}{2}}$

Unfortunately, we constructed counter-examples for conjecture 4-4 in [14], and finally all these conjectures false [15].

Based on our experimental results we state the following conjecture.

Conjecture 5.1. *The Kasami-Welch exponent* $\frac{2^{kt}+1}{2^k+1}$ *is almost bent if and only if* $t = 3$ *and* $(k, m) = 1$.

As explained by Voloch during the SAGA conference, this weaker version of Dobbertin's conjecture is partially proved by works of Jedlicka [11] in the sense that for all fixed $t > 3$ and all fixed k the exponent $(2^{kt}+1)/(2^k+1)$ is not almost perfect nonlinear for all m large enough.

References

1. C. Bracken *Designs, codes, spin models and the Walsh transform* Ph.D. Thesis, National University of Ireland, Maynooth (2004).
2. A. Canteaut, P. Charpin P, H. Dobbertin, Binary M-sequences with three-valued crosscorrelation: a proof of theWelch conjecture. *IEEE trans IT.*, 46(1):4–8, 2000.
3. Canteaut A., Charpin P. and Dobbertin H. Weight divisibility of cyclic codes, highly nonlinear functions and crosscorrelation of maximum-length sequences. *SIAM J. Discrete Math.*, 13:105–138, 2000.

4. F. Chabaud and S. Vaudenay. Links between differential and linear cryptanalysis, *Advances in Cryptology -EUROCRYPT'94, Lecture Notes in Computer Science*, Springer-Verlag, New York, 950, pp. 356-365, 1995.
5. Cusick T., Dobbertin H. Some new 3-valued crosscorrelation functions of binary m-sequences. *IEEE Transactions on Information Theory*, 42:1238–1240, 1996.
6. H. Dobbertin One-to-one highly nonlinear power functions on finite fields. *AAECC*, 9(2):139–152, 1998.
7. H. Dobbertin Almost Perfect nonlinear power functions on $GF(2^n)$: the Welch case. *IEEE Trans. on Info. Theory*, 45:1271–1275, 1999.
8. H. Dobbertin Almost Perfect nonlinear power functions on $GF(2^n)$: the Niho case. *Inform. and Comput.*, 151:57–72, 1999.
9. H. Dobbertin Another Proof of Kasami's Theorem. *Des. Codes Crypt.*, 13:177–180, 1999.
10. H.D.L. Hollmann and Q. Xiang, *A Proof of the Welch and Niho Conjectures on Crosscorrelations of Binary m-sequences*, Finite Fields Appl. **7** (2001), 253–286.
11. D. Jedlicka, Classifying APN Monomials preprint 2005, http://eprint.iacr.org/2005/096.pdf
12. Kasami T. The weight enumerators for several classes of subcodes of the 2-nd Reed-Muller codes. *Information and Control*, 18:369–394, 1971.
13. Niho Y. *Multi-valued cross correlation functions between two maximal linear recursive sequences*. PhD thesis, University of Southern California, 1972.
14. Ph. Langevin, G. Leander, G. Mcguire, Gary A Counter-Example to a Conjecture of Niho preprint
15. Ph. Langevin Numerical Experiments on Power Functions `http://langevin.univ-tln.fr/project/spectrum/`
16. P. Langevin, P. Véron, *On the Non-linearity of Power Functions* Des. Codes Cryptography **37** 31-43 (2005)
17. Y. Niho, *Multi-valued cross-correlation functions between two maximal linear recursive sequences* Ph.D. Thesis, University of Southern California (1972).
18. Sidel'Nikov V. M. On the mutual correlation of sequences. *Soviet Math. Dokl.*, 12:197–201, 1971.
19. J. Stickelberger. Uber eine verallgemeinerung der kreistheilung. *Math. Ann.*, 37:321–367, 1890.

On the number of resilient Boolean functions

Sihem Mesnager
MAATICAH, University Paris 8
France
E-mail: hachai@math.jussieu.fr

Boolean functions are very important primitives of symmetric cryptosystems. To increase the security of such cryptopsystems, these Boolean functions have to fit several security criteria. In particular, they have to be m-resilient, that is, to be balanced and m-correlation immune. This class of Boolean function has been widely studied by cryptographers. Nevertheless, the problem of counting the number of m-resilient n-variables Boolean functions is still challenging. In this paper, we propose a new approach to this question. We reword this question in that to count integer solutions of a system of linear inequalities. This allows us to deduce two representation formulas for the number of m-resilient n-variables Boolean functions.

Keywords: Boolean functions, resiliency, multivariate generating function, residue calculus

Introduction

Symmetric cryptosystems are commonly used for encrypting and decrypting messages owing to their efficiency. The encryption and the decryption consists in adding bitwisely the input stream and a pseudo random sequence generated by a pseudo-random generator from a secret information, the secret key. Classical tools to produce such pseudo-random sequences, that are called the keystream, are Linear Feedback Registers (LFSR). Because of the linear properties of the LFSR's, Boolean functions are used to combine several LSFR or filter one LFSR. Thus, the security of such cryptosystems deeply relies on the choice of the Boolean function. Indeed, some properties of the Boolean function can be exploited by cryptanalysts, that is, people who wishes to deduce the plaintext (the decrypted message) from a ciphertext (an encrypted message), without knowledge of the secret key. These Boolean functions need to have some important characteristics to resists to several types of attacks that are called security criteria.

Among all the security criteria, two of them are crucial : the Boolean function has be to *balanced*, that is takes the value 1 with probability 1, and m-correlation immune, that is, the output distribution does not change if we fix at most m inputs. The Boolean function that are both balanced and m-correlation immune are said to be m-resilient. The literature about correlation immune or resilient Boolean functions is very rich and a lot of problems on this subject remains open. Notably, the problem of counting the number of m-resilient n-variables Boolean functions is still challenging. Indeed, this number is only known for $m = 1$ up to 7 variables (the number of 1-resilient 7-variables Boolean function have been found in 2007 [10]) and for $m \geq n - 3$ for every n [2]. This problem seems to be untractable.

In fact, the problem of efficiently lower and upper bounding the number of m-resilient n-variables Boolean functions remains also open for every positive integer m less than $n - 3$. Schneider [14] obtained an upper bound which seems efficient for resilient functions of low order. Some another results have been obtained in [8,13,17]. Their results are slightly better than [14] but are more complex to compute. Schneider's upper bound has been improved for high order in [4,7]. None of the upper bounds presented in [4,7,14] improves all the other upper bounds in all situations. Few efficient lower bounds were found. Mostly, they are obtained by building and counting restricted classes of resilient Boolean functions [11–13,16]. Recently, further improvement have been done by Le Bars and Viola [10] which presented the best lower bound and upper bound on the number of 1-resilient n-variables Boolean function.

In this paper, we present a new approach for the problem of counting the number m-resilient n-variables Boolean functions. For that, we use the numerical normal form of Boolean functions of [5,6], that is, the representation as multivariate polynomials over $\mathbb{Z}$. Carlet and Guillot [6] provided a characterization of m-resilient n-variables Boolean functions by means of the numerical normal form. This allows us to reword the problem of counting the number of m-resilient n-variables Boolean functions in that to count the number of integer solutions of a system of linear inequalities (Proposition 3.2). We then use a classical approach of enumerative combinatorics to count integer solutions of linear systems with integer coefficients, that is, we introduce a multivariate generating function and express the number of m-resilient n-variables Boolean functions with respect to the coefficients of this multivariate generating function. We then use multivariate residue calculus to derive a representation formula by means of the Cauchy integral (Proposition 3.4). In a second stage, we propose an alternative represen-

tation formula for the number of m-resilient n-variables Boolean function. More precisely, we show that this number can also be interpreted as a coefficient of a term of a multivariate polynomial with integer coefficients (Proposition 3.6).

The paper is organized as follows. In section 1, we briefly recall the main definitions and results that we need about Boolean functions. We next briefly mention some previous known results about then question of counting the number of resilient Boolean functions (section 2). We then present in section 3 our main results that are the two representation formulas for the number of m-resilient n-variables Boolean functions. For the sake of clearness, we give separately in two subsections these representation formulas together with their proofs.

Notation 0.1. In this paper, $\{1, \ldots, n\}$ stands for the set of all integers ranging from 1 to n and $\mathcal{P}_n$ stands for the set of subsets of $\{1, \ldots, n\}$. The cardinality of a subset I of $\{1, \ldots, n\}$ shall be denoted by $|I|$. We denote by Θ_n^m the subset of $\mathcal{P}_n$ of all subsets whose cardinality is at most $n-m-1$. We let Γ_n^m be the subset of $\mathcal{P}_n$ formed with all subsets of $\{1, \ldots, n\}$ of cardinality at least $n-m$.

1. Background on Boolean functions

Let $\mathbb{F}_2$ be the finite field $\{0, 1\}$. An n-variable Boolean function f is a map from $\mathbb{F}_2^n$, that is, the set of all binary vectors of length 2^n, to $\mathbb{F}_2$. The set of all n-variable Boolean functions shall be denoted by $\mathcal{B}_n$. The hamming weight of an n-variable Boolean function f, that we denote by $wt(f)$, is the number of 1 in its truth table, that is, $wt(f)$ is the cardinality of $\{x \in \mathbb{F}_2^n \mid f(x) = 1\}$. An n-variable f is said to be *balanced* if its Hamming weight equals 2^{n-1}.

Definition 1.1. Let m be a positive integer less than n. An n-variable Boolean function f is said to be *m-correlation immune* if the output distribution does not change when t input values (*i.e.* t coordinates of the input binary vector) are fixed. If f is moreover balanced then f is said to be m-resilient. The set of all n-variable m-resilient Boolean functions shall be denoted by Res_n^m.

An n-variable Boolean function f admits an unique representation as a multivariate polynomial over $\mathbb{F}_2$, that is called its *algebraic normal form* :

$$\forall x \in \mathbb{F}_2^n, \quad f(x) = \bigoplus_{I \in \mathcal{P}_n} a_I \prod_{i \in I} x_i \tag{1}$$

where the a_I's are in $\mathbb{F}_2^n$. The terms $\prod_{i\in I} x_i$ are called monomials. The algebraic degree $\deg(f)$ of an n-variable Boolean function equals the maximum degree of those monomials whose coefficient are nonzero in its algebraic normal form. It has been shown that the higher the order of resiliency is, the lower the maximum degree is. More precisely, it has been shown :

Proposition 1.1 (Siegenthaler [15]). *Suppose that $1 \leq m < n-1$. Then, the algebraic degree of an n-variable m-resilient Boolean function is at most $n-m-1$.*

Any n-variable Boolean function can be viewed as an integer-valued mapping taking values in the subset $\{0,1\}$ of $\mathbb{Z}$. Now, any integer-valued mapping f can be uniquely represented as a multivariate polynomial over $\mathbb{Z}$:

$$\forall x \in \mathbb{F}_2^n, \quad f(x) = \sum_{I\in\mathcal{P}_n} \lambda_I \prod_{i\in I} x_i \tag{2}$$

where the λ_I's are in $\mathbb{Z}$. The degree of the numerical normal form of an integer-valued map f is called its *numerical degree*. To ensure that f takes values in $\{0,1\}$, that is, that satisfies $f^2(x) = f(x)$ for every $x \in \mathbb{F}_2^n$, the coefficients λ_I's has to satisfy

$$\forall I \in \mathcal{P}_n, \quad \Big(\sum_{J\subset I} \lambda_J\Big)^2 - \sum_{J\subset I} \lambda_J = 0. \tag{3}$$

where $\sum_{J\subset I}$ denotes the summation over all the subsets J of $\{1,\ldots,n\}$ which are contained in the subset I. Carlet and Guillot [6] characterized resilient Boolean function by means of the numerical normal form. We give below this characterization.

Theorem 1.1. *Let f be an n-variable Boolean function f. Let g be the n-variable Boolean function defined as : $\forall x \in \mathbb{F}_2^n$, $g(x) = f(x) \oplus \bigoplus_{i=1}^n x_i$. Then, f is m-resilient if and only if the numerical degree of g is less than or equal to $n-m-1$.*

2. State of art

In this section, we present a short and non exhaustive survey of the question of counting or finding lower bound or upper bound for the number of m-resilient n-variable Boolean functions. We omit to speak about the lower bounds. In fact, they are mostly obtained by building classes of m-resilient n-variables Boolean function. For further details, we send the reader to the

recent work of Claude Carlet [3] which summarizes the previous known results related with the enumeration of resilient Boolean function.

The only n-variable Boolean functions which are $(n-1)$-resilient are the two affine n-variable Boolean functions : $\bigoplus_{i=1}^{n} x_i$ and its complement $1 \oplus \bigoplus_{i=1}^{n} x_i$. Thus, $\left|Res_n^{n-1}\right| = 2$. Siegenthaler's upper bound on the algebraic degree of a resilient Boolean function (Proposition 1.1) implies that only affine n-variable affine Boolean functions can be $(n-2)$-resilient. Now, a sub-function obtained by fixing at most $(n-2)$ in an affine Boolean function stays balanced if and only if this sub-function is not constant. That requires that the algebraic normal form of this function contains at least $n-1$ monomials x_i. The number of such affine n-variable Boolean functions equals $2\binom{n}{n-1}+2$, that is, we have $\left|Res_n^{n-2}\right| = 2(n+1)$. The first non trivial result about the number of m-resilient Boolean functions has been obtained by Camion and *al* ([2]) :

$$\left|Res_n^{n-3}\right| = \frac{n(n-1)(3n-2)(n+1)}{3}$$

Except those cases, the only other known values are the ones of $|Res_5^1|$, $|Res_6^1|$ (Harary and Palmer, [9]) and $|Res_7^1|$ (Le Bars and Viola, [10]) :

$$|Res_5^1| = 807980 \quad |Res_6^1| = 95259103924394$$
$$|Res_7^1| = 23478015754788854439497622689296$$

For the other values of $|Res_n^m|$, upper bounds have been shown. Before stating those uppers bounds, let us make a simple remark : according to Proposition 1.1, the algebraic degree of an n-variable m-resilient Boolean function cannot exceed $n-m-1$; that implies a simple upper bound of the number of n-variable m-resilient Boolean functions :

$$|Res_n^m| \leq 2^{1+n+\binom{n}{2}+\cdots+\binom{n}{n-m-1}} \tag{4}$$

that we shall refer in the sequel as the *naive bound* on the number of n-variable m-resilient Boolean function. The first general and efficiently computed upper bound was found by Schneider ([14]).

Proposition 2.1. *For every positive integers n and m such that $1 \leq m \leq n-1$, we have :*

$$|Res_n^m| \leq \prod_{j=1}^{n-m} \binom{2^j}{2^{j-1}}^{\binom{n-j-1}{m-1}}$$

This bound is weak for high orders of resiliency. For instance, the exact number of $(n-3)$-resilient which equals $n(n-1)(3n-2)(n+1)/3$ is much less

than $\prod_{j=1}^{3}\binom{2^j}{2^{j-1}}^{\binom{n-j-1}{n-4}}$. This upper bound has been partially improved for high orders, firstly, by Carlet and Klapper and, next, by Carlet and Gouget.

Proposition 2.2 (Carlet and Klapper, [7]). *For every positive integers n and m such that $1 \leq m \leq n-1$, we have :*

$$|Res_n^m| \leq \frac{2^{\sum_{i=0}^{n-m-1}\binom{n}{i}} - 2^{\sum_{i=0}^{n-m-2}\binom{n}{i}}}{2^{2^{2m+1}-1}} + 2^{\sum_{i=0}^{n-m-2}\binom{n}{i}}$$

for $2 \leq m < \frac{n}{2}$ and

$$|Res_n^m| \leq \frac{2^{\sum_{j=0}^{n-m-1}\binom{n}{j}}(1+\epsilon)}{2^{\sum_{j=0}^{n-m-1}\binom{m-1}{i}}} + 2^{\sum_{j=0}^{n-m-2}\binom{n}{j}} \text{ where } \epsilon = \frac{1}{2^{\Omega((2^n/n)^{1/2})}}.$$

for $\frac{n}{2} \leq m < n-2$.

Proposition 2.3 (Carlet and Gouget, [4]). *For every positive integers n and m such that $1 \leq m \leq n-1$, we have :*

$$|Res_m^n| \leq 2^{\sum_{i=0}^{n-m-2}\binom{n}{i}} + \frac{\binom{n}{n-m-1}}{2^{\binom{m+1}{n-m-1}+1}} \prod_{i=1}^{n-m}\binom{2^j}{2^{j-1}}^{\binom{n-j-1}{m-1}}$$

3. Representation formulas for the number of resilient Boolean function

Throughout this section, we shall use the following notation in order to allow compact description of our result. Let $\mathcal{X}$ be a set of numbers. We shall denote by $\mathcal{X}^{\mathcal{I}}$, where $\mathcal{I}$ is a finite set, the set $\{\boldsymbol{x} = (x_I)_{I\in\mathcal{I}} \mid \forall I \in \mathcal{I}, x_I \in \mathcal{X}\}$. Throughout this section, n denotes any positive integer greater than 4 and m is a positive integer such that $m < n-3$. We shall denote by $\oint$ the Cauchy integral, that is, in the sequel, the expression $\oint F(z)dz$, where F is a k-variables complex-valued mapping and where we adopt the multivariate notation $dz = \prod_{i=1}^{k} z_i$, has to be understood as a multiple integral; each integral is over a circle of radius < 1 centered at 1; all appearing radii should be different. We shall also use the following result that expresses terms of a multi-indexed integer sequences by means of a residue formula

Proposition 3.1. *Let $K = (K_{n_1,\ldots,n_k})_{(n_1,\ldots,n_k)\in\mathbb{N}^k}$ be a multi-indexed integer sequence. Let G_K be the associated multivariate generating function, that is, the multivariate mapping defined as*

$$G_K(z) = \sum_{(n_1,\ldots,n_k)\in\mathbb{N}^k} K_{n_1,\ldots,n_k} \prod_{i=1}^{k} z_i^{n_i}.$$

for $z = (z_1, \ldots, z_k) \in \mathbb{C}^k$. *Then, for every* $n = (n_1, \ldots, n_k) \in \mathbb{N}^k$, *we have*

$$K_{n_1,\ldots,n_k} = \frac{1}{(2i\pi)^k} \oint G_K(z) z^{-n} \frac{dz}{z}$$

where we use the multivariate notation : $z^{-n} = \prod_{i=1}^k z_i^{-n_i}$ *and* $dz = \prod_{i=1}^k dz_i$.

3.1. *A first representation formula*

The problem of counting the number of m-resilient n-variable Boolean functions can be reworded into the problem of counting the integer solutions of a system of linear inequalities. For that, we use Theorem 1.1 that characterizes the n-variable Boolean function which are m-resilient by using the numerical normal form (2).

Proposition 3.2. *Let* $\mathfrak{R}_n^m$ *be the subset of* $\mathbb{R}^{\Theta_n^m}$ *defined as*

$$\mathfrak{R}_n^m = \Big\{ (x_J)_{J \in \Theta_n^m} \in \mathbb{R}^{\Theta_n^m} \Big| \forall I \in \mathcal{P}_n, 0 \leq \sum_{\substack{J \in \Theta_n^m \\ J \subset I}} x_J \leq 1 \Big\}. \tag{5}$$

Then,

$$\big| Res_n^m \big| = \big| \mathbb{Z}^{\Theta_n^m} \cap \mathfrak{R}_n^m \big|.$$

Proof. Take $f \in \mathcal{B}_n$. Define $g \in \mathcal{B}_n$ as : $\forall x \in \mathbb{F}_2^n$, $g(x) = f(x) \oplus \bigoplus_{i=1}^n x_i$. By Theorem 1.1, f is m-resilient *if and only if* the numerical degree of g is at most $n - m - 1$. Now, the map from $\mathcal{B}_n$ to itself which maps f to g is one-to-one. That implies in particular that the number of m-resilient n-variables Boolean function equals the number of integer-valued mappings taking values in $\{0, 1\}$ whose numerical normal forms is of numerical degree at most $n-m-1$. Now, because of the uniqueness of the numerical normal form and according to (3), $|Res_n^m|$ is equal to the number of $(\lambda_J)_{J \in \Theta_n^m} \in \mathbb{Z}^{\Theta_n^m}$ which satisfies : $\forall I \in \mathcal{P}_n$, $\sum_{\substack{J \in \Theta_n^m \\ J \subset I}} \lambda_J \in \{0, 1\}$. □

We now state without proof and with our notation a classical result that is called the Mobius-Rota inversion formula.

Lemma 3.1. *We have :*

$$\Big(\forall I \in \Theta_n^m, z_I = \sum_{\substack{J \in \Theta_n^m \\ J \subset I}} x_J \Big) \iff \Big(\forall I \in \Theta_n^m, x_I = \sum_{\substack{J \in \Theta_n^m \\ J \subset I}} (-1)^{|I|-|J|} z_J \Big). \tag{6}$$

We then derive from Lemma 3.1 that the elements of $\mathfrak{R}_n^m$ belongs to a bounded domain of $\mathbb{R}^{\Theta_n^m}$.

Lemma 3.2. *Let $\mathfrak{R}_n^m$ be the subset of $\mathbb{R}^{\Theta_n^m}$ defined in Proposition 3.2. Let $(x_J)_{J\in\Theta_n^m} \in \mathfrak{R}_n^m$. Then $x_\emptyset \in [0,1]$ and,*

$$\forall J \in \Theta_n^m, \quad -2^{|J|-1} \leq x_J \leq 2^{|J|-1}.$$

Proof. Firstly, note that, if $I = \emptyset$, the summation $\sum_{\substack{J\in\Theta_n^m \\ J\subset I}} x_J$ reduces to $x_\emptyset$ and thus the condition $0 \leq \sum_{\substack{J\in\Theta_n^m \\ J\subset\emptyset}} x_J \leq 1$ simply says that $x_\emptyset \in [0,1]$.

For the other cases, we use Lemma 3.1 in the particular case where $(x_J)_{J\in\Theta_n^m}$ belongs to $\mathfrak{R}_n^m$, that is, in the case where $z_I = \sum_{\substack{J\in\Theta_n^m \\ J\subset I}} x_J \in \{0,1\}$ for every $I \in \mathcal{P}_n$. That gives, for every $I \in \Theta_n^m$,

$$x_I = \sum_{\substack{J\in\Theta_n^m \\ J\subset I}} (-1)^{|I|-|J|} z_J \leq \big|\{J \in \Theta_n^m \mid J \subset I,\ |I|-|J| \text{ is even}\}\big| = 2^{|J|-1}$$

and

$$x_I = \sum_{\substack{J\in\Theta_n^m \\ J\subset I}} (-1)^{|I|-|J|} z_J \geq -\big|\{J \in \Theta_n^m \mid J \subset I,\ |I|-|J| \text{ is odd}\}\big| = -2^{|J|-1}.$$

□

Remark 3.1. On can recover the *naive upper bound* (4) by using Mobius-Rota inversion formula (6). Indeed, introduce the linear mapping φ from $\mathbb{R}^{\Theta_n^m}$ to itself which maps $(x_J)_{J\in\Theta_n^m}$ to $\left(\sum_{\substack{J\in\Theta_n^m \\ J\subset I}} x_J\right)_{J\in\Theta_n^m}$. Lemma 3.1 implies that φ is one-to-one. Now, by definition, the subset $\mathfrak{R}_n^m$ is contained in the preimage of $\{0,1\}^{\Theta_n^m}$ under φ. Now, the linear mapping φ being is one-to-one, we have : $|\mathfrak{R}_n^m| \leq \big|\{0,1\}^{\Theta_n^m}\big| = 2^{|\Theta_n^m|} = 2^{\sum_{j=0}^{n-m-1}\binom{n}{i}}$.

We now use Lemma 3.2 to slightly reword the statement of Proposition 3.2. The idea is to translate $\mathfrak{R}_n^m$ by an integer vector so that its image under this translation lies in the non-negative orthant $\mathbb{R}_+^{\Theta_n^m}$. At this stage, an important point is to note that translating by an integer vector does change the integer solution count.

Corollary 3.1. *Let $\mathfrak{S}_n^m$ be the subset of $\mathbb{R}_+^{\Theta_n^m}$ defined as*

$$\mathfrak{S}_n^m = \left\{(y_J)_{J\in\Theta_n^m} \in \mathbb{R}_+^{\Theta_n^m} \;\middle|\; \forall I \in \mathcal{P}_n,\, b_I \leq \sum_{\substack{J\in\Theta_n^m \\ J\subset I}} y_J \leq b_{I+1}\right\} \tag{7}$$

where

$$b_I = \sum_{j=1}^{\min(|I|,n-m-1)} \binom{|I|}{j} 2^{j-1} \text{ if } I \in \mathcal{P}_n \setminus \{\emptyset\} \text{ and } b_\emptyset = 0.$$

Then,

$$\left|Res_n^m\right| = \left|\mathbb{N}^{\Theta_n^m} \cap \mathfrak{S}_n^m\right|.$$

Proof. Define $(v_J)_{J\in\mathcal{P}_n}$ as : $v_\emptyset = 0$ and, for every $J \in \mathcal{P}_n\setminus\{\emptyset\}$, $v_J = 2^{|J|-1}$. Set $\mathfrak{S}_n^m = \{(y_J) \in \mathbb{R}^{\Theta_n^m} \mid \exists (x_J)_{J\in\Theta_n^m} \in \mathfrak{R}_n^m, \forall J \in \mathcal{P}_n, y_J = x_J + v_J\}$. Lemma 3.2 says that $\mathfrak{S}_n^m \subset \mathbb{R}_+^{\Theta_n^m}$ since : $\forall J \in \Theta_n^m$, $x_J \geq -2^{|J|-1} = -v_J$. Now, replacing x_J by $y_J - v_J$ in all the linear inequalities defining $\mathfrak{R}_n^m$ yields to the new system of linear inequalities whose unknowns are the y_J's

$$\forall I \in \mathcal{P}_n, \quad \sum_{\substack{J\in\Theta_n^m \\ J\subset I}} v_J \leq \sum_{\substack{J\in\Theta_n^m \\ J\subset I}} y_J \leq \sum_{\substack{J\in\Theta_n^m \\ J\subset I}} v_J + 1.$$

The result follows then from the identity $\sum_{\substack{J\in\Theta_n^m \\ J\subset I}} v_J = b_{|I|}$. □

Let $\mathfrak{S}_n^m$ be the subset of $\mathbb{R}_+^{\Theta_n^m}$ defined by Corollary 3.1. We then split $\mathfrak{S}_n^m$ into disjoint subsets. For that, we introduce a collection of subsets of $\mathbb{R}_+^{\Theta_n^m}$ indexed by $\mathbb{N}^{\mathcal{P}_n}$ whose terms are defined as

$$\forall c = (c_I)_{I\in\mathcal{P}_n} \in \mathbb{N}^{\mathcal{P}_n}, \quad \mathfrak{T}_n^{c,m} = \{(y_J)_{J\in\Theta_n^m} \in \mathbb{R}_+^{\Theta_n^m} \mid \forall I \in \mathcal{P}_n, \sum_{\substack{J\in\Theta_n^m \\ J\subset I}} y_J = c_I\}.$$

We then have

$$\mathfrak{S}_n^m = \bigcup_{\epsilon\in\{0,1\}^{\mathcal{P}_n}} \mathfrak{T}_n^{b+\epsilon,m}. \tag{8}$$

where the terms of $b = (b_I)_{I\in\mathcal{P}_n}$ are the b_I's defined in Corollary 3.1. Next, let $K = (K_c)_{c\in\mathbb{N}^{\mathcal{P}_n}}$ be the integer sequence index by $\mathcal{P}_n$ whose terms are : $\forall c \in \mathbb{N}^{\mathcal{P}_n}$, $K_c = \left|\mathbb{N}^{\Theta_n^m} \cap \mathfrak{T}_n^{c,m}\right|$. But above, we have

$$|Res_n^m| = \left|\mathbb{N}^{\Theta_n^m} \cap \mathfrak{S}_n^m\right| = \sum_{\epsilon\in\{0,1\}^{\mathcal{P}_n}} \left|\mathfrak{T}_n^{b+\epsilon,m} \cap \mathbb{N}^{\Theta_n^m}\right| = \sum_{\epsilon\in\{0,1\}^{\mathcal{P}_n}} K_{b+\epsilon}. \tag{9}$$

A classical approach in enumerative combinatorics is to introduce the multivariate generating function associated to K that we denote by G_K and that is defined as :

$$z \in \mathbb{C}^{\mathcal{P}_n}, \quad G_K(z) = \sum_{c\in\mathbb{N}^{\Theta_n^m}} K_c z^c \tag{10}$$

where we use the multivariate notation $z^c = \prod_{I\in\mathcal{P}_n} z_I^{c_I}$ for $z = (z_I)_{I\in\mathcal{P}_n}$ and $c = (c_I)_{I\in\mathcal{P}_n}$. We then prove the following key result.

Proposition 3.3. *The power series* (10) *converges provided that, for every* $I \in \mathcal{P}_n$*, the modulus* $|z_I|$ *is small enough. Moreover, when the power series converges, we have*

$$G_K(z) = \prod_{J\in\Theta_n^m} \frac{1}{1-\prod_{\substack{I\in\mathcal{P}_n\\ J\subset I}} z_I}.$$

Proof. Firstly

$$\sum_{\mathbf{c}\in\mathbb{N}^{\mathcal{P}_n}} K_c z^c = \sum_{\mathbf{c}\in\mathbb{N}^{\mathcal{P}_n}} \sum_{y\in\mathbb{N}^{\Theta_n^m}\cap\mathfrak{T}_n^{c,m}} \prod_{I\in\mathcal{P}_n} \prod_{\substack{J\in\Theta_n^n\\ J\subset I}} z_I^{y_J}$$

because

$$z^{\mathbf{c}} = \prod_{I\in\mathcal{P}_n} z_I^{c_I} = \prod_{I\in\mathcal{P}_n} z_I^{\sum_{\substack{J\in\Theta_n^m\\ J\subset I}} y_J} = \prod_{I\in\mathcal{P}_n} \prod_{\substack{J\in\Theta_n^m\\ J\subset I}} z_I^{y_J}$$

whenever $y = (y_I)_{I\in\mathcal{P}_n} \in \mathfrak{T}_n^{c,m}$. Thus

$$\begin{aligned}\sum_{c\in\mathbb{N}^{\mathcal{P}_n}} K_c z^c &= \sum_{y\in\mathbb{N}^{\Theta_n^m}} \prod_{J\in\Theta_n^m} \prod_{\substack{I\in\mathcal{P}^n\\ J\subset I}} z_I^{y_J} = \prod_{J\in\Theta_n^m} \left(\sum_{y_J=0}^{+\infty} \prod_{\substack{I\in\mathcal{P}^n\\ J\subset I}} z_I^{y_J} \right)\\ &= \prod_{J\in\Theta_n^m} \frac{1}{1-\prod_{\substack{I\in\mathcal{P}^n\\ J\subset I}} z_I}\end{aligned}$$

provided that, for every $I \in \mathcal{P}_n$, $\left|\prod_{\substack{I\in\mathcal{P}^n\\ J\subset I}} z_I\right| = \prod_{\substack{I\in\mathcal{P}^n\\ J\subset I}} |z_I|$ is small enough. A sufficient condition is that, for every $I \in \mathcal{P}_n$, $|z_I|$ is small enough. □

We finally deduce from Proposition 3.1 a representation formula for $|Res_n^m|$

Lemma 3.3. *we have*

$$\forall c \in \mathbb{N}^{\mathcal{P}_n}, \quad K_c = \frac{1}{(2i\pi)^{2^n}} \oint G_K(z) z^{-c} \frac{dz}{z}.$$

where adopt the multivariate notation $dz = \prod_{I\in\mathcal{P}_n} dz_I$.

A straightforward consequence of the preceding Lemma is a representation formula for $|Res_m^n|$.

Proposition 3.4. *We have*

$$\left|Res_n^m\right| = \frac{1}{(2i\pi)^{2^n}} \oint G(z)\frac{dz}{z} \tag{11}$$

where G is defined for $z = (z_I)_{I\in\mathcal{P}_n}$ as

$$G(z) = \prod_{I\in\mathcal{P}_n} (1+z_I) z_I^{-b_I-1} \cdot \prod_{J\in\Theta_n^m} \frac{1}{1-\prod_{\substack{I\in\mathcal{P}^n\\ J\subset I}} z_I}.$$

Proof. Replacing each term $K_{b+\epsilon}$ by its expression given by Lemma 3.3 in (9) yields to

$$|\mathbb{N}^{\Theta_n^m} \cap \mathfrak{S}_n^m| = \sum_{\epsilon\in\{0,1\}^{\mathcal{P}_n}} \frac{1}{(2i\pi)^{2^n}} \oint F(z) z^{-(b+\epsilon)} \frac{dz}{z}.$$

Exchange the summation $\sum$ and the integration $\oint$:

$$|\mathbb{N}^{\Theta_n^m} \cap \mathfrak{S}_n^m| = \frac{1}{(2i\pi)^{2^n}} \oint F(z) z^{-b} \left(\sum_{\epsilon\in\{0,1\}^{\mathcal{P}_n}} z^{-\epsilon} \right) \frac{dz}{z}.$$

We then get the result by noting that

$$\sum_{\epsilon\in\{0,1\}^{\mathcal{P}_n}} z^{-\epsilon} = \prod_{I\in\mathcal{P}_n} z_I^{-1} \sum_{\epsilon\in\{0,1\}^{\mathcal{P}_n}} z^{\epsilon} = \prod_{I\in\mathcal{P}_n} z_I^{-1} \cdot \prod_{I\in\mathcal{P}_n} (1+z_I).$$

□

3.2. *Another representation formula for the number of resilient Boolean functions*

In this section, we state an alternative representation formula for $|Res_m^n|$. To this end, we begin with noting that the set $\mathfrak{R}_n^m$ of Proposition 3.2 can be written as

$$\mathfrak{R}_n^m = \mathfrak{R}_n^{1,m} \cap \mathfrak{R}_n^{2,m} \tag{12}$$

where

$$\mathfrak{R}_n^{1,m} = \left\{ (x_J)_{J\in\Theta_n^m} \in \mathbb{R}^{\Theta_n^m} \,\middle|\, \forall I \in \Theta_n^m,\, 0 \le \sum_{\substack{J\in\Theta_n^m\\ J\subset I}} x_J \le 1 \right\}$$

and

$$\mathfrak{R}_n^{2,m} = \left\{ (x_J)_{J\in\Theta_n^m} \in \mathbb{R}^{\Theta_n^m} \,\middle|\, \forall I \in \Gamma_n^m,\, 0 \le \sum_{\substack{J\in\Theta_n^m\\ J\subset I}} x_J \le 1 \right\}.$$

Thus,

$$\mathfrak{R}_n^m \cap \mathbb{Z}^{\Theta_n^m} = \left(\mathfrak{R}_n^{1,m} \cap \mathbb{Z}^{\Theta_n^m}\right) \cap \mathfrak{R}_n^{2,m} \tag{13}$$

Then, consider again the linear mapping φ introduced in Remark 3.1, that is, let φ be the linear mapping from $\mathbb{R}^{\Theta_n^m}$ to itself which maps $(x_J)_{J\in\Theta_n^m}$ to $\left(\sum_{\substack{J\in\Theta_n^m\\ J\subset I}} x_J\right)_{J\in\Theta_n^m}$. We then have $\mathfrak{R}_n^{1,m}\cap\mathbb{Z}^{\Theta_n^m}=\left\{x=(x_J)_{J\in\Theta_n^m}\in\mathbb{R}^{\Theta_n^m}\,\middle|\,\forall I\in\Theta_n^m,\,\varphi(x)\in\{0,1\}\right\}=\varphi^{-1}\left(\{0,1\}^{\Theta_n^m}\right)$. Now, Mobius-Rota inversion formula recalled in Lemma 3.1 implies that φ is one-to-one. We thus deduce firstly that the intersection (12) can be rewritten as

$$\mathfrak{R}_n^m\cap\mathbb{Z}^{\Theta_n^m}=\varphi^{-1}\left(\{0,1\}^{\Theta_n^m}\cap\varphi(\mathfrak{R}_n^{2,m})\right) \tag{14}$$

but above we have

$$|Res_n^m|=\left|\mathfrak{R}_n^m\cap\mathbb{Z}^{\Theta_n^m}\right|=\left|\{0,1\}^{\Theta_n^m}\cap\varphi(\mathfrak{R}_n^{2,m})\right|. \tag{15}$$

We now compute the image of $\mathfrak{R}_n^{2,m}$ under φ.

Lemma 3.4. *$(y_J)\in\varphi(\mathfrak{R}_n^{2,m})$ if and only if*

$$\forall I\in\Gamma_n^m,\quad 0\le\sum_{\substack{J\in\Theta_n^m\\ J\subset I}}(-1)^{n-m-|J|-1}\binom{|I|-|J|-1}{n-m-|J|-1}y_J\le 1.$$

Proof. Take $y=(y_J)_{J\in\Theta_n^m}\in\varphi(\mathfrak{R}_n^{2,m})$. By definition there exists $x=(x_J)_{J\in\Theta_n^m}$ such that

$$\forall I\in\Gamma_n^m,\quad 0\le\sum_{\substack{J\in\Theta_n^m\\ J\subset I}}x_J\le 1 \tag{16}$$

and such that

$$\forall I\in\Theta_n^m,\quad y_I=\sum_{\substack{J\in\Theta_n^m\\ J\subset I}}x_J.$$

Mobius-Rota inversion formula (Lemma 3.1) implies that

$$\forall I\in\Theta_n^m,\quad x_I=\sum_{\substack{J\in\Theta_n^m\\ J\subset I}}(-1)^{|I|-|J|}y_J.$$

Thus,

$$\begin{aligned}\forall I\in\Gamma_n^m,\quad \sum_{\substack{J\in\Theta_n^m\\ J\subset I}}x_J&=\sum_{\substack{J\in\Theta_n^m\\ J\subset I}}\sum_{\substack{K\in\Theta_n^m\\ K\subset J}}(-1)^{|J|-|K|}y_K\\&=\sum_{\substack{K\in\Theta_n^m\\ K\subset I}}y_K\sum_{\substack{J\in\Theta_n^m\\ K\subset J\subset I}}(-1)^{|J|-|K|}.\end{aligned} \tag{17}$$

Now,

$$\sum_{\substack{K\in\Theta_n^m\\ K\subset J\subset I}}(-1)^{|J|-|K|}=\sum_{j=|K|}^{n-m-1}(-1)^{j-|K|}\binom{|I|-|K|}{j-|K|}$$

The result follows then from the identity : for every positive integers $1<p<n$, we have

$$\sum_{j=0}^{p}(-1)^j\binom{n}{j}=(-1)^p\binom{n-1}{p}.$$

Indeed, if we use the Pascal identity, that is, the identity $\binom{n}{j}=\binom{n-1}{j-1}+\binom{n-1}{j}$ then we get

$$\sum_{j=0}^{p}(-1)^j\binom{n}{j}=\sum_{j=0}^{p}(-1)^j\binom{n-1}{j}+\sum_{j=0}^{p-1}(-1)^{j+1}\binom{n-1}{j}$$

Hence, all the terms at the right-hand side cancel out except the last one, that is, the term $(-1)^p\binom{n-1}{p}$. □

Combining (15) and Lemma 3.4, we get

Proposition 3.5. *$|Res_n^m|$ equals the number of elements (y_J) in $\{0,1\}^{\Theta_n^m}$ which satisfy*

$$\forall I\in\Gamma_n^m,\quad 0\le\sum_{\substack{J\in\Theta_n^m\\ J\subset I}}(-1)^{n-m-|J|-1}\binom{|I|-|J|-1}{n-m-|J|-1}y_J\le 1.$$

We slightly reword the preceding proposition

Corollary 3.2. *$|Res_n^m|$ equals the number of elements (z_J) in $\{0,1\}^{\Theta_n^m}$ which satisfy*

$$\forall I\in\Gamma_n^m,\quad b_I\le\sum_{\substack{J\in\Theta_n^m\\ J\subset I}}\binom{|I|-|J|-1}{n-m-|J|-1}z_J\le b_I+1$$

where

$$\forall I\in\Gamma_n^m,\quad b_I=\sum_{\substack{j=0\\ n-m-1-j\ \text{is odd}}}^{n-m-1}\binom{|I|}{j}\binom{|I|-j-1}{n-m-1-j}.$$

Proof. The proof follows from the change of variables : $z_J = y_J$ if $n-m-|J|-1$ is even and $z_J = 1-y_J$ otherwise. □

We now deduce from Corollary 3.2 that $|Res_n^m|$ can be interpreted as a coefficient of a multivariate polynomial.

Proposition 3.6.

$$|Res_n^m| = \frac{1}{(2i\pi)^{|\Gamma_n^m|}} \oint P(z) \prod_{I\in\Gamma_n^m} z_I^{-(b_I+1)} \frac{dz}{z}$$

where P is the multivariate polynomial defined as

$$\forall z \in \mathbb{C}, \quad P(z) = \prod_{I\in\Gamma_n^m} (1+z_I) \prod_{J\in\Theta_n^m} \left(1 + \prod_{\substack{I\in\Gamma_n^m \\ J\subset I}} z_I^{a_{I,J}}\right)$$

with

$$\forall (I,J) \in \Gamma_n^m \times \Theta_n^m, \quad a_{I,J} = \binom{|I|-|J|-1}{n-m-|J|-1}.$$

Proof. Note that

$$\prod_{I\in\Gamma_n^m} (1+z_I) \prod_{J\in\Theta_n^m} \left(1 + \prod_{\substack{I\in\Gamma_n^m \\ J\subset I}} z_I^{a_{I,J}}\right) = \prod_{I\in\Gamma_n^m} \left(\sum_{\epsilon_I=0}^{1} z^{\epsilon_I}\right) \prod_{J\in\Theta_n^m} \left(\sum_{\eta_J=0}^{1} z_I^{\eta_J a_{I,J}}\right)$$

$$= \sum_{(\epsilon,\eta)\in\{0,1\}^{\Gamma_n^m\times\Theta_m^n}} \prod_{I\in\Gamma_n^m} z_I^{\sum_{I\in\Gamma_n^m}\epsilon_I + \sum_{\substack{J\in\Theta_n^m \\ J\subset I}} a_{I,J}\eta_J}$$

Now, the coefficient of the monomial $\prod_{I\in\Gamma_n^m} z_I^{b_I+1}$ is

$$\sum_{\epsilon\in\{0,1\}^{\Gamma_n^m}} \left|\{\eta = (\eta_J)_{J\in\Theta^n} \in \{0,1\}^{\Theta^n} \mid \forall I \in \Gamma_n^m, \sum_{\substack{J\in\Theta_n^m \\ J\subset I}} a_{I,J}\eta_J = b_I + \epsilon_I\}\right|,$$

where $a_{I,J} = \binom{|I|-|J|-1}{n-m-|J|-1}$, which is equal to $|Res_n^m|$ according to Corollary 3.2. The result follows then from Proposition 3.1. □

References

1. Matthias Beck and Dennis Pixton : The Ehrhart polynomial of the Birkhoff polytope, Discrete Comput. Geom. **30** (2003), 623-637.

2. P. Camion and C. Carlet and P. Charpin and N. Sendrier : On correlation-immune functions. Crypto'91, Advance in Cryptology, Lecture Notes in Computer Science **576** (1991), 86–100.
3. C. Carlet : Boolean functions for cryptography and error correcting codes. Boolean Methods and Models. Cambridge University Press, 2007. To appear.
4. C. Carlet and A. Gouget : An upper bound on the number of m-resilient Boolean functions. ASIACRYPT 2002, Advances in Cryptology, Lecture Notes in Computer Science **2501** (2002), 484–496.
5. C. Carlet and P. Guillot : A new representation of Boolean functions. AAECC'13, Lecture Notes in Computer Science **1719** (1999), 94–103.
6. C. Carlet and P. Guillot : Bent, resilient Functions and the Numerical Normal Form. DIMACS, Discrete Mathematics and Theoretical Computer Science (2001), 87–96.
7. C. Carlet and A. Klapper : Upper bounds on the numbers of resilient functions and of bent functions. 23rd Symposium on Information Theory in the Benelux, Louvain-La-Neuve, Belgique, Mays, 2002.
8. B. Guo and Y. Yang : Further enumerating Boolean functions of cryptographic significance. Journal of Cryptology **8** (1995), 115–122
9. F. Harary and E. M. Palmer :*Graphical Enumeration*. Academic, New York, 1973.
10. J-M. Le Bars and A. Viola : Equivalence classes of Boolean functions for first-order correlation.
11. S. Maitra and P. Sarkar : Enumeration of correlation immune Boolean functions. ACISP (1999), 12-25.
12. C. J. Mitchell : Enumerating Boolean functions of cryptographic significance. Journal of Cryptology **2** (1990), 155–170.
13. S.M. Park, S. Lee, S. H. Sung and K. Kim : Improving bounds for the number of correlation-immune Boolean functions. Information Processing Letters **61** (1997), 209–212.
14. M. Schneider : A note on the construction and upper bounds of correlation-immune functions. 6th IMA Conference, 1997, 295–306.
15. T.Siegenthaler : Correlation-immunity of nonlinear combining functions for cryptographic applications. IEEE Transactions on Information Theory **30** (1984), 776–780.
16. Y. X. Yang and B. Guo : Further enumerating Boolean funvtions of cryptographic significance. Journal of Cryptology **8** (1995), 115–122.
17. Jian-Zhoua Zhang and Zhi-Shenga You and Zheng-Liang Li : Enumeration of binary orthogonal arrays of strength 1. Discrete Mathematics **239** (2001), 191–198.

On Quadratic Extensions of Cyclic Projective Planes

Hiu Fai Law

Department of Mathematics,
University of Hong Kong, Hong Kong
E-mail : hflaw@graduate.hku.hk

Philip P. W. Wong

Department of Mathematics,
University of Hong Kong, Hong Kong
E-mail : ppwwong@maths.hku.hk

1. Introduction

A finite Desarguesian projective plane can be characterized by properties of its collineation group. The classical result of Ostrom-Wagner [1959] asserts that a finite projective plane is Desarguesian if it admits a doubly transitive collineation group. Building on this, Ott [1975] proves that two cyclic collineation groups are sufficient for a plane to be Desarguesian, and later, Ho [1998] generalizes the result and proves that two abelian collineation groups are sufficient. On the other hand, Singer [1938] proves that a Desarguesian plane is cyclic, i.e. it admits a cyclic collineation group. In view of the results above it is important to ask whether the converse is true, namely, whether a cyclic projective plane is Desarguesian. This is an open problem. Bruck [1960] obtained the following partial results: using the fact that equivalent cyclic difference sets give rise to isomorphic cyclic projective planes, Bruck demonstrates that the converse is true for small square orders by showing the uniqueness of a quadratic extension of a cyclic difference set. It is remarkable that in his paper, Bruck noted that the existence of a foliation structure for cyclic projective planes of square order provided the motivation for his work, although he did not make use of it to obtain the results.

In this article, our aim is to show the foliation structure of a cyclic projective plane of square order, and devise from it a canonical approach

to recover Bruck's results. Furthermore, we are able to prove two new cases, namely, cyclic projective planes of order 121 or 256 are Desarguesian.

We first study the structure of a cyclic difference set of order m^2. Let $s = m^2 - m + 1, t = m^2 + m + 1$. We introduce the key notion of *good points* defined by a cyclic difference set, and define a function $g : \mathbb{Z}_s \longrightarrow \mathbb{Z}_t$, such that $g(\mathbb{Z}_s \backslash \{0\}) = GP$ is the set of good points. We prove the following version of Bruck's result.

Theorem 1.1. (Lemma 3.1, (8).) *Let D_{m^2} be a cyclic difference set of order m^2 fixed by all multipliers. Then D_{m^2} has the unique normal form given by $D_{m^2} = sD_m \cup E(g)$, where $D_m = \{d_0, d_1, \cdots, d_m \mid d_1 - d_0 \equiv 1 \ (mod\ t)\}$ is a cyclic difference set of order m, and $E(g)$ is a set defined by a function $g : \mathbb{Z}_s \longrightarrow \mathbb{Z}_t$, which satisfies Condition C and $g(0) = d_0$.*

Geometrically, this means that a cyclic projective plane π' of order m^2 consists of s mutually isomorphic cyclic projective planes of order m, each of which is called a *leaf*. Every line in π' restricts to a line in a leaf, and intersects each of the other leaves in exactly one point. Condition C is a translation of the projective plane axioms with respect to the foliation structure of π' just described.

The function g plays the following role. Call a translate of D_{m^2} a *long line*. Arrange all the long lines cyclically in an $(m^2 + 1) \times st$ array, and take the first and every other s columns from it to form a smaller array. The columns of this smaller array restrict to lines in the same leaf π. The function g describes how the remaining long lines can be recovered from this array, by matching each leaf to a good point in π.

Condition C is not only necessary for $sD_m \cup E(g)$ to be a cyclic difference set; it is also sufficient.

Theorem 1.2. (Theorem 5.1, Theorem 5.2.) *Let*

$$D_m = \{d_0, d_1, \cdots, d_m \mid d_1 - d_0 \equiv 1 \ (mod\ t)\}$$

be a cyclic difference set D_m of order m, and a function $g : \mathbb{Z}_s \longrightarrow \mathbb{Z}_t$ satisfying Condition C and $g(0) = d_0$. Then the set $sD_m \cup E(g)$ is a cyclic difference set of order m^2.

Hence, the problem is reduced to finding solutions for g. The combinatorial complexity is intractable with Condition C as the only criterion. We make use of the theory of multipliers to reduce the construction of g to the cycle level: if the sets GP and $\mathbb{Z}_s \backslash \{0\}$ are partitioned into cycles by a multiplier, then g must respect such cycle structures.

While the complexity is significantly reduced by the use of multiplier, we derive yet further necessary conditions imposed by the foliation structure of a quadratic extension. We study the distribution of the good points on the lines of a cyclic projective plane of order m. The lines can be classified into three types by the good points. In a cyclic projective plane of order m^2 containing $\pi(sD_m)$, there exists a special point called the *singular point* on each long line which must lie on the extension of a line of a certain type called a *singular line*. This gives rise to a further condition which a solution g must satisfy. With this, the combinatorial complexity is brought down to a manageable level.

Next we address the uniqueness issue. Given a cyclic difference set of order m, we introduce a notion of equivalence between two solutions g and g', and show that equivalent solutions define equivalent cyclic difference sets of order m^2. Thus, if there is a unique solution for a given order m, then there is a unique extension. It follows that if there exists a unique cyclic difference set of order m, then all cyclic difference sets of order m^2 are equivalent (see Theorem 6.1). We show that this is the case for m a prime power at most 16 other than 13.

Theorem 1.3. (Theorem 8.1) *A cyclic difference set of order m or m^2 is unique, where $m = 2, 3, 4, 5, 7, 8, 9, 11, 16$.*

From this we conclude the main result:

Theorem 1.4. (Theorem 8.2) *A cyclic projective plane of order m or m^2 is Desarguesian, where $m = 2, 3, 4, 5, 7, 8, 9, 11, 16$.*

In Section 1, we summarize the results needed from the theory of cyclic difference sets for our purpose.

In Section 2, we study the foliation structure of a cyclic projective plane of square order and Condition C. These are summarized in the *Structure Theorem* (Theorem 3.1).

In Section 3, we obtain further necessary conditions through the study of good points and singular lines.

In Section 4, we prove that the necessary conditions on g are in fact sufficient for the construction of a quadratic extension of a cyclic projective plane.

In Section 5, we treat the problem of uniqueness.

In Section 6, we make use of the theory of multipliers to develop a search strategy for the solutions g. We also illustrate this strategy through a thorough discussion of the case $m = 5$.

In Section 7, we summarize our main results and present the data obtained from the search for cases $m = 2, 3, 4, 5, 7, 8, 9, 11, 16$.

2. Planar Difference Sets and Cyclic Projective Planes

A finite projective plane is a system consisting of a finite set of points, and a collection of its subsets, called lines, together with an incidence relation, such that (i) two distinct points lie on a unique line; (ii) two distinct lines intersect at a unique point; (iii) there exist four points no three of which are collinear. Every line has the same number of points and every point point is on the same number of lines. A projective plane has *order* m if there are $m + 1$ points on a line. A collineation is a permutation of the points of the plane sending a line to a line.

A finite projective plane is said to be *cyclic* if it admits a collineation which permutes the points in one cycle. We call such a collineation a *cyclic collineation*. It is a standard result that this collineation also permutes the lines in one cycle (see, for example, Hughes and Piper [1973], p.256).

Definition 2.1. A planar cyclic difference set D of order m is a $(m + 1)$-subset of $\mathbb{Z}_{m^2+m+1}$ such that every nonzero element in $\mathbb{Z}_{m^2+m+1}$ can be expressed as a unique difference between elements in D.

In this article, a cyclic difference set is a planar cyclic difference set. Given a cyclic difference set, we can define a cyclic projective plane. Indeed, if D is a cyclic difference set of order m, then $\pi(D)$ is a cyclic projective plane whose points are $\mathbb{Z}_{m^2+m+1}$ and lines are $D, D+1, \cdots, D+(m^2+m)$.

Definition 2.2. Two cyclic difference sets D, D' of order m are said to be *equivalent* if there exist integers a, b such that $D = aD' + b$.

Definition 2.3. Two projective planes are said to be *isomorphic* if there exists a one-to-one correspondence between the points of the planes which sends collinear points to collinear points.

Lemma 2.1. *Equivalent cyclic difference sets define isomorphic cyclic projective planes.*

Proof. If $D = aD' + b$, then $x \mapsto ax + b$ is an isomorphism from $\pi(D)$ to $\pi(D')$. □

Thus, to prove that a cyclic projective plane of order m is unique, it suffices to show that cyclic difference sets of order m are all equivalent.

On the other hand, given a cyclic projective plane π of order m with cyclic collineation α. Let the points of π be labelled by elements of $\mathbb{Z}_{m^2+m+1}$, so that α is given by $(0\ 1\ 2\ \cdots m^2+m)$. Consider a line $l = \{d_0, d_1, \cdots, d_m\}$ in π. The axioms of a projective plane show that l is a cyclic difference set D of order m. It is easy to see that $\pi(D)$ is isomorphic to π. Note that there exists a unique translate D' of D whose elements sum to zero.

To study difference sets further, we need the notion of a multiplier.

Definition 2.4. Let D be a cyclic difference set of order m. A multiplier of D is an integer p such that $pD = D + k$ for some $k \in \mathbb{Z}_{m^2+m+1}$.

Theorem 2.1. *(Multiplier Theorem) Let D be a cyclic difference set of order m. If p is a prime divisor of m, then p is a multiplier of D.*

Proof. See Hall [1947], p.411. □

Let p be a multiplier of D. Then p induces a collineation on $\pi(D)$ by sending $x \in \pi(D)$ to $px \in \pi(D)$. A cyclic difference set D is said to be *fixed* by a multiplier p if $pD = D$.

Lemma 2.2. *Given a cyclic difference set D_m of order m, there exists at least one, and at most three translates of D_m which are fixed by all multipliers.*

Proof. See Baumert [1971], p.79. □

Let $\mathcal{D}(m)$ be the set of all cyclic difference sets of order m each of whose elements sum to zero. It is readily verified that every element in $\mathcal{D}(m)$ is fixed by all its multipliers.

The following lemma is useful in the construction of a cyclic difference set.

Lemma 2.3. *Let p be a multiplier of D_m which lies in $\mathcal{D}(m)$. Then $D_m = \bigcup_i (x_i)$, where each (x_i) is an p-orbit. Moreover, $\sum_i |(x_i)| = m + 1$.*

Proof. Since $pD_m = D_m$, it follows that if $x \in D_m$, then $px \in D_m$. Repeating the argument, we see that D_m contains the whole p-orbit of x. Thus, D_m is a union of disjoint p-orbits. □

An important class of cyclic projective planes is provided by the Desarguesian planes. A projective plane is said to be *Desarguesian* if centrally

perspective triangles are axially perspective. By a result of Singer [1938], a finite Desarguesian plane is cyclic. In this article, we study to what extent the converse holds in the case of a cyclic projective plane of square order.

3. Structure Theorem

Let m be an integer greater than 1, $t = m^2 + m + 1, s = m^2 - m + 1$. Then $st = m^4 + m^2 + 1$. Let $\pi(D_{m^2})$ be the cyclic projective plane of order m^2, whose points are $\mathbb{Z}_{st}$ and lines are the translates of a cyclic difference set D_{m^2}, and with cyclic structure given by the collineation $(0\ 1\ \cdots st-1)$. We shall assume that $D_{m^2} \in \mathcal{D}(m^2)$. This implies that D_{m^2} is fixed by all multipliers.

Since $(s, m-1) = 1$, $m^3 x \equiv x \pmod{st}$ if and only if $x \equiv 0 \pmod{s}$. From this we deduce the following fact. Let $su \in \mathbb{Z}_{st}\backslash\{0\}$. Then $su \equiv d - d' \pmod{st}$ for a unique pair $d, d' \in D_{m^2}$. By multiplying both sides of the equation by m^3, we obtain $su \equiv m^3 d - m^3 d' \pmod{st}$. By uniqueness, $m^3 d \equiv d, m^3 d' \equiv d' \pmod{st}$. Thus, $d \equiv 0 \pmod{s}$ and $d' \equiv 0 \pmod{s}$. We now give a different proof to the following lemma of [Bruck 1960]:

Lemma 3.1. *Suppose $D_{m^2} \in \mathcal{D}(m^2)$. Then $D_{m^2} = sD_m \cup E$, where $D_m = \{d_0, d_1, \cdots, d_m \mid d_1 - d_0 = 1 \ (mod\ t)\}$ is a cyclic difference set of order m, and $E\ (mod\ s) \equiv \mathbb{Z}_s\backslash\{0\}$.*

Proof. Let x_i be the number of elements of D_{m^2} which are congruent to $i \pmod{s}$, for $i = 0, 1, \cdots, s-1$. Only elements in the same congruence class $\pmod{s}$ give rise to differences in $\{s, 2s, \cdots, (t-1)s\}$. Thus, $\sum_{i=0}^{s-1} x_i(x_i - 1) = t - 1 = m^2 + m$. By the fact obtained preceding the lemma, we see that $\sum_{i=1}^{s-1} x_i(x_i - 1) = 0$. Thus, $x_0(x_0 - 1) = m^2 + m$, i.e. $x_0 = m + 1$. These $m + 1$ elements form the set sD_m. It is readily verified that D_m is a cyclic difference set of order m. We may arrange to have $d_1 - d_0 = 1 (\text{mod}\ t)$. Furthermore, since D_{m^2} has $m^2 + 1$ elements, $\sum_{i=0}^{s-1} x_i = m^2 + 1$. Then $\sum_{i=0}^{s-1}(x_i - 1)^2 = \sum_{i=0}^{s-1} x_i(x_i - 1) - \sum_{i=0}^{s-1} x_i + \sum_{i=0}^{s-1} 1 = m^2$. Putting $x_0 = m+1$, we obtain $\sum_{i=1}^{s-1}(x_i - 1)^2 = 0$. Thus, $x_i = 1$ for all $i = 1, 2, \cdots, s-1$. These $s - 1$ elements form the set E. □

Henceforth in this section, we shall assume D_{m^2} is of the form given in Lemma 3.1.

We arrange the lines of $\pi(D_{m^2})$ in an $(m^2 + 1)$ by st array, called the *full array* (Fig. 1). The bottom row of the array consists of the points of the plane. In the first columns, $e_i \equiv i \pmod{s}$ and $\{d_0, d_1, \cdots, d_m\} = D_m$.

Each of the remaining columns is obtained by adding 1 to the preceding column.

$$\begin{array}{|ccccc|}\hline e_{s-1} & e_{s-1}+1 & e_{s-1}+2 & \cdots & e_{s-1}+st-1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ e_2 & e_2+1 & e_2+2 & \cdots & e_2+st-1 \\ e_1 & e_1+1 & e_1+2 & \cdots & e_1+st-1 \\ sd_m & sd_m+1 & sd_m+2 & \cdots & sd_m+st-1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ sd_1 & sd_1+1 & sd_1+2 & \cdots & sd_1+st-1 \\ sd_0 & sd_0+1 & sd_0+2 & \cdots & sd_0+st-1 \\ \hline\end{array}$$

Fig. 1 *The full array*

Let $S_0 = \{0, s, 2s, \cdots, (t-1)s\} \subset \mathbb{Z}_{st}$. Then S_0 is a subplane of $\pi(D_{m^2})$ whose lines are given by $sD_m + su \pmod{st}$, $u = 0, 1, \cdots, t-1$, with cyclic structure given by $(0\ s\ 2s \cdots (t-1)s)$.

With respect to S_0, the plane $\pi(D_{m^2})$ has a foliation structure consisting of s subplanes, namely, $S_0, S_1, \cdots, S_{s-1}$, where each $S_i = \{i + su \mid u = 0, 1, \cdots, t-1\}$, with lines given by $\{i + sD_m + su \mid u = 0, \cdots, t-1\}$. Note that each S_i is isomorphic to S_0 via the isomorphism $x \mapsto x - i$. We shall refer to these s subplanes as the *leaves* of the foliation, and in particular we shall refer to the leaf S_0 as the *basic leaf*.

In the form of an array, this foliation structure is represented by an m^2+1 by t array obtained by extracting the first and every other s columns from the full array. We shall refer to this array as the *standard array* (Fig. 2).

$$\begin{array}{rc}
S_{s-1} & \left|\begin{array}{cccc} (0,s-1) & (1,s-1) & \cdots & (t-1,s-1) \\ \vdots & \vdots & \vdots & \vdots \end{array}\right. \\
S_2 & \left|\begin{array}{cccc} (0,2) & (1,2) & \cdots & (t-1,2) \end{array}\right. \\
S_1 & \left|\begin{array}{cccc} (0,1) & (1,1) & \cdots & (t-1,1) \end{array}\right. \\
 & \left|\begin{array}{cccc} sd_m & sd_m+s & \cdots & sd_m+(t-1)s \\ \vdots & \vdots & \vdots & \vdots \end{array}\right. \\
S_0 & \left|\begin{array}{cccc} sd_1 & sd_1+s & \cdots & sd_1+(t-1)s \\ sd_0 & sd_0+s & \cdots & sd_0+(t-1)s \end{array}\right.
\end{array}$$

Fig. 2 *The standard array*

Remark 3.1. We have used the ordered pairs (u, i) to denote the entries $e_i + su$, and $(u, 0)$ for the entries $sd_0 + su$. Note that u indicates the column

and i indicates the row. In the standard array, we shall refer to the columns as column 0, column 1, $\cdots$, column $t-1$.

Remark 3.2. Arithmetic in the first coordinate of an ordered pair is in $\mathbb{Z}_t$, while that in the second coordinate is in $\mathbb{Z}_s$. We reserve the symbol $\frac{a}{b}$ or a/b, for the division of a by b as integers.

A line in $\pi(D_{m^2})$ which restricts to a line in the subplane S_i is called an S_i-*leaf line*. A line in $\pi(D_{m^2})$ which restricts to a line in the subplane S_0 is called an S_0-*leaf line* or a *basic leaf line*. The S_i-leaf line defined by $(0,i)$ and $(1,i)$ is denoted by l_i. Note that by Lemma 3.1, every line in $\pi(D_{m^2})$ is either an S_i-leaf line or a basic leaf line, and it intersects each of the other leaves at one point.

We describe the points on l_i. Consider the line $D_{m^2}+(0,i)-sd_0$. Since it contains $(0,i)$ and $(1,i)$, it is equal to l_i. It follows that $(0,j)+(0,i)-sd_0$ is on l_i. Switching i and j, we see that

$$l_i \cdot l_j = (0,i)+(0,j)-sd_0 = (\frac{(0,i)+(0,j)-(0,i+j)-sd_0}{s}, i+j). \quad (1)$$

Lemma 3.2. *$l_i = \{(0,i),(1,i),(d_2-d_0,i),\cdots,(d_m-d_0,i);\ l_i\cdot l_j;\ \tilde{i} \mid j = 1,2,\cdots,s-1, j\neq i\}$, where $l_i\cdot l_j$ is given by (1), and*

$$\tilde{i} = (0,i)+(0,i)-sd_0 = (\frac{(0,i)+(0,i)-(0,i+i)-sd_0}{s}, i+i).$$

Proof. Since D_{m^2} is a cyclic difference set, no three leaf lines l_i, l_j, l_k are concurrent, and so all $l_i\cdot l_j, j\neq i$ are distinct. As the size of l_i is m^2+1, there is a point $\tilde{i}$ on l_i which lies neither on S_i nor any other leaf lines l_j. Consequently, $\tilde{i}$ is an element of S_{2i} and is given by substituting $j=i$ in (1). □

We now introduce the notion of good points and show that $l_i\cdot l_j$ is determined by the good points via a map g as follows.

Definition 3.1. Let $g:\mathbb{Z}_s \mapsto \mathbb{Z}_t$ be defined by $g(i) = (l_i\cdot l_{-i})/s$ and $g(0)=d_0$.

Since $(0,i)=e_i \equiv i \pmod s$, $l_i\cdot l_{-i}\equiv 0 \pmod s$ and g is well defined.

Definition 3.2. $GP = \mathbb{Z}_t \backslash \bigcup_{i=0}^{m}(D_m+d_i-d_0)$ is the set of *good points*.

Lemma 3.3. $g(\mathbb{Z}_s\backslash\{0\}) = GP$.

Proof. By (1), $sg(i) = (((0,i)+(0,-i)-(0,0)-sd_0)/s, 0) = (((0,i)+(0,-i)-2sd_0)/s, 0)$. Since D_{m^2} is a cyclic difference set, the column number of $sg(i)$ cannot be $d_v - d_0$, for any $v = 0, 1, \cdots, m$. It follows that $g(i) \in GP$. By the same reason, $g(i) \neq g(j)$ unless $i = \pm j$. Now by definition $g(i) = g(-i)$, and since the size of GP is $(s-1)/2$, g is surjective. □

Lemma 3.4. $(0,i)+(0,j)-(0,i+j)-sd_0 \equiv (0,-i)+(0,-j)-(0,-i-j)-sd_0$ *(mod st).*

Proof. Since $m^3 \equiv 1 \pmod{t}$, we have

$$(0,i)+(0,j)-(0,i+j)-sd_0 \equiv m^3((0,i)+(0,j)-(0,i+j)-sd_0) \pmod{t}.$$

The congruence holds in $\mathbb{Z}_{st}$ since both sides are congruent to 0 (mod s) and $(s,t) = 1$. Since m^3 fixes D_{m^2} and $m^3 \equiv -1 \pmod{s}$, $m^3(0,i) \equiv (0, s-i) \pmod{st}$. It follows that the RHS of the congruence is equivalent to $(0,-i)+(0,-j)-(0,-i-j)-sd_0 \pmod{st}$. □

Using Lemma 3.4 and the definition of g, we can rewrite (1) as follows:

$$l_i \cdot l_j = (2^{-1}(g(i)+g(j)-g(i+j)-d_0), i+j). \tag{2}$$

When $i = j$, we have

$$\tilde{i} = (2^{-1}(g(i)+g(i)-g(i+i)-d_0), i+i). \tag{3}$$

Definition 3.3. $\tilde{i}$ given by (3) is called the *singular point* of l_i.

We say that row i is *dual* to row j if $i + j \equiv 0 \pmod{s}$. By (2), $l_i \cdot l_j$ and $l_{-i} \cdot l_{-j}$ lie in the same column but in dual rows. By (3), the same is true for the singular points $\tilde{i}, \widetilde{-i}$. This leads to the following interesting configuration of the points of the leaf lines: for l_i and l_{-i}, there is mirror symmetry across the separation between row $2^{-1}(s-1)$ and row $2^{-1}(s-1)+1$.

We now have a description of the lines l_i in terms of the mapping g and the good points.

Lemma 3.5. $l_i = \{(d_v - d_0, i), (2^{-1}(g(i)+g(j)-g(i+j)-d_0), i+j) \mid v = 0, 1, \cdots, m, j \in \mathbb{Z}_s \backslash \{0\}\}$.

Proof. Substitute (2), (3) in Lemma 3.2, and note that $d_1 - d_0 = 1 \pmod{t}$. □

We shall refer to the position of the points of l_i in the standard array as its *configuration*.

We summarize the above results in the following structure theorem for a cyclic projective plane of order m^2.

Theorem 3.1 (Structure Theorem). *Let m be an integer greater than 1, $s = m^2 - m + 1, t = m^2 + m + 1$. A cyclic projective plane $\pi(D_{m^2})$ of order m^2 consists of the following:*

(1) The full array given by Fig. 1 and its associated standard array given by Fig. 2.
(2) The rows $S_0, S_1, \cdots, S_{s-1}$ in Fig. 2 constitute the st points of $\pi(D_{m^2})$.
(3) $\{S_0, S_1, \cdots, S_{s-1}\}$ is a set of mutually isomorphic cyclic projective planes of order m, which are called the leaves of $\pi(D_{m^2})$; in particular S_0 is called the basic leaf.
(4) The columns of the standard array are the t lines in $\pi(D_{m^2})$ called basic leaf lines; the first column is the cyclic difference set D_{m^2} of order m^2 given by Lemma 3.1, with the bottom $m+1$ rows congruent to $0 \pmod s$, and the remaining rows congruent to $1, 2, \cdots, s-1$, respectively; the columns of the standard array restricted to the bottom $m+1$ rows are the lines of S_0.
(5) The remaining $t(s-1)$ lines of $\pi(D_{m^2})$ are the $(s-1)$-leaf lines l_i given by Lemma 3.2, together with their translates $l_i + s, l_i + 2s, \cdots, l_i + (t-1)s$. The configuration of each l_i is determined by Lemma 3.5 via the mapping g given by Definition 3.1 and Lemma 3.3.

We shall henceforth refer to the structure of $\pi(D_{m^2})$ given in Theorem 3.1 as its *foliation structure*.

In order to construct the quadratic extension of a cyclic projective plane, we proceed to derive a set of necessary conditions for its existence in terms of g by extracting from the structure theorem the requirement on the configuration of lines imposed by the axioms for a projective plane. Let

$$c(i,j) = 2^{-1}(g(i) + g(j) - g(i+j) - d_0), \tag{4}$$

where g is given by Definition 3.1.

First we study the axiom: every pair of distinct lines must intersect at a unique point. There are three cases to consider:

(i) the intersection between two basic leaf lines,
(ii) the intersection between a basic leaf line and a non-basic leaf line,
(iii) the intersection between two non-basic leaf lines.

For (i), the axiom imposes no condition on g since any two basic leaf lines restrict to lines in the basic leaf S_0, and S_0 is a projective plane.

For (ii), in the case where the non-basic leaf line is l_i, the axiom is equivalent to the condition that for each i, $\{d,\ g(i)-d_0-d,\ c(i,j) \mid j \neq -i; d \in D_m - d_0\} = \mathbb{Z}_t$. Note that there is no loss of generality in considering only l_i, since all other non-basic leaf lines are translates of it.

For (iii), we may assume as in (ii) that one of the non-basic leaf lines is l_i. First consider the case where the other non-basic leaf line is an S_i-leaf line. Here the axiom imposes no condition on g since each S_i is a projective plane. The other case is when the other non-basic leaf line is an S_j-leaf line, where $j \neq i$. Here the condition is that for each pair of distinct i, j, $\{c(i,j-i)-d,\ d-c(j,i-j),\ c(i,k-i)-c(j,k-j) \mid k = 0,1,\cdots,s-1, k \neq i,j; d \in D_m - d_0\} = \mathbb{Z}_t$. It is straightforward to verify that these conditions are equivalent to the conditions in (ii).

Similarly, one verifies that the other projective plane axioms do not yield further conditions. Thus, in view of the foliation structure of a cyclic projective plane of order m^2, the necessary conditions for the construction of a quadratic extension of a given cyclic projective plane of order m are given as follows:

Condition 3.1. For $i = 1, 2, \cdots, s-1$,

$$\{d,\ g(i)-d_0-d,\ c(i,j) \mid j = 1,2,\cdots,s-1, j \neq -i;\ d \in D_m - d_0\} = \mathbb{Z}_t,$$

where $c(i,j) = 2^{-1}(g(i)+g(j)-g(i+j)-d_0)$, and g is defined in Definition 3.1.

4. Distribution of Good Points and Further Necessary Conditions

In this section we investigate further necessary conditions on the mapping g. Here the multipliers play an essential role, as follows. Recall from Section 3 that l_i is the line defined by $(0,i)$ and $(1,i)$.

Lemma 4.1. *Let p be a prime divisor of m. Let $\Phi_p : \mathbb{Z}_{st} \longrightarrow \mathbb{Z}_{st}$ be defined by $x \mapsto px + s(p-1)d_0$. Then $\Phi_p(l_i) = l_{pi}$.*

Proof. Note that by Theorem 2.1, p is a multiplier of D_m and D_{m^2}, and thus Φ_p is a collineation of $\pi(D_{m^2})$. Now $p(D_m - d_0) + (p-1)d_0 = D_m - d_0$. Thus $0 + (p-1)d_0$ and $p + (p-1)d_0$ lie on $D_m - d_0$. Hence, $\Phi_p(l_i) = \Phi_p((0,i)\cdot(1,i)) = ((p-1)d_0, pi)\cdot(p+(p-1)d_0, pi) = l_{pi}$. □

Let $\varphi_p : \mathbb{Z}_t \longrightarrow \mathbb{Z}_t$ be defined by

$$\varphi_p : x \mapsto px + (p-1)d_0. \tag{5}$$

Corollary 4.1. $g(pi) = \varphi_p(g(i))$.

Proof. Since Φ_p fixes S_0 and takes l_i to l_{pi}, it takes $l_i \cap S_0 = sg(i)$ to $l_{pi} \cap S_0 = sg(pi)$, i.e. $\Phi_p(sg(i)) = sg(pi)$. Divide the equation by s to complete the proof. □

Since $(p, s) = 1$, we may write $\mathbb{Z}_s \backslash \{0\}$ as a union of p-orbits (x). By Corollary 4.1, the good points defined by the leaf lines $l_i, i = 1, 2, \cdots, s-1$, are permuted by φ_p, i.e. $g(pi) = \varphi_p(g(i))$. Thus, g is determined once it is defined for an element in each orbit of $\mathbb{Z}_s$ under p.

Next we study the distribution of the good points on the lines of $\pi(D_m)$. Recall that the set of good points GP is the complement in $\pi(D_m)$ of $\{D_m, D_m + d_1 - d_0, \cdots, D_m + d_m - d_0\}$.

Lemma 4.2. *The t lines in $\pi(D_m)$ are divided into three groups:*

(1) m odd:

- *(a)* $A = \{D_m, D_m + d_1 - d_0, \cdots, D_m + d_m - d_0\}$ *and* $|A| = m + 1$.
- *(b)* $B = \{$*the set of lines containing* $(m-1)/2$ *good points*$\}$ *and* $|B| = (t-1)/2$.
- *(c)* $C = \{$*the set of lines containing* $(m+1)/2$ *good points*$\}$ *and* $|C| = (s-1)/2$.

(2) m even:

- *(a)* $A = \{D_m, D_m + d_1 - d_0, \cdots, D_m + d_m - d_0\}$ *and* $|A| = m + 1$.
- *(b)* $B = \{$*the set of lines containing* $m/2$ *good points*$\}$ *and* $|B| = m^2 - 1$.
- *(c)* $C = \{2D_m - d_0\}$ *and* $|C| = 1$.

Proof. Since D_m is a cyclic difference set, the lines in A are in general position. Call x a single (resp. double) point, if it lies on one (resp. two) line in A. If a line not in A contains k single points, then it contains $(m+1-k)/2$ double points and $(m+1-k)/2$ good points. Let x_k be the number of such lines. Clearly, for odd m, $x_k = 0$ if k is odd, as $m + 1$ is even; for even m, $x_k = 0$ if k is even, as $m + 1$ is odd. By counting the number of lines containing at least two single points out of $m + 1$ of them and at least two double points out of $m(m + 1)/2$ respectively, we have,

(1) m odd:

$$x_0 + x_2 + x_4 + \cdots + x_{m+1} = m^2$$

$$C_2^m(m+1) + C_2^{\frac{m+1}{2}}x_0 + C_2^{\frac{m-1}{2}}x_2 + C_2^{\frac{m-3}{2}}x_4 + \cdots + C_2^2 x_{m-3} = C_2^{\frac{m(m+1)}{2}}$$
$$C_2^2 x_2 + C_2^4 x_4 + \cdots + C_2^{m+1}x_{m+1} = C_2^{m+1}.$$

Since each x_i must be a non-negative integer, the solution to this system is $x_0 = (s-1)/2, x_2 = (t-1)/2, x_k = 0$ for all other k.

(2) m even:

$$x_1 + x_3 + x_5 + \cdots + x_{m+1} = m^2$$
$$C_2^m(m+1) + C_2^{\frac{m}{2}}x_1 + C_2^{\frac{m}{2}-1}x_3 + \cdots + C_2^2 x_{m-3} = C_2^{\frac{m(m+1)}{2}}$$
$$C_2^3 x_3 + C_2^5 x_5 + \cdots C_2^{m+1}x_{m+1} = C_2^{m+1}.$$

All the single points form the set $2D_m - d_0$, which is a line in the plane since 2 is a multiplier. Thus, any line containing two or more single points is $2D_m - d_0$, hence $x_{m+1} = 1$, and so $x_1 = m^2 - 1$ and $x_k = 0$ for all other k. □

To illustrate, we list the composition of lines for the cases where $m = 5$ and $m = 4$, using s, d, g to indicate a single point, a double point and a good point respectively.

Line type	Composition of line	Type size
A	$d\ d\ d\ d\ d\ s$	6
B	$g\ g\ d\ d\ s\ s$	15
C	$g\ g\ g\ d\ d\ d$	10

$m = 5$

Line type	Composition of line	Type size
A	$d\ d\ d\ d\ s$	5
B	$g\ g\ d\ d\ s$	15
C	$s\ s\ s\ s\ s$	1

$m = 4$

We now introduce the concept of a singular line.

Definition 4.1. The extension of a line in C of Lemma 4.2 to a basic leaf line in $\pi(D_{m^2})$ is called a *singular line*.

Definition 4.2. $SL = \{x \mid D_m + x$ is a line in $C\}$.

Thus, $x \in SL$ is the column number of a singular line. In other words, each singular line is given by

$$D_{m^2} + sx = \{sd_0 + sx, \cdots, sd_m + sx, (x,1), \cdots, (x, s-1)\},$$

for some $x \in SL$.

We study the relation between the singular lines and the singular points.

Lemma 4.3. *Let $m > 1$ be a positive odd integer.*

(1) Each singular line contains two dual singular points.
(2) Each singular point is on a singular line.

Proof. (1) Each basic leaf line must intersect the $s-1$ leaf lines, $l_1, l_2, \cdots, l_{s-1}$. Since a singular line contains $(m+1)/2$ good points, it intersects $2 \times (m+1)/2$ of these leaf lines at S_0. Note that each l_i intersects a singular line either at a singular point or at an intersection with other leaf lines. Moreover, if a singular line contains $l_i \cdot l_j$ then it also contains $l_{-i} \cdot l_{-j}$; this accounts for a multiple of four leaf lines. But 4 does not divide $(s-1)-(m+1)$ for odd m. Therefore, a singular line must contain a singular point, as well as its dual. As there are twice as many singular points as singular lines, therefore every singular line contains exactly two singular points.

(2) By (1), each singular line contains two singular points. By Lemma 4.2, the number of singular lines is $(s-1)/2$. Since there are $s-1$ singular points, the result follows. □

The above lemma gives rise to the following important necessary condition for g.

Define

$$\gamma(i) = c(i,i). \tag{6}$$

Corollary 4.2. $\gamma(\mathbb{Z}_s \backslash \{0\}) = SL$.

Proof. The singular point of l_i is given by $(\gamma(i), 2i)$. By Lemma 4.3 (2), $(\gamma(i), 2i)$ lies on a singular line $D_{m^2} + sx$ for some $x \in SL$. Hence, $\gamma(i) \in SL$. By Lemma 4.3 (1), for each $x \in SL$, $D_{m^2} + sx$ contains $(\gamma(i), 2i)$ and $(\gamma(-i), -2i)$ for some $i \in \mathbb{Z}_s \backslash \{0\}$. Thus, $x = \gamma(i) = \gamma(-i)$, i.e. γ is surjective. □

By Corollary 4.2, $\gamma(i) = 2^{-1}(2g(i) - g(2i) - d_0) \in SL$. Thus, g satisfies the following

Singular Condition. For $i = 1, 2, \cdots, s-1$,

$$g(2i) \in 2g(i) - d_0 - 2SL.$$

Similar to the case of the good points, the singular points also behave well under φ_p. By Lemma 4.1, we have the following

Lemma 4.4. $\gamma(pi) = \varphi_p(\gamma(i))$.

Proof. Since Φ_p takes S_{2i} to S_{2pi} and l_i to l_{pi}, it takes $l_i \cap S_{2i} = \tilde{i}$ to $l_{pi} \cap S_{2pi} = \widetilde{pi}$, i.e. $\Phi_p(\gamma(i), 2i) = (\gamma(pi), 2pi)$. Note that since $\Phi_p(D_{m^2} + sa) = D_{m^2} + s\varphi_p(a)$, it follows that $\Phi_p(\gamma(i), 2i) = (\varphi_p(\gamma(i)), 2pi)$. In particular, $\gamma(pi) = \varphi_p(\gamma(i))$. □

Thus, as in the case for the function g and GP, γ must also respect the p-cycle structure of $\mathbb{Z}_s\backslash\{0\}$ and the φ_p-cycle structure of SL.

The singular condition gives rise to no new condition when m is even.

Lemma 4.5. *For even m, all singular points lie on the extension of $2D_m - d_0$.*

Proof. Taking $p = 2$ in Corollary 4.1, we have $g(2i) = 2g(i) + d_0$. Hence, $\gamma(i) = 2^{-1}(g(i) + g(i) - g(2i) - d_0)$ is the constant $-d_0$, which represents the column containing the extension $2D_{m^2} - sd_0$ of the line $2D_m - d_0$. In particular, $\gamma(\mathbb{Z}_s\backslash\{0\}) = SL$. Geometrically, there is a basic leaf line which intersects each of the leaf lines at a singular point, which is not in S_0. Therefore, the restriction of the basic leaf line must not contain any good points, which is where a leaf line intersects S_0. By Lemma 4.2, the basic leaf line must be the extension of $2D_m - d_0$. □

It follows from Lemma 4.5 that when m is even, the singular condition becomes

$$g(2i) = 2g(i) - d_0 + 2d_0,$$

and is therefore already satisfied by g.

As an application, we prove a known result concerning extraneous multiplier, though it is unnecessary for the sequel. An integer k is called an *extraneous multiplier* of a cyclic difference set if k is a multiplier of the difference set and k does not divide the order of the difference set.

Lemma 4.6. *The integer 2 is never an extraneous multiplier of a planar cyclic difference set.*

Proof. Suppose 2 is an extraneous multiplier of a planar cyclic difference set D_m of order m. Then m is odd. Consider the set $2D_m - d_0$. By the

definition of a multiplier, $2D_m - d_0$ is a line of $\pi(D_m)$. On the other hand, the set of single points of $\pi(D_m)$ is $2D_m - d_0$. By Lemma 4.2 (1), this cannot be a line of $\pi(D_m)$. We have a contradiction. □

5. Sufficient Conditions

We are now prepared to construct a cyclic projective plane π' of order m^2 which contains a given cyclic projective plane π of order m as a subplane.

Let $\pi = \pi(D_m)$ be a given cyclic projective plane of order m, whose points are $\mathbb{Z}_t$, and the lines are translates of $D_m = \{d_0, d_1, \cdots, d_m\}$, such that $(0\ 1\ \cdots\ t-1)$ is a cyclic collineation. Assume $d_1 - d_0 \equiv 1 \pmod{t}$.

Let g be any function $g : \mathbb{Z}_s \longrightarrow \mathbb{Z}_t$, which satisfies Condition C, and $g(0) = d_0$. We construct a cyclic projective plane of order m^2 as follows. Let $\pi' = \{su, (u,i) \mid (0,i) \equiv i \pmod{s}, (u,i) = (0,i) + su, u = 0, 1, \cdots, t-1, i = 1, 2, \cdots, s-1\} = \mathbb{Z}_{st}$.

We arrange the points of π' in the following *small array*:

$(0, s-1)$	$(1, s-1)$	$\cdots$	$(t-1, s-1)$
$(0, s-2)$	$(1, s-2)$	$\cdots$	$(t-1, s-2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$(0, 2)$	$(1, 2)$	$\cdots$	$(t-1, 2)$
$(0, 1)$	$(1, 1)$	$\cdots$	$(t-1, 1)$
sd_0	$sd_0 + s$	$\cdots$	$sd_0 + (t-1)s$

Fig. 3 *The small array*

We expand the small array into an $(m^2+1) \times t$ array M of the form given by Fig. 2, without the condition that the first column is a cyclic difference set of order m^2.

Declare the lines of π' to be the columns of M, and for $i = 1, 2, \cdots, s-1$,

$$l_i = \{(0,i), (1,i), \cdots, (d_m - d_0, i), (c(i,j), i+j) \mid j = 1, 2, \cdots, s-1\},$$

and their translates, i.e. $\{(u,i), (1+u,i), \cdots, (d_m - d_0 + u, i), (c(i,j) + u, i+j) \mid j = 1, 2, \cdots, s-1\}$, for $u = 0, 1, \cdots, t-1$, and where $c(i,j) = 2^{-1}(g(i) + g(j) - g(i+j) - d_0)$.

By the foliation structure of cyclic projective plane of order m^2, we mean the structure given in Theorem 3.1. Let $\psi = (0\ 1\ 2\ \cdots st-1)$.

Theorem 5.1. *Let $\pi(D_m)$ be a cyclic projective plane of order m. If $g : \mathbb{Z}_s \longrightarrow \mathbb{Z}_t$ is a function satisfying $g(0) = d_0$ and Condition C, then π' defined above is a cyclic projective plane of order m^2 with foliation structure and cyclic collineation ψ.*

Proof. With Condition C, π' constructed above satisfies the axioms for a projective plane. Moreover, by the above construction, $S_0, S_1, \cdots, S_{s-1}$ are mutually isomorphic cyclic projective planes of order m. If we can show that $(0\ 1\ \cdots\ st-1)$ is a cyclic collineation of π', then we have a cyclic projective plane satisfying the requirement of the theorem.

First we prepare the following series of lemmas.

Note that by construction, the mapping α defined by going horizontally across M is a collineation of order t. If we can construct a collineation that commutes with α and has order s, then we can generate a cyclic collineation since $(s,t)=1$.

Our aim is to construct a collineation in the vertical direction with the required properties. By a collineation in the vertical direction, we mean a mapping applied to the small array. Note that by going in the vertical direction in the small array as it is, we do not get a collineation. Thus, we need to perform an operation on each of the rows, called sliding, which is defined as follows.

By a *slide* on the small array, we mean a map $L : \mathbb{Z}_s \longrightarrow \mathbb{Z}_t$ where $L(i)$ is defined to be the number of columns row i is shifted cyclically to the left, for $i \neq 0$. We set $L(0) = 0$.

A slide L is *proper* if it brings the singular point of l_i to the column containing its good point.

Lemma 5.1. *A slide L is proper if and only if $L(i) = 2^{-1}(d_0 - g(i))$.*

Proof. Since the singular point of the leaf line $l_{2^{-1}i}$ lies on row i, column $\gamma(2^{-1}i)$ in the small array, and its good point lies on column $g(2^{-1}i) - d_0$, we have $L(i) = 2^{-1}(g(2^{-1}i) + g(2^{-1}i) - g(i) - d_0) - (g(2^{-1}i) - d_0) = 2^{-1}(d_0 - g(i))$. □

Lemma 5.2. *If a slide L is proper, then $L(i+j) - L(i) - L(j) = c(i,j)$.*

Proof. This follows from (4) and Lemma 5.1. □

Let S be the small array and, by abuse of notation, $L(S)$ the small array with L applied. Let $\lambda : L(S) \longrightarrow L(S)$ be defined by $\lambda(u,i) = (u, i+1)$. Thus, the first column of $L(S)$ is given by $\{sd_0, (L(1),1), \cdots, (L(s-1), s-1)\}$.

Lemma 5.3. *λ is a collineation if and only if L is proper.*

Proof.

Suppose L is proper. By Lemma 5.1, row i is shifted cyclically to the left by $2^{-1}(d_0 - g(i))$ steps. Note that in the shifted small array, $l_i + sL(i)$ consists of the entries $(d_v - d_0, i)$, $(c(i,j) + L(i) - L(i+j), i+j)$, for $j \neq 0, v = 0, 1, \cdots, m$. On the other hand, $l_{i+1} + sL(i+1)$ consists of the entries $(d_v - d_0, i+1)$, $(c(i+1,j) + L(i+1) - L(i+j+1), i+j+1)$, for $j \neq 0, v = 0, 1, \cdots, m$; this is precisely the image of $l_i + sL(i)$ under λ since, by Lemma 5.2, $c(i,j)+L(i)-L(i+j) = -L(j) = c(i+1,j)+L(i+1)-L(i+1+j)$. Thus, λ is a collineation.

Suppose λ is a collineation on $L(S)$. Let $l_0 = D_{m^2}$. We have $\lambda(l_0) = l_1 + sL(1)$. It follows that for any $j \neq 0$, $\lambda(l_0 \cap S_j) = \lambda(l_0) \cap \lambda(S_j) = (l_1 + sL(1)) \cap S_{j+1}$. Since the point $l_0 \cap S_j$ must lie on the same column as its image under λ in $L(S)$,

$$-L(j) = c(1,j) + L(1) - L(j+1). \tag{7}$$

In particular, when $j = s-1$, we get $-L(s-1) = c(1,s-1) + L(1) - L(s-2)$, or $0 = L(0) = c(1,s-1)+L(1)+L(s-1)$. By repeated application of (7), we have

$$\begin{aligned} 0 &= c(1,s-1) + L(1) + c(1,s-2) + L(1) + L(s-2) \\ &= c(1,s-1) + L(1) + c(1,s-2) + L(1) + c(1,s-3) + L(1) + L(s-3) \\ &= \cdots \\ &= \sum_{j \neq 0} c(1,j) + sL(1). \end{aligned}$$

In other words, $-sL(1) = \sum_{j\neq 0} c(1,j)$. Hence,

$$\begin{aligned} -sL(1) &= 2^{-1}[\sum_{j\neq 0}(g(1) - d_0) + \sum_{j\neq 0}(g(j) - g(1+j))] \\ &= 2^{-1}[(s-1)(g(1) - d_0) + (g(1) - g(0))] \\ &= 2^{-1}[s(g(1) - d_0)], \end{aligned}$$

i.e., $L(1) = 2^{-1}(d_0 - g(1))$.

By (7),

$$\begin{aligned} L(2) &= c(1,1) + L(1) + L(1) \\ &= 2^{-1}(2g(1) - g(2) - d_0) + d_0 - g(1) \\ &= 2^{-1}(d_0 - g(2)). \end{aligned}$$

Repeating this argument, we obtain $L(i) = 2^{-1}(d_0 - g(i))$ for each i. Thus L is proper. □

We now complete the proof of Theorem 5.1. After application of the proper slide to the small array, we obtain in the vertical direction a collineation λ of order s which commutes with α. Let q be such that $qs \equiv 1 \pmod{t}$; thus, $(q,t) = 1$. As α has order t, therefore so does α^q. Furthermore, since α^q commutes with λ, the order of $\phi = \alpha^q \circ \lambda$ is st, i.e. ϕ is a cyclic collineation.

In fact, $\phi = \psi$. This we prove as follows. By the definition of ϕ, $\phi(x) \equiv x+1 \pmod{s}$, and so $\phi(x) \equiv x+1+sf(x) \pmod{st}$, for some $f : \mathbb{Z}_{st} \longrightarrow \mathbb{Z}_t$.

We claim that f is constant. First we note that if $x \equiv y \pmod{s}$, then $f(x) \equiv f(y) \pmod{t}$. Suppose, $x \equiv y + sr \pmod{st}$, for some r. Then, $x + 1 + sf(x) \equiv \phi(x) \equiv \alpha^r \phi(y) \equiv y + 1 + sf(y) + sr \equiv x + 1 + sf(y) \pmod{st}$. Since $(s,t) = 1$, $f(x) \equiv f(y) \pmod{t}$.

Since $\phi((0,1)) = (0,1) + 1 + sf(1)$ and $\phi((1,1)) = (1,1) + 1 + sf(1)$, it follows that $\phi(l_1) = l_1 + 1 + sf(1)$. In particular, for any $i \neq 1$, $\phi(l_1 \cap S_i) = (l_1 \cap S_i) + 1 + sf(1) = (c(1, i-1), i) + 1 + sf(1)$. On the other hand, $\phi(l_1 \cap S_i) = \phi((c(1, i-1), i)) = (c(1, i-1), i) + 1 + sf(i)$. Equating the two expressions for $\phi(l_1 \cap S_i)$, we conclude that $f(i) = f(1)$.

Thus, f is constant. It follows that $x+(1+fs)s \equiv \phi^s(x) \equiv (\alpha^q \circ \lambda)^s(x) \equiv \alpha(x) \equiv x+s \pmod{st}$. As $(s,t) = 1$, $f \equiv 0 \pmod{t}$, i.e. $\phi(x) \equiv x+1 \equiv \psi(x) \pmod{st}$.

Note that D_{m^2} is completely determined by D_m and g. Indeed, let $w \equiv 2^{-1} \pmod{t}$ and $E(g)$ be given by

$$E(g) = \{i - isq + sw(g(i) + d_0) \mid i = 1, 2, \cdots, s-1\}. \tag{8}$$

We have the following

Theorem 5.2. *The quadratic extension π' constructed in Theorem 5.1 is given by $\pi(D_{m^2})$, where $D_{m^2} = sD_m \cup E(g)$, and $E(g)$ is given by (8).*

Proof. As $D_{m^2} = sD_m \cup E$, it suffices to determine $E = \{(0,1), (0,2), \cdots, (0,s-1)\}$ in terms of g. Let $\phi = \alpha^q \circ \lambda$, where $sq \equiv 1 \pmod{t}$. Thus, ϕ is a collineation of π which moves a point one row upwards and q columns to the right. ϕ permutes the points as follows:

$$sd_0 \overset{\phi}{\to} (L(1)+q, 1) \overset{\phi}{\to} \cdots \overset{\phi}{\to} (L(i)+iq, i) \overset{\phi}{\to} \cdots \overset{\phi}{\to} (L(s-1)+(s-1)q, s-1) \overset{\phi}{\to} sd_0+s \overset{\phi}{\to} \cdots$$

Thus, $\phi^i(sd_0) = (L(i) + iq, i)$. Since ϕ is the collineation $(0\ 1\ 2\ \cdots\ s-1)$, the LHS $= sd_0 + i$. On the other hand, the RHS $= (0,i) + s(L(i) + iq)$. Therefore, $(0,i) = sd_0 + i - s(L(i) + iq)$. Now, L is proper if and only if $L(i) = 2^{-1}(g(i) + d_0)$. It follows that $(0,i) = i - isq + sw(g(i) + d_0)$, where $w \equiv 2^{-1} \pmod{t}$. □

Note that $sD_m \cup E(g)$, where $E(g)$ is given by (8), is the first column of M.

6. Equivalence Problem

We now turn our attention to the question of uniqueness. Recall that by Lemma 2.1, to prove that a cyclic projective plane of order m is unique, it suffices to show that cyclic difference sets of order m are all equivalent.

We consider cyclic projective planes of square order. We shall make use of the foliation structure of such planes (see Sections 3, 5). In terms of difference sets, we are therefore considering cyclic difference sets D_{m^2} of order m^2 of the form $sD_m \cup E(g)$, where $D_m = \{d_0, d_1, \cdots, d_m\}$ is a cyclic difference set of order m whose elements sum to 0, with $d_1 - d_0 \equiv 1 \pmod{t}$, and g is a function satisfying Condition C. We study conditions which imply their uniqueness, i.e. all cyclic difference sets of square order are equivalent.

Recall from Section 2 that $\mathcal{D}(m)$ is the set of all cyclic difference sets of order m each of whose elements sum to zero. Define an equivalence relation on $\mathcal{D}(m)$ as follows: $D_m \sim D'_m$ if and only if there exists an integer a such that $D'_m = aD_m$. It is readily verified that this is an equivalence relation.

Let $D_m \in \mathcal{D}(m)$. Write $D_{m^2}(g) = sD_m \cup E(g)$, where g satisfies Condition C, and call it the *quadratic extension* of D_m. Let

$$\mathcal{G}(D_m) = \{g \mid D_{m^2}(g) \text{ is a quadratic extension of } D_m\}.$$

Define an equivalence relation on $\mathcal{G}(D_m)$ as follows: $g \sim g'$ if and only if there exist a multiplier a of D_m, and an integer b, $(b, s) = 1$, such that $g'(i) = ag(bi) + (a-1)d_0, \forall i \in \mathbb{Z}_s \backslash \{0\}$, and $g'(0) = d'_0$. It is straightforward to check that the relation is an equivalence relation, if we can show that such a g' is in $\mathcal{G}(D_m)$. Indeed, the proof of the following lemma contains a more general result.

Lemma 6.1. *If D_m, D'_m are equivalent elements of $\mathcal{D}(m)$, then*

$$|\mathcal{G}(D'_m)| = |\mathcal{G}(D_m)|$$

and

$$|\mathcal{G}(D'_m)/\sim| = |\mathcal{G}(D_m)/\sim|.$$

Proof. Since D_m, D'_m are equivalent, there exists an integer a such that $D'_m = aD_m$. For any $g \in \mathcal{G}(D_m)$, Condition C is readily verified for $g' = a(g \circ b + d_0 - d_x)$ with respect to aD_m, where $ad_x = d'_0$, b an integer

satisfying $(b, s) = 1$, and $g'(0) = d'_0$. (Note that we use b to denote also the mapping defined by multiplication by b). Indeed, for any i, j, $c'(i, j) = ac(bi, bj) + a(d_0 - d_x)$. Thus,

$$\begin{aligned}
&\{d'_v - d'_0, g'(i) - d'_v, c'(i,j) \mid j \neq -i, v = 0, 1, \cdots m\} \\
&= \{ad_v - ad_0 + ad_0 - ad_x, ag(bi) - ad_v + ad_0 - ad_x, ac(bi, bj) + ad_0 - ad_x\} \\
&= a\{d_v - d_0, g(bi) - d_v, c(bi, bk)\} + ad_0 - ad_x.
\end{aligned}$$

It follows that Condition C is satisfied by g with respect to D_m if and only if it is satisfied by g' with respect to aD_m. Hence, $g' \in \mathcal{G}(aD_m) = \mathcal{G}(D'_m)$.

Let $\psi : \mathcal{G}(D_m) \longrightarrow \mathcal{G}(D'_m)$ be defined by $\psi(g) = g'$, where g' is given above. We have just shown that ψ is a bijection. Let $\Psi : \mathcal{G}(D_m)/\sim \longrightarrow \mathcal{G}(D'_m)/\sim$ be defined by $\Psi([g]) = [\psi(g)]$, where $[g]$ denotes the equivalence class of g. It is straightforward to check that Ψ is well defined and is a bijection. □

Lemma 6.2. (see also Bruck [1960]). *Let D_m be an element of $\mathcal{D}(m)$. Let $D_{m^2}(g), D_{m^2}(g')$ be quadratic extensions of D_m, respectively, with $g(0) = g'(0) = d_0$. Let a be a multiplier of D_m, and b be an integer prime to s. If $g'(i) = ag \circ b(i) + (a-1)d_0$, $\forall i \in \mathbb{Z}_s \backslash \{0\}$, then $D_{m^2}(g') = rD_{m^2}(g)$, where r satisfies the following system of equations:*

$$\begin{cases} r \equiv a \pmod{t} \\ rb \equiv 1 \pmod{s}. \end{cases}$$

Proof.

Note that by Lemma 6.1, if g and g' are related as in the hypothesis, then $g \in \mathcal{G}(D_m)$ if and only if $g' \in \mathcal{G}(D_m)$. Let r, b denote also multiplication by r, b, respectively. Then we have,

$$\begin{aligned}
rD_{m^2}(g) &= saD_m \cup rE(g) \\
&= sD_m \cup r\{bi - bisq + sw(g(bi) + d_0) \mid i = 1, 2, \cdots, s-1\} \\
&= sD_m \cup \{i - isq + sw(ag(bi) + (a-1)d_0 + d_0)\} \\
&= sD_m \cup E(ag \circ b) \\
&= D_{m^2}(g').
\end{aligned}$$

□

Lemma 6.3. *Let D_m, D'_m be equivalent elements of $\mathcal{D}(m)$, i.e. $D'_m = aD_m$, for an integer a. Let $D_{m^2}(g), D_{m^2}(g')$ be quadratic extensions of D_m, D'_m, respectively, with $g(0) = d_0$, and $g'(0) = d'_0$. If $g'(i) = a(g(i) + d_0 - d_x), \forall i \in$*

$\mathbb{Z}_s\backslash\{0\}$, *and* $ad_x = d'_0$, *then* $D'_{m^2}(g') = rD_{m^2}(g)$, *where* r *satisfies the following system of equations:*

$$\begin{cases} r \equiv a \pmod{t}, \\ r \equiv 1 \pmod{s}. \end{cases}$$

Proof.

Note that by Lemma 6.1, if g and g' are related as in the hypothesis, then $g \in \mathcal{G}(D_m)$ if and only if $g' \in \mathcal{G}(aD_m) = \mathcal{G}(D'_m)$. Let r, b denote also multiplication by r, b, respectively. Then we have,

$$\begin{aligned} rD_{m^2}(g) &= saD_m \cup rE(g) \\ &= s(D'_m) \cup r\{i - isq + sw(g(i) + d_0) \mid i = 1, 2, \cdots, s-1\} \\ &= (sD'_m) \cup \{i - isq + swa(g(i) + d_0)\} \\ &= (sD'_m) \cup \{i - isq + sw(g'(i) + d'_0)\} \\ &= D'_{m^2}(g'). \end{aligned}$$

□

We are now ready to state the main result of this section.

Theorem 6.1. *Suppose* $|\mathcal{D}(m)/\sim| = 1$, *and* $|\mathcal{G}(D_m)/\sim| = 1$ *for some* $D_m \in \mathcal{D}(m)$. *Then* $|\mathcal{D}(m^2)/\sim| = 1$.

Proof. Let $D_{m^2}, D'_{m^2} \in \mathcal{D}(m^2)$. By Theorem 3.1, there exist uniquely determined g, g', D_m, D'_m such that $D_{m^2} = D_{m^2}(g) = sD_m \cup E(g)$, and $D'_{m^2} = D'_{m^2}(g') = sD'_m \cup E(g')$.

Since $|\mathcal{D}(m)/\sim| = 1$, there exists an integer a such that $D'_m = aD_m$. Let $h(i) = a(g(i) + d_0 - d_x), i = 1, 2, \cdots, s-1$, $ad_x = d'_0$, and $h(0) = d'_0$. Consider the quadratic extension $D'_{m^2}(h)$ of D'_m. By Lemma 6.3, there exists an integer r_1 such that $D'_{m^2}(h) = r_1 D_{m^2}(g)$, where r_1 satisfies

$$\begin{cases} r_1 \equiv a \pmod{t}, \\ r_1 \equiv 1 \pmod{s}. \end{cases}$$

Since $D'_m \sim D_m$, Lemma 6.1 gives $|\mathcal{G}(D'_m)/\sim| = |\mathcal{G}(D_m)/\sim|$. Then by the hypothesis, $|\mathcal{G}(D'_m)/\sim| = |\mathcal{G}(D_m)/\sim| = 1$. Hence there exist integers a', b, where a' is a multiplier of D'_m, $(b, s) = 1$, such that $h = a'g' \circ b + (a' - 1)d'_0$. By Lemma 6.2, there exists an integer r_2 such that $D'_{m^2}(h) = r_2 D'_{m^2}(g')$, where r_2 satisfies

$$\begin{cases} r_2 \equiv a' \pmod{t}, \\ r_2 b \equiv 1 \pmod{s}. \end{cases}$$

Then $D'_{m^2}(g') = r_2^{-1} D'_{m^2}(h) = r_2^{-1} r_1 D_{m^2}(g)$. □

7. Search for Solutions

Given a cyclic difference set of order m, we are interested in finding all its quadratic extensions. More precisely, using notations introduced in Section 6, this amounts to determining $\mathcal{G}(D_m)$ and $\mathcal{G}(D_m)/\sim$ for a given $D_m \in \mathcal{D}(m)$. By Section 3, we must therefore search for all solutions $g : \mathbb{Z}_s \longrightarrow GP \cup \{d_0\}$ satisfying Condition C, where GP is the set of good points determined by D_m. In case all solutions are equivalent, i.e. $|\mathcal{G}(D_m)/\sim| = 1$, then the given cyclic difference set admits, up to equivalence, a unique quadratic extension.

On the other hand, for a given order m, we can study the equivalence problem for cyclic difference sets of that order. We are interested in those orders where there is a unique cyclic difference set, i.e. $|\mathcal{D}(m)| = 1$. If, for such orders, there is only a unique solution $g \in \mathcal{G}(D_m)$, i.e. $|\mathcal{G}(D_m)/\sim| = 1$, then by Theorem 6.1, we can conclude that cyclic difference sets of order m^2, and hence cyclic projective planes of those orders, are unique.

Our implementation of the search strategy is as follows: Let D_m be a cyclic difference set of order m where $m = p^\alpha$, where p is a prime. We shall perform a depth-first-search for $g \in \mathcal{G}(D_m)$ when $m \leq 9$, and a breadth-first-search when $m = 11, 16$. In view of Lemma 4.3 and Lemma 4.5, we distinguish between the case when m is odd and when m is even.

(1) For odd m:
By Corollary 4.1 and 4.4, we see that a basic criterion is given by the orbit size. Indeed, consider the p-orbits of $\mathbb{Z}_s\backslash\{0\}$. Since $p^{3\alpha} = m^3 \equiv -1 \pmod{s}$, it follows that i lies in the same p-orbit of $-i$. Since $g(j) = g(i)$ if and only if $j = \pm i$, it follows that the orbit size of $g(i)$ under φ_p must be half of the orbit size of i under p. Moreover, distinct p-orbits are mapped under g to distinct orbits of GP, since g is a two-to-one function.
Next, we consider the orbits of SL. By Lemma 4.3, a singular line contains dual singular points. Therefore, $\gamma(i) = \gamma(j)$ if and only if $i = \pm j$. Thus, distinct p-orbits are mapped to distinct φ_p-orbits of SL under γ.
Now, given any $i \in \mathbb{Z}_s\backslash\{0\}$, we choose an φ_p-orbit $(r) \in GP$ whose size is half that of the orbit (i) under p.
We may assume $g(i) = r$. Indeed, by Corollary 4.1, there exists an integer a, such that $g(p^a i) = r$. From the discussion in the last Section,

we know that $g \circ p^a$ is equivalent to g as elements of $\mathcal{G}(D_m)$. Thus we may replace g by $g \circ p^a$ as they give rise to equivalent quadratic extensions. By Corollary 4.1, the rest of the orbit of i under p is then determined.

Compute $c(j,k)$ whenever possible; in this case, for $j, k, j+k \in (i)$. By Condition C, $c(j,k)$ must not lie in $(D_m - d_0) \cup (g(j) - D_m) \cup (g(k) - D_m)$. Moreover, the singular condition implies that $\gamma(j)$ must lie in SL. If either of these fails, then $g(i) = r$ will not lead to a solution, and we have to choose r from another orbit of GP.

Suppose $g(i) = r$ poses no obstruction at this stage. Now, since $g(2i) \in GP$, and by the singular condition, it follows that $g(2i) \in (2r - d_0 - 2SL) \cap GP$. If $(2r - d_0 - 2SL) \cap GP$ is empty, then $g(i) = r$ does not lead to a solution, and we have to return to the beginning. Otherwise, choose $r' \in (2r - d_0 - 2SL) \cap GP$ such that r' lies in an orbit of size double that of $(2i)$. Repeat the procedure, this time for $g(2i) = r'$.

Having defined g on the orbits $(i), (2i), \cdots, (2^{x-1}i)$, where $(2^x i) = (i)$, we proceed to define g on a distinct orbit (j) by repeating the above procedure. Whenever g fails a condition, return to the last decision step and choose another value.

After g is completely determined, we use Condition C to test the validity of the solution. We check whether for each i, $\{d_v - d_0, g(i) - d_v, c(i,j) \mid j \neq -i, v = 0, \cdots, m\} = \mathbb{Z}_t$. Return to the last decision step and resume the search regardless of the result of the test of Condition C. Continuing this way backwards until we have reached the first decision, i.e. the assignment of $g(i)$, we get all candidates passing the test of Condition C. Using these candidates, we construct quadratic extensions following our established procedure.

(2) For even m, the strategy is basically the same, the only difference being that the singular condition does not give rise to any further condition.

The search above determines $\mathcal{G}(D_m)$. Next, we compute $\mathcal{G}(D_m)/\sim$. Given $g \in \mathcal{G}(D_m)$, we determine its equivalence class in $\mathcal{G}(D_m)/\sim$, by computing $ag \circ b + (a-1)d_0$ as a varies over all multipliers of D_m and b varies over all integers prime to s. If we can construct as many equivalent g's in this way as there are elements in $\mathcal{G}(D_m)$, then we have $|\mathcal{G}(D_m)/\sim| = 1$.

The remaining step is to determine the number of equivalence classes of a cyclic difference set of order m. Without loss of generality, we may restrict our attention to $\mathcal{D}(m)$.

By Lemma 2.3, the problem of finding all cyclic difference set of order m is reduced to finding all correct combinations of p-orbits of $\mathbb{Z}_t$ such that

they satisfy the definition. Note that the combinatorial complexity using this method increases very quickly, and thus we cannot apply this method to prove the uniqueness of cyclic difference sets of high square orders.

To illustrate our methodology, we present the details for the case where $m = 5$.

Here, $t = 31, s = 21, q = 3, w = 16$.

We first look at the equivalence of cyclic difference sets D_5 of order 5. We determine $\mathcal{D}(5)$ using Lemma 2.3. Since 5 is a multiplier, we decompose $\mathbb{Z}_{31}$ into 5-cycles. They are (0) and

$$\begin{array}{llllll} (1\ \ 5\ \ 25) & (2\ 10\ 19) & (4\ \ 20\ \ 7) & (8\ \ \ 9\ \ 14) & (16\ 18\ \ 28) \\ (3\ 15\ 13) & (6\ 30\ 26) & (12\ 29\ 21) & (24\ 27\ 11) & (17\ 23\ 22). \end{array}$$

Now for any $D_5 \in \mathcal{D}(5)$, $|D_5| = 6$. Therefore, D_5 must consist of two out of these ten cycles. Note that by multiplying 6 successively, any cycle can be brought to (1 5 25), which may be assumed to be in D_5. Then we must verify which of the 9 combinations gives a genuine D_5.

Computations show that only two of them form cyclic difference sets of order 5. All the other elements in $\mathcal{D}(5)$ can be obtained by multiplying 2 successively to these two. They are

$$\begin{array}{ll} \{1\ \ 5\ \ 25\ 24\ 27\ 11\} & \{1\ \ 5\ \ 25\ 17\ 23\ 22\} \\ \{2\ \ 10\ 19\ 17\ 23\ 22\} & \{2\ \ 10\ 19\ \ 3\ \ 15\ 13\} \\ \{4\ \ 20\ \ 7\ \ \ 3\ \ 15\ 13\} & \{4\ \ 20\ \ 7\ \ \ 6\ \ 30\ 26\} \\ \{8\ \ \ 9\ \ 14\ \ 6\ \ 30\ 26\} & \{8\ \ \ 9\ \ 14\ 12\ 29\ 21\} \\ \{16\ 18\ 28\ 12\ 29\ 21\} & \{16\ 18\ 28\ 24\ 27\ 11\} \end{array}$$

Note that these 10 cyclic difference sets can be obtained by multiplying successively 17 to $D_5 = \{1, 5, 25, 24, 27, 11\}$. So $|\mathcal{D}(5)/\sim| = 1$.

Next we consider the quadratic extension of D_5. Note that $d_0 = 24$, and by (5), $\varphi_5 : x \mapsto 5x + 3$. The 5-orbits of $\mathbb{Z}_{21}\backslash\{0\}$ and the φ_5-orbits of GP and SL are respectively,

$$\mathbb{Z}_{21}\backslash\{0\} = \begin{cases} (1\ \ \ 5\ \ \ 4\ \ 20\ 16\ 17) \\ (2\ \ 10\ \ \ 8\ \ 19\ 11\ 13) \\ (3\ \ 15\ \ 12\ 18\ \ \ 6\ \ \ 9) \\ (7\ 14) \end{cases},$$

$$GP = \begin{cases} (0\ \ \ 3\ \ 18) \\ (10\ 22\ 20) \\ (21\ 15\ 16) \\ (7) \end{cases}, SL = \begin{cases} (9\ \ 17\ 26) \\ (14\ 11\ 27) \\ (20\ 10\ 22) \\ (7) \end{cases}.$$

First of all, note that the size of the 5-orbit (7) is 2, and there is only one φ_5-orbit of size 1. Therefore, $g(7) = g(14) = 7$.

Next, consider the 5-orbit (3). This orbit is particularly convenient as we can compute all $c(3i, 3j)$, where i, j run from 1 to 6, once $g(3)$ is defined. (Note that this follows from the fact that $s = 21$ is a multiple of 3, and all multiples of 3 lie in (3)). Suppose $g(3) \in (a)$. We may assume $g(3) = a$.

Suppose $a = 0$, i.e. $g(3) = 0$. By Corollary 4.1, $g(5\times 3) = g(15) = g(6) = 5\times 0+3 = 3$. Therefore, $\gamma(3) = 2^{-1}(2g(3)-g(6)-24) = 16\times(2\times 0-3-24) = 2 \notin SL$. This shows that $g(3) \notin (0)$, by the singular condition. Suppose $a = 10$, i.e. $g(3) = 10$. Then $g(6) = 22$, and $\gamma(3) = 18 \notin SL$. Therefore, $g(3) \notin (10)$. Suppose $g(3) = 21$. Then $g(6) = 15$, and $\gamma(3) = 17 \in SL$. Thus, $g(3)$ must lie in (21).

Next, we proceed to determine g for the orbit (2×3). But $(2 \times 3) = (3)$, and this has already been done. Therefore, we move on to a different orbit, say (1). Suppose $g(1) = 0$. Then, $g(2) \in (2g(1) - d_0 - 2SL) \cap GP = \{10, 15, 16, 18, 20\}$. Since $g(3) = 21$ and $15, 16 \in (21)$, $g(2)$ can only be $20, 18$ or 10. Suppose $g(2) = 20$. Then, $\gamma(1) = 9$. This means the singular point of l_1 lies on column 9. But by Corollary 4.4, $\gamma(12) = 9$. This contradicts the fact that every singular line contains only two dual singular points. Hence $g(2)$ cannot be 20. Note that $g(1) \in (18)$, therefore $g(2) \neq 18$. Thus, $g(2)$ can only be 10. We then have a candidate g_1:

i	1	2	3	4	5	6	7	8	9	10
$g_1(i)$	0	10	21	18	3	15	7	20	16	22
i	11	12	13	14	15	16	17	18	19	20
$g_1(i)$	22	16	20	7	15	3	18	21	10	0

Next, we return to the last decision step, that is, the assignment of $g(2)$. But it has been shown that there is no other choice at this stage. Hence, we return one more step backward to the assignment of $g(1)$. It is straightforward to show that we get only one more candidate, g_2:

i	1	2	3	4	5	6	7	8	9	10
$g_2(i)$	20	3	21	22	10	15	7	0	16	18
i	11	12	13	14	15	16	17	18	19	20
$g_2(i)$	18	16	0	7	15	10	22	21	3	20

We check Condition C for these two candidates. Using g_1, we get a genuine cyclic difference set which is given in Section 7. Similarly, we can show that g_2 gives rise to a genuine $D_{25}(g_2)$. Since $g \circ 5^3 = g \circ (-1) = g$, we can construct two more solutions equivalent to g once a solution g is defined, namely, $g \circ 5$ and $g \circ 5^2$. Thus, we have altogether six elements, namely, $g_1, g_1 \circ 5, g_1 \circ 5^2, g_2, g_2 \circ 5, g_2 \circ 5^2$, in $\mathcal{G}(D_5)$. Upon closer inspection, we see that $g_2 = g_1 \circ 8$. Thus, all these 6 solutions are equivalent to each other, i.e. $|\mathcal{G}(D_5)/\sim| = 1$.

We have now shown that $|\mathcal{D}(5)/\sim| = 1$, and $|\mathcal{G}(D_5)/\sim| = 1$, for $D_5 = \{24, 25, 27, 1, 5, 11\}$. According to Theorem 6.1, D_{25} is unique.

We now discuss the equivalence of all the quadratic extensions of the cyclic difference sets of order 5. We have $\mathcal{D}(5) = \{D_5^1, D_5^2, \cdots, D_5^{10}\}$, with $D_5^{i+1} = 17D_5^i$, where $D_5 = \{24, 25, 27, 1, 5, 11\}$. Let $D_{25}(i,j) = 21D_5^i \cup E(g_j^i)$, $i = 1, \cdots, 10, j = 1, \cdots, 6$, where $g_1^1 = g_1$ as given in the above discussion. The rest of the g_j^i's are determined as follows. Since $5 \times 8 \equiv 19 \pmod{21}$ and $(2,3) = 1$, $\mathcal{G}(D_5^1) = \{g_1 \circ 5^k \circ 8^h \mid k = 0,1,2, h = 0,1\} = \{g_1 \circ 19^k \mid k = 0,1,2,3,4,5,\}$. Together with Lemma 6.1, this gives the following system, for $i = 1, \cdots, 10, j = 1, \cdots, 6$:

$$\begin{cases} g_{j+1}^i &= g_j^i \circ 19, \\ g_j^{i+1} &= 17(g_j^i + d_0^i) - d_0^{i+1}. \end{cases}$$

Hence, by Lemma 6.2, $D_{25}(i, j+1) = 94 \times D_{25}(i,j)$, and by Lemma 6.3, $D_{25}(i+1, 1) = 358 \times D_{25}(i,1)$.

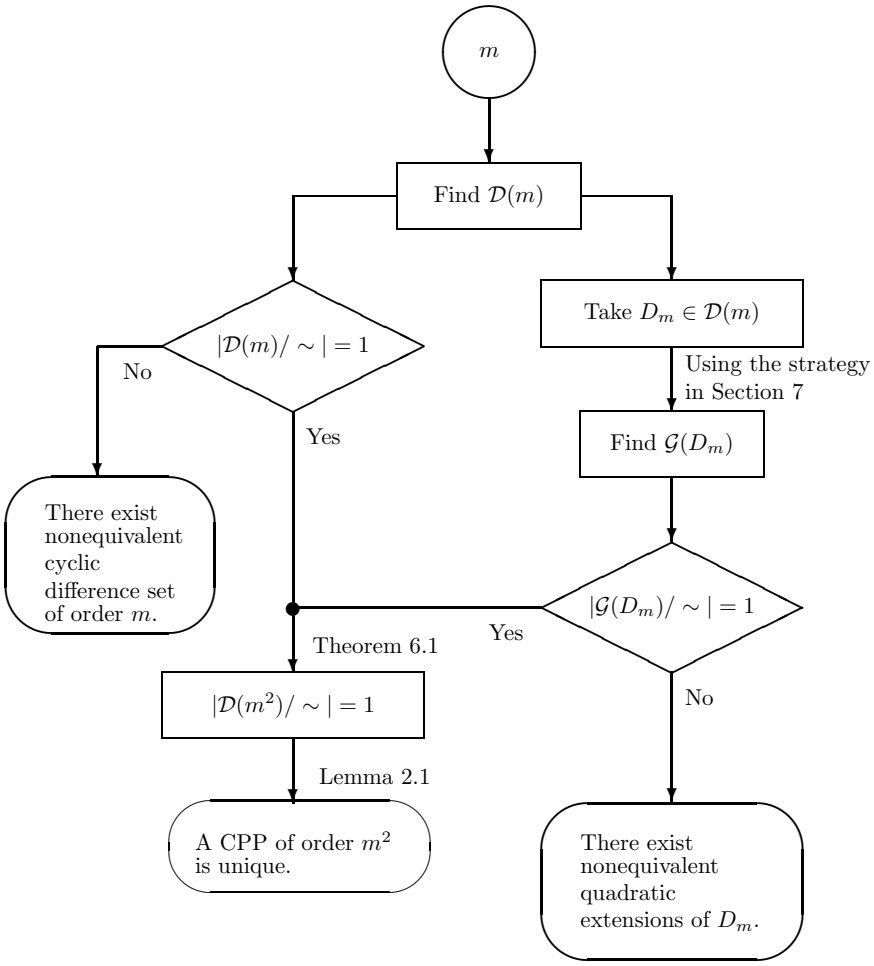
Thus, $\mathcal{D}(25) = \{94^p \times 358^h D_{25}(1,1) \mid p = 0,1,\cdots,5; h = 0,1,\cdots,9\}$.

8. Conclusion and Data

Recall the following notation:

- m: an integer > 1
- $t = m^2 + m + 1$
- $s = m^2 - m + 1$
- $D_m \subset \mathbb{Z}_t$: a planar cyclic difference set of order m
- $D_{m^2} \subset \mathbb{Z}_{st}$: a planar cyclic difference set of order m^2
- $\mathcal{D}(m) = \{D_m \mid \text{elements in } D_m \text{ sum to zero } (\text{mod } t)\}$.

We summarize our methodology in the following flow chart.



The following are the mains results of this article.

Theorem 8.1. *A cyclic difference set of order m or m^2 is unique, where $m = 2, 3, 4, 5, 7, 8, 9, 11, 16$.*

Theorem 8.2. *A cyclic projective plane of order m or m^2 is Desarguesian, where $m = 2, 3, 4, 5, 7, 8, 9, 11, 16$.*

We now present the sets $\mathcal{D}(m)$ and $\mathcal{D}(m^2)$, for the above values of m. In particular, this corrects an error in Bruck's [1960] data for a cyclic difference set of order 49. Note that elements of each D_{m^2} are arranged according to their residues modulo s. In the set description of $\mathcal{D}(m^2)$, numbers whose

powers are p or q define equivalence between quadratic extensions of the same $D_m \in \mathcal{D}(m)$, while numbers whose powers are h or k define equivalence of quadratic extensions of different D_m's $\in \mathcal{D}(m)$. All indices run in integers.

8.1. $m = 2$

$s = 3,\ t = 7,\ st = 21.$

$$D_2 = \{\ 1\ 2\ 4\ \}.$$

$\mathcal{D}(2) = \{3^h D_2 \mid 0 \le h \le 1\}.$

$$D_4 = \{\ 3\ 6\ 12\ 7\ 14\ \}.$$

$\mathcal{D}(4) = \{10^h D_4 \mid 0 \le h \le 1\}.$

8.2. $m = 3$

$s = 7,\ t = 13,\ st = 91.$

$$D_3 = \{\ 0\ 1\ 3\ 9\ \}.$$

$\mathcal{D}(3) = \{2^h D_3 \mid 0 \le h \le 3\}.$

$$D_9 = \{\ 0\ 7\ 21\ 63\ 71\ 2\ 31\ 18\ 54\ 6\ \}.$$

$\mathcal{D}(9) = \{40^p \times 15^h D_9 \mid 0 \le p \le 2;\ 0 \le h \le 3\}.$

8.3. $m = 4$

$s = 13,\ t = 21,\ st = 273.$

$$D_4 = \{\ 6\ 7\ 12\ 14\ 3\ \}.$$

$\mathcal{D}(4) = \{10^h D_4 \mid 0 \le h \le 1\}.$

$$D_{16} = \{\ 78\ 91\ 156\ 182\ 39\ 157\ 41\ 55\ 82\ 122\ 110\ 215\ 164\ 61\ 244\ 167\ 220\}.$$

$\mathcal{D}(16) = \{85^p \times 157^h D_{16} \mid 0 \le p \le 5;\ 0 \le h \le 1\}.$

8.4. $m = 5$

$s = 21,\ t = 31,\ st = 651.$

$$D_5 = \{\ 24\ 25\ 27\ 1\ 5\ 11\ \}.$$

$\mathcal{D}(5) = \{17^h D_3 \mid 0 \le h \le 9\}.$

$$D_{25} = \{\ 504\ 525\ 567\ 21\ 105\ 231\ 190\ 233\ 612\ 193\ 299\ 363\ 217\ 617\ 513\ 514\ 452\ 327\ 307\ 434\ 456\ 268\ 38\ 333\ 481\ 314\}.$$

$\mathcal{D}(25) = \{94^p \times 358^h D_{25} \mid 0 \le p \le 5;\ 0 \le h \le 9\}.$

8.5. $m = 7$

$s = 43,\ t = 57,\ st = 2451.$

$$D_7 = \{\ 38\ 39\ 45\ 5\ 17\ 19\ 30\ 35\ \ \}.$$

$\mathcal{D}(7) = \{2^h \times 10^k D_7 \mid 0 \le h \le 5;\ 0 \le k \le 1\}.$

$$\begin{aligned} D_{49} = \{\ & 1634\ 1677\ 1935\ 215\ 731\ 817\ 1290\ 1505\ 1893\ 561 \\ & 1981\ 4\ 134\ 2070\ 996\ 653\ 1514\ 612\ 656\ 528 \\ & 2120\ 1476\ 2251\ 1263\ 576\ 1480\ 1051\ 794\ 1612\ 1441 \\ & 281\ 196\ 283\ 1488\ 1833\ 28\ 1362\ 1664\ 2181\ 1967 \\ & 1581\ 2141\ 938\ 939\ 1671\ 1844\ 1372\ 556\ 1245\ 2235\}. \end{aligned}$$

$\mathcal{D}(49) = \{1825^p \times 173^h \times 1549^k D_{49} \mid 0 \le p \le 20;\ 0 \le h \le 5;\ 0 \le k \le 1\}.$

8.6. $m = 8$

$s = 57,\ t = 73,\ st = 4161.$

$$D_8 = \{\ 1\ 2\ 4\ 8\ 16\ 32\ 64\ 55\ 37\}.$$

$\mathcal{D}(8) = \{5^h D_8 \mid 0 \le h \le 7\}.$

$$\begin{aligned} D_{64} = \{\ & 57\ 114\ 228\ 456\ 912\ 1824\ 3648\ 3135\ 2109\ 742 \\ & 1484\ 1941\ 2968\ 2456\ 3882\ 1717\ 1775\ 1491\ 751\ 2234 \\ & 3603\ 2293\ 3434\ 3606\ 3550\ 2639\ 2982\ 1387\ 1502\ 3555 \\ & 307\ 1847\ 3045\ 2647\ 425\ 1737\ 2707\ 371\ 3051\ 1228 \\ & 2939\ 2826\ 1117\ 3227\ 1803\ 3400\ 2774\ 3858\ 3004\ 3404 \\ & 2949\ 2266\ 614\ 1413\ 3694\ 1700\ 1929\ 1702\ 1133\ 2787 \\ & 850\ 851\ 3474\ 2506\ 1253\}. \end{aligned}$$

$\mathcal{D}(64) = \{3943^p \times 1027^h D_{64} \mid 0 \le p \le 17;\ 0 \le h \le 7\}.$

8.7. $m = 9$

$s = 73,\ t = 91,\ st = 6643.$

$$D_9 = \{\ 6\ 7\ 18\ 21\ 31\ 54\ 63\ 71\ 0\ 2\ \ \}.$$

$\mathcal{D}(9) = \{2^h D_9;\ h = 0, 1, \cdots, 11\}.$

$$\begin{aligned} D_{81} = \{\ & 438\ 511\ 1314\ 1533\ 2263\ 3942\ 4599\ 5183\ 0 \\ & 146\ 1\ 2411\ 3\ 4749\ 6575\ 590\ 1321\ 81 \\ & 9\ 4609\ 3442\ 961\ 1035\ 4467\ 6439\ 2644\ 1623 \\ & 1770\ 4764\ 5714\ 3963\ 2139\ 4768\ 243\ 3018\ 4406 \\ & 27\ 5576\ 1489\ 541\ 4119\ 6018\ 3683\ 4925\ 2006 \\ & 2883\ 2519\ 914\ 3105\ 1135\ 2742\ 115\ 2452\ 2672 \\ & 6031\ 6397\ 3405\ 1289\ 4429\ 1583\ 4869\ 5965\ 345 \\ & 5310\ 1588\ 713\ 1006\ 4073\ 1373\ 3856\ 3054\ 4807 \\ & 5246\ 6561\ 5905\ 6417\ 4958\ 3572\ 1018\ 2187\ 3867 \\ & 729\}. \end{aligned}$$

$\mathcal{D}(81) = \{5825^p \times 366^h D_{81} \mid 0 \le p \le 35;\ 0 \le h \le 11\}.$

8.8. $m = 11$

$s = 111,\ t = 133,\ st = 14763.$

$$D_{11} = \{71\ 72\ 79\ 85\ 101\ 116\ 118\ 127\ 4\ 44\ 47\ 67\ \ \}.$$

$\mathcal{D}(11) = \{2^h \times 45^k D_{11} \mid 0 \le h \le 17;\ 0 \le k \le 1\}.$

$$\begin{array}{rccccccccc}
D_{121} = \{ & 7881 & 7992 & 8769 & 9435 & 11211 & 12876 & 13098 & 14097 & 444 \\
& 4884 & 5217 & 7437 & 11212 & 5552 & 2112 & 1891 & 6554 & 8775 \\
& 12661 & 4781 & 1119 & 13219 & 5228 & 11112 & 4786 & 10448 & 7230 \\
& 460 & 9341 & 8343 & 3349 & 7457 & 4128 & 2020 & 7238 & 801 \\
& 14344 & 14567 & 5355 & 8353 & 9353 & 4581 & 5803 & 8357 & 8469 \\
& 5473 & 11912 & 11025 & 4921 & 5810 & 3813 & 7366 & 1151 & 8811 \\
& 11587 & 6038 & 7149 & 2932 & 9149 & 13146 & 11371 & 10595 & 4824 \\
& 12928 & 10154 & 5715 & 13042 & 12377 & 3720 & 6829 & 8273 & 13602 \\
& 3280 & 2726 & 3171 & 12607 & 5060 & 7947 & 5506 & 9725 & 5619 \\
& 11392 & 1514 & 11394 & 12061 & 9842 & 14616 & 14173 & 6404 & 8070 \\
& 6628 & 2744 & 192 & 3634 & 1304 & 11739 & 4858 & 3305 & 3195 \\
& 8302 & 1754 & 2532 & 4531 & 13856 & 2757 & 2425 & 6977 & 12417 \\
& 14305 & 7313 & 12309 & 5095 & 11756 & 13089 & 658 & 7208 & 1992 \\
& 13204 & 7211 & 6102 & 8212 & 12542\}. & & & &
\end{array}$$

$\mathcal{D}(121) = \{7715^p \times 400^q \times 667^h \times 14542^k D_{121} \mid 0 \le p \le 5;\ 0 \le q \le 5;\ 0 \le h \le 17;\ 0 \le k \le 1\}.$

8.9. $m = 16$

$s = 241, t = 273, st = 65793.$

$$\begin{array}{rl}
D_{16} = \{ & 78\ \ 91\ 156\ 182\ \ \ 39\ \ \ 157\ 41\ 55\ 82\ 122\ 110\ 215 \\
& 164\ 61\ 244\ 167\ 220\}.
\end{array}$$

$\mathcal{D}(16) = \{85^p \times 157^h D_{16};\ 0 \le p \le 5;\ 0 \le h \le 1\}$.

$$
\begin{aligned}
D_{256} = \{\ & 9158\ 9399\ 59768\ 4579\ 7471\ 14942\ 17593\ 18316\ 18798 \\
& 21931\ 29884\ 35186\ 36632\ 37596\ 41693\ 43862\ 53743\ 41212 \\
& 16631\ 65555\ 33262\ 246\ 65317\ 8924\ 731\ 60018\ 492 \\
& 15194\ 64841\ 8207\ 17848\ 23392\ 1462\ 18574\ 54243\ 39302 \\
& 984\ 50149\ 30388\ 35209\ 63889\ 15690\ 16414\ 6534\ 35696 \\
& 3885\ 46784\ 34012\ 2924\ 61970\ 37148\ 26545\ 42693\ 37151 \\
& 12811\ 14981\ 1968\ 23418\ 34505\ 18841\ 60776\ 4865\ 4625 \\
& 12579\ 61985\ 47285\ 31380\ 14270\ 32828\ 30660\ 13068\ 34036 \\
& 5599\ 48257\ 7770\ 54525\ 27775\ 32837\ 2231\ 36695\ 5848 \\
& 42722\ 58147\ 34530\ 8503\ 55981\ 53090\ 54055\ 19593\ 36464 \\
& 8509\ 62976\ 25622\ 41529\ 29962\ 15744\ 3936\ 5142\ 46836 \\
& 20568\ 3217\ 40814\ 37682\ 55758\ 55759\ 15754\ 9730\ 16479 \\
& 9250\ 30941\ 25158\ 12868\ 58177\ 36729\ 28777\ 31670\ 62760 \\
& 21550\ 28540\ 19142\ 65656\ 47582\ 61320\ 25653\ 26136\ 64697 \\
& 2279\ 25657\ 11198\ 9512\ 30721\ 63016\ 15540\ 51691\ 43257 \\
& 38920\ 55550\ 20606\ 65674\ 123\ 4462\ 30009\ 7597\ 37000 \\
& 11696\ 9287\ 19651\ 57971\ 50501\ 7845\ 3267\ 34839\ 17006 \\
& 30985\ 46169\ 51472\ 40387\ 11709\ 42317\ 35329\ 39186\ 56539 \\
& 7135\ 15330\ 17018\ 57025\ 60159\ 49315\ 51244\ 21361\ 17265 \\
& 60887\ 59924\ 18232\ 31488\ 53661\ 7872\ 2571\ 10284\ 20407 \\
& 27879\ 7877\ 41136\ 48367\ 6434\ 51261\ 15835\ 10775\ 9571 \\
& 23791\ 45723\ 65245\ 45725\ 4756\ 31508\ 58742\ 19460\ 10303 \\
& 32958\ 47901\ 18500\ 37540\ 61882\ 36819\ 50316\ 48389\ 25736 \\
& 38751\ 50561\ 61166\ 7665\ 61409\ 57554\ 43577\ 63340\ 9116 \\
& 59727\ 34182\ 43100\ 36835\ 57080\ 58527\ 38284\ 44792\ 65519 \\
& 2378\ 29371\ 38048\ 56847\ 18770\ 51306\ 57091\ 52272\ 30583 \\
& 63601\ 54685\ 4558\ 17091\ 51314\ 62160\ 22396\ 1189\ 19024 \\
& 9385\ 61442\ 48188\ 60239\ 41442\ 31080\ 33491\ 37589\ 24094 \\
& 20721\ 49642\ 12047\ 24821\ 45307\}.
\end{aligned}
$$

$\mathcal{D}(256) = \{58150^p \times 40489^h \times 56395^k D_{256} \mid 0 \le p \le 119;\ 0 \le h \le 5;\ 0 \le k \le 1\}$.

References

1. Baumert, L. D. (1971). *Cyclic Difference Sets*, Lectures Notes in Mathematics **182**, Springer-Verlag, Berlin.
2. Bruck, R. H. (1960). Quadratic extensions of cyclic planes, *Combinatorial Analysis, Proceedings of Symposia in Applied Mathematics* **X**, 15-44. American Mathematical Society, Providence, Rhode Island.
3. Hall, M. Jr. (1959). *The Theory of Groups*, The Macmillan Company, New York.
4. Ho, C. Y. (1998). Finite projective planes with abelian transitive collineation groups, *J. Algebra* **208**, 533-550.

5. Hughes, D. R., Piper, F. (1973). *Projective Planes*, Springer-Verlag, New York.
6. Ostrom, T. G., Wagner, A. (1959). On projective and affine planes with transitive collineation groups, *Math. Z.* **71**, 186–199.
7. Ott, U. (1975). Endliche zyklische Ebenen, *Math. Z.* **144**, 195-215.
8. Singer, J. (1938). A theorem in finite projective geometry and some applications to number theory, *Trans. Amer. Math. Soc.* **43**, 377-385.

Some integral representations of finite groups and their arithmetic applications

Dmitry A.Malinin

Middle East Technical University,
Northern Cyprus Campus,
Kalkanli, Güzelyurt, KKTC
Mersin 10, Turkey,
E-mail: dmalinin@gmail.com

For a given global field F we consider a finite normal extension E/F with the maximal order $\mathcal{O}_E$ and finite subgroups $G \subset GL_n(E)$. We assume that G is stable under the natural action of the Galois group of E/F and consider fields $E = F(G)$ that are obtained via adjoining all matrix coefficients of all matrices $g \in G$ to F. We study the conditions of Galois stability for finite subgroups $G \subset GL_n(\mathcal{O}_E)$. Some results for the possible realization fields $E = F(G)$ of characteristic $p > 0$ and 0 are given. Using some explicit estimates for integral polynomials it is proved that contrary to realizability of fields $E = F(G)$ in characteristic $p > 0$ the condition $F \neq F(G)$ is a rarity in characteristic 0.

Keywords: integral representations, finite groups, global fields, positive characteristic, Galois group, algebraic integers, large sieve method

1. Introduction

In this paper we study the conditions of Galois stability for finite subgroups of $GL_n(E)$ with entries in a global field E.

Let E/F be a Galois extension of finite degree of global fields, i.e. E, F are finite extensions of the field of rationals $\mathbb{Q}$ or a field of rational functions $R(x)$ with a finite field R, and let Γ be the Galois group of E/F for a finite subgroup $G \subset GL_n(E)$, where we assume that G is stable under the natural coefficientwise Γ-action.

Throughout this paper $\mathcal{O}_E$ is the maximal order of E, and $F(G)$ denotes a field that is obtained via adjoining to F all matrix coefficients of all matrices $g \in G$.

The main objective of this paper is to prove the existence of Γ-stable subgroups G such that $F(G) = E$ and to study the interplay between the

existence of Γ-stable groups G over global fields and over their rings of integers. Using some explicit estimates for integral polynomials it is proved that contrary to realizability of fields $E = F(G)$ in characteristic $p > 0$ the condition $F \neq F(G)$ is a rarity in characteristic 0 for $F = \mathbb{Q}$ or its quadratic extensions (see Theorem 2.3 below).

The results related to the Galois stability of finite groups in the situation similar to ours arise in the theory of definite quadratic forms and Galois cohomologies of certain arithmetic groups if F is an algebraic number field and G is realized over its maximal order ([3], see also [20]). In our context we study whether a given field E normal over F can be realized as a field $E = F(G)$, and if this is so what are the possible fields of realization and the structure of G. Some similar questions for Γ-stable orders in simple algebras are considered in [21], see also [22] for some applications. Some special Galois operation on finite groups (mainly on quaternion group) was considered in [5], and also some applications to the classification of curves of genus 2 with automorphism group isomorphic to $\tilde{S}_4$.

In an earlier paper [2] we studied finite groups $G \subset GL_n(\mathcal{O}_E)$ stable under the operation of the Galois group $\Gamma = Gal(E/\mathbb{Q})$ for finite Galois extensions E of $\mathbb{Q}$. We proved, that necessarily $G \subset GL_n(\mathcal{O}_{E_{ab}})$ holds, where E_{ab} denotes the maximal abelian subextension of E over $\mathbb{Q}$ (compare [2], Main Theorem). For totally real extensions $E/\mathbb{Q}$, or even extensions not containing nontrivial roots of 1, the above result can be expressed in a more spectacular form:
For $G \subset GL_n(\mathcal{O}_E)$ stable under the operation of the Galois group Γ the inclusion $G \subset GL_n(\mathbb{Z})$ holds, $\mathbb{Z}$ the ring of rational integers. The most important step was the reduction of the problem to the case of abelian groups G of prime exponent p.

Translated to the language of finite flat group schemes defined over $\mathbb{Z}$ this mentioned Theorem implies the complete classification of finite flat commutative group schemes over $\mathbb{Z}$ annihilated by a prime p (see Corollary 1 in [2]), and this answers a question of J. Tate [24]. The above mentioned facts are relevant to
1) the behaviour of classes of positive definite quadratic $\mathbb{Z}$-lattices under scalar extensions (compare [3] and [20]) and
2) for the existence of abelian varieties with good reduction everywhere (see [1,2,9,23]).

It is known that if F has unramified extensions then there exist examples of Galois stable finite groups $G \subset GL_n(\mathcal{O}_E)$ which are not fixed elementwise by the commutator subgroup of $Gal(E/F)$ (see [19]). How-

ever, in the situation studied in [2], where $E = \mathbb{Q}$ and there do not exist unramified extensions of the ground field, and there exist only cyclotomic fields $E = \mathbb{Q}(G)$. We also consider the role of the group of units in E for the existence of finite $Gal(E/F)$-stable groups G.

The methods used in the proofs, namely the detailed study of the operation of the higher ramification groups of the Galois group on the given Galois stable group G for the ramified primes in the field extension E over F using trivial action of higher ramification groups, together with some Odlyzko estimates, are similar to the methods used in [1] and [9].

We are interested in 2 basic conditions for the Γ-operation on G and the integrality of G.

A. G *is* Γ*-stable.*

B. $G \subset GL_n(\mathcal{O}_E)$

We intend to study the following questions:

Question 1.1. Do the conditions A., B. imply $G \subset GL_n(F)$?

Question 1.2. Classify the possible fields $E = F(G)$.

It is known that for unramified normal extensions E/F it is possible ([19], theorem 1) to construct a $Gal(E/F)-$ stable subgroup $G \subset GL_n(\mathcal{O}_E)$ such that $F(G) \neq F$. But if we assume that E/F has no unramified subextensions E_1/F, the following question can generalize the Main Theorem in [2]:

Question 1.3. Do the conditions A., B. imply $G \subset GL_n(FE_{ab})$, where E_{ab} is the maximal abelian subextension of $E/\mathbb{Q}$?

The above questions were still not considered for fields F of a positive characteristic p. However, contrary to a positive answer to the Question 1.1 in the case $F = \mathbb{Q}$, the answer for fields F of a positive characteristic p is negative, and it is surprisingly simple. If $E = R_1(x)$ is a finite extension of the field of rational functions $F = R(x)$ with a finite field R, and $R_1 \supset R$ such that $R_1 \neq R$, then for any finite group $G \subset GL_n(R_1), G \not\subset GL_n(R)$ the condition $F(G) \neq F$ is true. Note that in this situation $F(G)/F$ is unramified. Even in the case $R_1 = R$ the following theorem 2.1 gives a construction of a prescribed field E as $E = F(G)$ for an appropriate group G.

Acknowledgement The author is grateful to the referee for many useful remarks and suggested improvements.

2. Some results on the existence of Galois stable groups G

Theorem 2.1. *Let F be a global field of a positive characteristic p, and let E be a splitting field of some irreducible polynomial $f(y) \in F[y]$ whose roots are the conjugates of some element $t \in E$. Then $E = F(G)$ for any positive integer n and an appropriate group $G \subset GL_n(E)$. Moreover, if $t \in E$ is an element of $\mathcal{O}_E$ then $G \subset GL_n(\mathcal{O}_E)$.*

Proof. Let

$$g_t := \begin{vmatrix} 1 & t & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \ddots & \dots & 0 \\ \dots & \dots & \dots & \ddots & \\ 0 & 0 & \dots & \dots & 1 \end{vmatrix}.$$

Then $g_t^p = I_n$, the identity $n \times n$-matrix, and for any automorphism σ of E

$$g_t^\sigma = \begin{vmatrix} 1 & t^\sigma & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \ddots & \dots & 0 \\ \dots & \dots & \dots & \ddots & \\ 0 & 0 & \dots & \dots & 1 \end{vmatrix}.$$

We have $(g_t^\sigma)^p = I_n$, and the product of any 2 matrices g_t^σ for any automorphisms σ of E is still a $n \times n$ unitriangular matrix of order p. Therefore, a group G generated by all matrices g_t^σ is a finite abelian group of exponent p with non-trivial Galois operation of Γ such that $E = F(G) \neq F$ provided $t \notin F$. □

The reason for this constructive realizability of the above field E of characteristic p is that elements in G are not semisimple, the situation is completely different for fields E, F of characteristic 0, and even for extensions E/F of fields of characteristic $p > 0$, provided the order of G is not divisible by p. Under these conditions we can use the following criterion (Theorem 2.2) of the existence of groups G with $E = F(G) \neq F$ and entries in $\mathcal{O}_E$ which is true also for extensions E/F of fields in characteristic $p > 0$, and groups G of order not divisible by p. However, Lemma 2.2 below shows that for ramified extensions E/F with $E = F(G)$ the order of G is divisible by p.

The the following 2 lemmata have some extra motivation in papers [15,18].

Let us denote by $\mathcal{O}'_E$ the semilocal ring which is obtained by intersection of valuation rings of all ramified prime ideals in the rings $\mathcal{O}_E$.

Lemma 2.1. *Let J be an ideal of a Dedekind ring D of characteristic χ ($D = \mathcal{O}_E$ or its localization, or $D = \mathcal{O}'_E$), let $\{0\} \neq J \neq D$, and let $g \in GL_n(D)$ be a matrix of finite order, $g \equiv I_n(modJ)$. Then if $\chi = p > 0$, then $g^{p^j} = I_n$ for some positive integer j; if $\chi = 0$, then there is a prime $p \in J$, and $g^{p^i} = I_n$ for some positive integer i.*

Proof. We can assume that $h = g^a \neq I_n$ is a matrix of a prime order q. Pick J_0 to be a minimal ideal containing all entries of the matrix $B = g - I_n$. Since $h^q = I_n$, we have:

$$qB + \frac{q(q-1)}{2}B^2 + \cdots + B^q = 0_n$$

where 0_n is the zero $n \times n$-matrix. If $\chi = p > 0$ then necessarily $q = p$, otherwise the above identity would imply $J = J_0 = D$. If $\chi = 0$ then the same identity implies that $q \in J_0$, so we can change notation and denote $p = q$. This completes the proof. □

Lemma 2.2. *Let D be the same as in Lemma 2.1, let E/F be a Galois extension of global fields with the residue field of characteristic $p > 0$, and let $G \subset GL_n(D)$ be a finite Γ-stable subgroup. If the inertia subgroup $\Gamma_0 \subset \Gamma$ of some prime ideal $\mathfrak{p}$ in D operates non-trivially on G, i.e. $g^\gamma \neq g$ for some $\gamma \in \Gamma_0$ and some $g \in G$, then the order of G is divisible by p.*

Proof. We can assume $g^\gamma \neq g$ for some $\gamma \in \Gamma_0$ and some $g \in G$, so $h = g^{-1}g \equiv I_n(mod\mathfrak{p})$, and by Lemma 2.1 $h \in G$ is a matrix of order p^j. Then the order of G is divisible by p. This completes the proof of Lemma 2.2. □

If E, F are functional fields of characteristic $p > 0$, it is easy to apply the following criterion to a group G generated by all Galois conjugates of any semisimple matrix g to obtain an answer to Question 1.1. This criterion was earlier proven for fields E, F of characteristic 0 but can be literally extended to the functional fields of characteristic $p > 0$, it can be formulated in the following Theorem 2.2.

Let E, L be finite extensions of a global field F. Let $\mathcal{O}'_E$, $\mathcal{O}'_F$, $\mathcal{O}'_L$ be the rings that are obtained by intersection of valuation rings of all ramified

prime ideals in the rings $\mathcal{O}_E$, $\mathcal{O}_F$, $\mathcal{O}_L$. If $F = \mathbb{Q}$ or $R(x)$ we can define $\mathcal{O}_F$ to be the intersection of F and $\mathcal{O}_E$. Let $w_1, w_2, \ldots, w_d$ be a basis of $\mathcal{O}'_E$ over $\mathcal{O}'_F$, and let d be a square root of the discriminant of this basis. By the definition $d^2 = \det[\mathrm{Tr}_{E/F}(w_i w_j)]_{ij}$. It is known that

$$d = det[w_m^{\sigma_k}]_{k,m} = \det \begin{vmatrix} w_1 & \ldots & w_k & \ldots & w_t \\ w_1^{\sigma_2} & \ldots & w_k^{\sigma_2} & \ldots & w_t^{\sigma_2} \\ \vdots & & & & \\ w_1^{\sigma_t} & \ldots & w_k^{\sigma_t} & \ldots & w_t^{\sigma_t} \end{vmatrix}.$$

Let us suppose that some matrix $g \in GL_n(E)$ has order t ($g^t = I_n$) and all Γ-conjugates g^γ, $\gamma \in \Gamma$ generate a finite subgroup $G \subset GL_n(E)$ of exponent t. Let $\sigma_1 = 1, \sigma_2, \ldots, \sigma_d$ denote all automorphisms of the Galois group Γ of E over F. Assume that $L = E(\zeta_{(1)}, \zeta_{(2)}, \ldots, \zeta_{(n)})$ where $\zeta_{(1)}, \zeta_{(2)}, \ldots, \zeta_{(n)}$ are the eigenvalues of the matrix g, therefore $L = E(\zeta_p)$, ζ_p a primitive p-th root of unity. We will reserve the same notations for some extensions of σ_i to L, and the automorphisms of L/F will be denoted $\sigma_1, \sigma_2, \ldots, \sigma_r$ for some $r \geqslant t$. Let $E = F(G)$ be obtained by adjoining to F all coefficients of all $g \in G$. For a suitable choice of t elements of $\{\zeta_{(1)}, \zeta_{(2)}, \ldots, \zeta_{(n)}\}$ say $\zeta_{(1)}, \zeta_{(2)}, \ldots, \zeta_{(t)}$ the following theorem is true (the proof is given in [15], Theorem 1, see also [2]). This theorem is true also for integers in global fields of characteristic p. For convenience of the reader below we give the proof using the eigenvalues of semisimple matrices.

Theorem 2.2. *Let G be generated by all $g^\gamma, \gamma \in \Gamma_{E/F}$ and irreducible under $\mathrm{GL}_n(F)$-conjugation. Then G is conjugate in $\mathrm{GL}_n(F)$ to a subgroup of $\mathrm{GL}_n(\mathcal{O}'_E)$ if and only if all determinants*

$$d_k = \det \begin{vmatrix} w_1 & \ldots & w_{k-1} & \zeta_{(1)} & w_{k+1} & \ldots & w_t \\ w_1^{\sigma_2} & \ldots & w_{k-1}^{\sigma_2} & \zeta_{(2)}^{\sigma_2} & w_{k+1}^{\sigma_2} & \ldots & w_t^{\sigma_2} \\ \vdots & & & & & & \\ w_1^{\sigma_t} & \ldots & w_{k-1}^{\sigma_t} & \zeta_{(t)}^{\sigma_t} & w_{k+1}^{\sigma_t} & \ldots & w_t^{\sigma_t} \end{vmatrix}$$

are divisible by d in the ring $\mathcal{O}'_L$.

Proof. The proof is constructive. The proof is based on the commutativity of L-algebra LG and uses a system of linear equations that arise from commuting matrices

$$g = \sum_{i=1}^{t} w_i B_i \quad \text{and} \quad g^\sigma,\ \sigma \in \Gamma_{E/F}.$$

The eigenvalues of the commuting matrices B_i, $i = 1, 2, \ldots, t$ are the solutions of this system. We also use the fact that each semisimple matrix $B_i \in M_n(F)$ is conjugate in $\mathrm{GL}_n(F)$ to a matrix from $M_n(\mathcal{O}'_F)$ if and only if all its eigenvalues are contained in $\mathcal{O}'_L$ for some field $L \supset F$. In fact, since $g = B_1w_1 + B_2w_2 + \cdots + B_tw_t$, $(B_i \in M_n(F))$ is a generator of G and $w_1, ..., w_t$ is a basis of $\mathcal{O}'_E$ over $\mathcal{O}'_F$, we need to determine whether or not matrices $B_i, i = 1, ..., t$ are conjugate in $GL_n(F)$ to matrices $B'_i \in M_n(\mathcal{O}'_F)$. The latter depends on whether or not the eigenvalues of B_i are contained in $\mathcal{O}'_L$ (see Lemma 2.3 below).

Since $g = B_1w_1 + B_2w_2 + \cdots + B_tw_t$ $(B_i \in M_n(F))$ and matrix $W = [w_i^{\sigma_j}]_{j,i}$ is nondegenerate, the following system of linear matrix equations allows to express B_i as a linear combination of g^{σ_j}, $i, j = 1, 2, \ldots, t$:

$$B_i = \sum_{j=1}^{t} m_{ij} g^{\sigma_j},$$

where $[m_{ij}] = W^{-1}$. By our assumption the matrices g^{σ_j} commute pairwise. Therefore, all matrices B_i also commute with each other.

Since $L = E(\zeta_{(1)}, \zeta_{(2)}, \ldots, \zeta_{(n)})$ contains all eigenvalues of the generators of G, L is a splitting field for G and a splitting field for the L-algebra LG. Since LG is semisimple and commutative, by Wedderburn's theorem it is a direct sum of fields: $LG = F_1 + ... + F_m$, $F_i = e_iLG$ for primitive idempotents $e_i \in LG, i = 1, ..., m$.

Simultaneous reduction by conjugation of all B_i to diagonal form allows us to obtain the following system of linear equations in variables x_{ij}, $i = 1, 2, \ldots, n$, and its solutions are the eigenvalues of B_j, $j = 1, 2, \ldots, t$:

$$\begin{cases} x_{11}w_1 + x_{12}w_2 + \cdots + x_{1t}w_t = \theta_1, \\ x_{21}w_1 + x_{22}w_2 + \cdots + x_{2t}w_t = \theta_2, \\ \vdots \\ x_{n1}w_1 + x_{n2}w_2 + \cdots + x_{nt}w_t = \theta_n \end{cases} \qquad (*)$$

In the system above θ_i, $i = 1, 2, \ldots, n$ are the eigenvalues of g, and the solutions $x_{ij} := \lambda_{ij}$ of the system (*) are the eigenvalues of B_j. Then all the matrices B_i are integral if and only if the solutions of (*), that can be calculated using the Cramer's rule (if we select appropriate t equations), are all in the extended ring $\mathcal{O}_L$.

The irreducibility of G implies that the minimal polynomial of B_i is irreducible over F for each i such that B_i is not zero. So if one of the eigenvalues of B_i is in $\mathcal{O}'_L$ then all of them are since they are Galois conjugate.

Using the dual basis $w_1^*, ..., w_t^*$ to $w_1, ..., w_t$ with respect to the trace form one can see that the inverse matrix W^{-1} to $W = [w_i^{\sigma_j}]_{j,i}$ is of the form $W^{-1} = [w_j^{*\sigma_i}]_{j,i}$. In order to prove the claim of the proposition, we need to determine whether or not matrices $B_i, i = 1, ..., t$ are conjugate in $GL_n(F)$ to matrices $B_i' \in M_n(\mathcal{O}_F')$, since for the generator g of G the equation

$$g = B_1 w_1 + B_2 w_2 + \cdots + B_t w_t,$$

holds with $B_i \in M_n(F)$ and $w_1, ..., w_t$ a basis of $\mathcal{O}_E'$ over $\mathcal{O}_F'$. In fact, each semisimple matrix $B_i \in M_n(F)$ is conjugate in $\mathrm{GL}_n(F)$ to a matrix from $M_n(\mathcal{O}_F')$ if and only if all its eigenvalues are contained in $\mathcal{O}_L'$, which is proven in Lemma 2.3 below.

Cramer's rule now implies that $w_i^{*\sigma_j} = (-1)^{i+j} W_{i,j} det(W)^{-1}$, where $W_{i,j}$ is the (i,j)-minor of W. Over the splitting field L there is a basis which consists of eigenvectors for G. Let u be one such common eigenvector with

$$g^{\sigma_i} u = t_i u.$$

Then $\zeta_{(i)} := t_i^{\sigma_i{}^{-1}}$ is an eigenvalue of g. It also follows, that u is an eigenvector for B_k with eigenvalue

$$\lambda_k = \sum_{j=1}^{t} m_{kj} t_j = \sum_{j=1}^{t} (-1)^{j+k} W_{j,k} \zeta_{(j)}^{\sigma_j} det(W)^{-1}.$$

The cofactor expansion for determinants implies $\lambda_k = d_k / detW$ and therefore the eigenvalues of B_k are in $\mathcal{O}_L'$ iff $detW$ divides d_k.

To complete the proof of Theorem 2.2 we need the fact (which was used earlier in our proof) that semisimple matrices $B_i \in GL_n(F)$ are simultaneously conjugate in $GL_n(F)$ to a matrix from $GL_n(\mathcal{O}_F')$ if and only if all their eigenvalues are contained in $\mathcal{O}_L'$:

Let us consider a commutative F-algebra $A = F[B_1, \ldots, B_t]$, the F-span of matrices $B_1, \ldots, B_t$. Then A is $GL_n(F)$-irreducible since otherwise G would not be $GL_n(F)$-irreducible. It is known that A is a field and the degree $[A : F] = n$, see e.g. [11], chapter 1, sect. 1, corollary 2. We use the following lemma to complete the proof. □

Lemma 2.3. *i) Let all eigenvalues λ_j, $j = 1, 2, \ldots, k$ of the semisimple matrices $B_i \in M_n(F), i = 1 \ldots, t$ be contained in the ring $\mathcal{O}_L'$ for some field $L \supset F$. Then B_i are conjugate in $GL_n(F)$ simultaneously to matrices that are contained in $M_n(\mathcal{O}_F')$. ii) Conversely, if the semisimple matrices B_i are contained in $M_n(\mathcal{O}_F')$ and B_i are diagonalizable over a field $L \supset F$, then their eigenvalues are contained in $\mathcal{O}_L'$.*

Proof. i) By the virtue of [11], chapter 1, sect. 1, corollary 2 we can consider A to be a field extending F. Let $a_1, a_2, \ldots, a_n$ be a basis of $\mathcal{O}'_A$ over $\mathcal{O}'_F$. Then for any $B \in A$ we have $B = b_1a_1 + \cdots + b_na_n$, and the elements $b_i \in F$ are contained in $\mathcal{O}'_F$ iff $B \in \mathcal{O}'_A$. But all coefficients k_{ij} of the characteristic polynomials $f_i(x) = k_{i0} + k_{i1}x + ... + k_{in}x^n$ of the matrices B_i are contained in $\mathcal{O}'_L$, and $k_{in} = 1$, so $B_i \in A$ are integral over F. It follows that $B_i = b_{i1}a_1 + \cdots + b_{in}a_n$, and $b_{ij} \in \mathcal{O}'_F$. If $v \in F^n$ is a non-zero vector in F^n, then $a_1v, a_2v, \ldots, a_nv$ is a basis of F^n, and $B_ia_jv = \sum_k c_{ijk}a_kv$, where $c_{ijk} \in \mathcal{O}'_F$. It follows that for any i the matrix $C_i = [c_{ijk}]_{k,j}$ belongs to $GL_n(\mathcal{O}'_F)$, and C_i is the matrix of the operator B_i in the basis $a_1v, a_2v, \ldots, a_nv$ of F^n. Therefore, B_i is conjugate in $GL_n(F)$ to C_i for any $i = 1, \ldots, t$.

ii) Consider the characteristic polynomials $f_i(x) = k_{i0} + k_{i1}x + ... + k_{in}x^n$ of the matrices B_i. Since $k_{in} = 1$ and all k_{ij} are in $\mathcal{O}'_F$ all roots of $f(x)$ are in $\mathcal{O}'_L$. This completes the proof of Lemma 2.3. □

Note that in the situation of Lemma 2.3, i) the F-algebra $A = F[B_1, \ldots, B_t]$ is isomorphic to the field $L = F[\lambda_1, \ldots, \lambda_k]$ where λ_j, $j = 1, 2, \ldots, k$ are all eigenvalues of the matrices $B_i, i = 1 \ldots, t$.

Let us suppose that $K = \mathbb{Q}$, the field of rationals, or $K = \mathbb{Q}(\sqrt{d})$ and d is a negative rational integer. We consider the set $\mathcal{O}(N) = \{\alpha \in \mathcal{O}_K \big| \, |N_{K/\mathbb{Q}}(\alpha)| \leqslant N\}$ where $N_{K/\mathbb{Q}}$ is the norm map. The proof of the following Theorem is based on the result by S. D. Cohen (Theorem 1 in [7]) using the large sieve method combined with some asymptotic estimates for the number of integral polynomials having bounded coefficients with respect to the norm and reducible over $K(\sqrt{b})$ (b is contained in a finite set of elements from $\mathcal{O}_K$):

Theorem 2.3. *Let $v(N)$ denote the total number of polynomials of degree m with coefficients in $\mathcal{O}(N)$, and let $\psi(N)$ denote the number of those polynomials whose splitting fields do not contain any fields $K(G) \neq K$ for $G \subset GL_n(\mathcal{O}_E), E \supset K$ and fixed n. Then*

$$\lim_{N \to \infty} \frac{\psi(N)}{v(N)} = 1.$$

The error term can be estimated in the case $K = \mathbb{Q}$ as

$$v(N) - \psi(N) = o(N^{m+0.5}(lnN)^2)$$

.

Theorem 2.3 shows that "almost all" fields are not realizable via adjoining matrix coefficients of Γ-stable groups to the field of rational numbers or

its imaginary quadratic extensions if this coefficients are contained in the rings of integers of algebraic number fields.

Remark that we can also consider other number fields, but it will be necessary to rearrange the definition of $\mathcal{O}(N)$, compare [7]. Note that proof below, specially in the case 1), can produce explicit estimates, and we can also use the estimates in [10,13,14].

Proof. We use properties of distribution of Galois groups of polynomials that were considered by S. D. Cohen [7], for the case $K = \mathbb{Q}$ see also [25]. According to [7] the number of polynomials in question having the symmetric Galois group S_m, divided by the total number of polynomials in question, approaches 1 when $N \to \infty$. Therefore, we can consider only the number of these K-irreducible polynomials that are reducible over $K(\sqrt{\alpha})$ for a finite number of α. The elements $\sqrt{\alpha}$ can be contained only in a finite number of extensions $K(G)$ that have no ramified primes $p \geqslant m! + 1$ (since p must divide the order of $\Gamma = S_n$, compare lemma 2.2 above) and have degree $m!$ over K. Let us estimate the number of these polynomials. However, if $K = \mathbb{Q}$, the situation is simplier, and we have to check only 2 possible extensions of $\mathbb{Q}$: the fields $\mathbb{Q}[i]$ and $\mathbb{Q}[\sqrt{-3}]$.

1) Let us consider the case $K = \mathbb{Q}$.

Note that in the virtue of the above result on the symmetric Galois group S_m and the Main Theorem in [2] (see also theorem 2 in [15]) which implies that only for fields $\mathbb{Q}(G)$ containing nontrivial roots of 1 it may happen that $\mathbb{Q}(G) \neq \mathbb{Q}$, we have to eliminate a possibility that $\mathbb{Q}(G)$ has nontrivial roots of 1 and simultaneously the Galois group of $Gal(\mathbb{Q}(G)/\mathbb{Q})$ is S_m. The latter is possible only if one of the primitive roots $\zeta_4 = i$ or $\zeta_3 = \frac{-1+\sqrt{-3}}{2}$ is in $K\mathbb{Q}(G)$.

Let us start from the case $i \in \mathbb{Q}(G)$. Let $k, l, k + l = m$ be positive integers such that an integral polynomial $A(x)$ satisfies the conditions of Theorem 2.3, $A(x) = a(x)b(x)$ with $a(x) = \sum_{i=0}^{k} a_i x^i, a_i \in \mathbb{Z}[i]$, and $b(x) = \sum_{j=0}^{l} b_j x^j, b_j \in \mathbb{Z}[i]$, and $a_0 \neq 0, a_k \neq 0, b_0 \neq 0, b_l \neq 0$. Since the number of possible polynomials $A(x)$ with either the first or the last coefficient equal 0 is $\sim N^m$ while the total number of polynomials in $\mathcal{O}(N)[x]$ is $\sim N^{m+1}$, so the polynomials $A(x)$ with either the first or the last coefficient equal 0 do not give any contribution asymptotically. Let us show that the number of the sets of coefficients $(a_0, a_1, \ldots, a_k, b_0, b_1, \ldots, b_l)$ admissible for polynomials $a(x), b(x)$ also do not contribute anything asymptotically. The ring $\mathbb{Z}[i]$ is euclidean, and $\pm 1, \pm i$ are the only invertible elements in $\mathbb{Z}[i]$, also for any integer D $|ab| \leqslant |D|$ imply $|b| \leqslant |D|$ or $|a| \leqslant |D|$. This implies

$|a_i| \leqslant C(m)N$ and $|b_j| \leqslant C(m)N$ where $C = C(m)$ depends only on m. Also we have $1 \leqslant |a_0 b_0| \leqslant N$ and $1 \leqslant |a_k b_l| \leqslant N$. Let us estimate the number $L(N)$ of pairs of Gaussian integers $a, b \in \mathbb{Z}[i]$ such that $1 \leqslant |ab| \leqslant N$. We can write $a = a_1' + a_2' i = c_1(\alpha_1 + \alpha_2 i)$ where c_1, α_1, α_2 are rational integers, α_1, α_2 are coprime, so c_1 is the greatest common divisor $c_1 = (\alpha_1, \alpha_2)$of α_1, α_2. Also, let $b = b_1' + b_2' i = c_2(\beta_1 + \beta_2 i)$ where c_2, β_1, β_2 are rational integers, and $c_2 = (\beta_1, \beta_2)$. It is known (see [8], ch. 4, sect. 68 or [6], ch. 9, sect. 6 and appendix B) that the number $F(t)$ of primitive representations of a positive integer t as a sum of 2 squares does not exceed $c_f 2^s$ where c_f is a constant depending only on the form $f(x_1, x_2) = x_1^2 + x_2^2$, the sum of 2 squares, $c_f = 4$ in our case, and s is the number of distinct prime divisors of t. Denote by $M(j)$ the number of all pairs of integers c_1, c_2 such that $|c_1 c_2| \leqslant j$ (note that both c_1 and c_2 can be positive or negative). Then (see e.g. [12], p.264) $M(j) \sim 4([j/1] + [j/2] + \cdots + [j/k] + \dots) = 4(j \cdot lnj + O(j))$, where $[x]$ denotes the greatest integer $\leqslant x$. Note that we can always write $F(t) \leqslant c_f t$. Let us estimate the number $L(N)$ of integers a, b introduced above. We can use that also $F(t) = c_f 2^s = o(t)$, and also $F(t) = c_f 2^s = o(t^{1/4})$ for $t \geqslant N^{1/4}$ (see e.g. [12], 18.7, p.270).

$$L(N) = \sum_{t=1}^{N} M(N/t)F(t) = o(\sum_{t=1}^{N^{\frac{1}{4}}} (N/t \cdot ln(N/t))t) + o(\sum_{t=N^{\frac{1}{4}}}^{N} (N/t \cdot ln(N/t))t^{\frac{1}{4}}) =$$

$$o(\int_1^{N^{\frac{1}{4}}} N \cdot ln(N/x)dx) + o(\int_{N^{\frac{1}{4}}}^{N} N \cdot ln(N/x)dx^{\frac{1}{4}}) = o(N^{\frac{5}{4}} lnN)$$

So the number of possible systems of (a_0, a_k, b_0, b_l) involving 2 couples (a_0, b_0) and (a_k, b_l) of coefficients is $o(N^{2.5}(lnN)^2)$. This estimate may be improved but this is not essential for our theorem. Finally, the number of polynomials $A(x)$ that are reducible in $\mathbb{Z}[i][x]$ is $o(N^{k-1}N^{l-1}N^{2.5}(lnN)^2) = o(N^{m+0.5}(lnN)^2) = o(N^{m+1})$, and we can combine this estimate with the estimate in [7] (see also [10]), which implies that the number of polynomials $A(x) = \sum_{i=0}^{m} p_i x^i \in \mathcal{O}(N)[x]$ whose Galois group is not symmetric is $O(N^{m+0.5} lnN)$. So our claim is true for polynomials in $\mathbb{Z}[i][x]$.

In a similar way we can consider the polynomials $A(x) \in \mathbb{Z}[\zeta_3][x]$. The number of these polynomials can be estimated using the quadratic form $f(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2$ corresponding to multiplication in the ring $\mathbb{Z}[\zeta_3]$, which is equivalent to the form $f(y_1, y_2) = y_1^2 + y_1 y_2 + y_2^2$, where

$x_1 = y_1 + y_2, x_2 = y_2$. The constant c_f for this form is $c_f = 6$ (see [8], ch. 4, sect. 70 or [6], ch. 9, sect. 6 and appendix B), and our argument can be used without changes in the case of the ring $\mathbb{Z}[\zeta_3]$ instead of $\mathbb{Z}[i]$.

2)Let us consider the case $K = \mathbb{Q}(\sqrt{d})$, $d < 0$, $d \in \mathbb{Z}$.

Let$f \in \mathcal{O}(N)[x]$ and $f = g \cdot g'$, $g, g' \in K(\sqrt{\alpha})[x], \sqrt{\alpha} \notin K$. Let $\mathcal{E} \in \mathcal{O}_{K(\sqrt{\alpha})}$ be a unit of infinite order. We can suppose that after some adjustment both the height $|g| = \max|a_i|$ of $g = \sum a_i x^i$ and the height $|g'|$ of $g' = \sum a'_i x^i$ are equal up to a constant $c = c(K, m)$. Indeed, let $|g| = A$, $|g'| = B$, $|f| = c_0 N$, $c_0 = c_0(K, m)$. Let $t = \log_{\mathcal{E}} \left(\frac{A}{\sqrt{N}}\right)$, then changing g and g' to $p = \mathcal{E}^{-[t]} g$ and $p' = \mathcal{E}^{[t]} g$ respectively we obtain $|p| \sim \sqrt{N}$, $|p'| \sim \sqrt{N}$, that is $|p| \leqslant c_1(K, m)\sqrt{N}$ and $|p'| \leqslant c_2(K, m)\sqrt{N}$. As $p = p_1 + \sqrt{\alpha} p_2$ and $p' = p'_1 + \sqrt{\alpha} p_2$ for $p_i, p'_i \in K[x]$ and $p' = p^\sigma$ for non-identical automorphism σ of $K(\sqrt{\alpha})$ over K, we can see that $|p_i| \leqslant c_3\sqrt{N}$ and $|p'_i| \leqslant c_3\sqrt{N}$ for $i = 1, 2$ and $c_3 = c_3(K, m)$. Therefore, there are only $(c_2\sqrt{N})^{2\cdot(m/2+1)} = c_4 N^{m+2}$, $c_4 = c_4(K, m)$, polynomials that are reducible over $K(\sqrt{\alpha})$. Likewise, there are $c_5 N^{2(m+1)}$, $c_5 = c_5(K, m)$, polynomials f in $\mathcal{O}(N)[x]$ and it is obvious that

$$\lim_{N\to\infty} \frac{c_4 N^{m+2}}{c_5 N^{2m+2}} = 0.$$

Note that the number of polynomials $f \in \mathcal{O}(N)[x]$ that are reducible already in $\mathcal{O}(N)[x]$ do not give any contribution asymptotically. Moreover, according to the result in [7], the number of polynomials in $O(N)[x]$ whose Galois group is not symmetric do not contribute asymptotically as well. So, we have shown that the number of polynomials whose splitting fields can contain any $K(G) \neq K$ is small asymptotically, and this completes the proof of Theorem 2.3. □

3. Some examples of Galois stable groups G.

Example 3.1. It is difficult to transfer the idea of reduction to abelian Galois stable groups G of composite order. For $p \neq 2$ the simplest example can be constructed as follows: Let

$$g_2 := \begin{vmatrix} 0 & \sqrt[p]{u} \\ (\sqrt[p]{u})^{-1} & 0 \end{vmatrix}$$

and $g := diag(g_2, I_{p-2}) \in GL_p(\mathcal{O}_K)$. Then $g^\gamma, \gamma \in \Gamma$ and $\zeta_p I_p$ generate a finite Γ-stable nonabelian subgroup of $GL_p(\mathcal{O}_K)$ of order divisible by 2 and p.

Example 3.2. Let

$$g := \begin{vmatrix} \sqrt{3+\sqrt{2}} & -\sqrt{2+\sqrt{2}} \\ \sqrt{2+\sqrt{2}} & -\sqrt{3+\sqrt{2}} \end{vmatrix},$$

let $E = F(\sqrt{3+\sqrt{2}}), F = \mathbb{Q}(\sqrt{3+\sqrt{2}} \cdot \sqrt{2+\sqrt{2}})$. Then E/F is ramified at 2, the ramification is wild, and $G = \{g, -g, I_2, -I_2\} \subset GL_2(\mathcal{O}_E)$ is a Γ-stable subgroup of order 4 and exponent 2.

Example 3.3. The difficulties to extend the results [2] to the case of relative extensions over a ground field R ramified over $\mathbb{Q}$ can be illustrated using the following construction:

If there exist an intermediate extension $L = F(\sqrt[p]{u}) \subset E$ for some unit $u \in \mathcal{O}_E$, we can put

$$g = \begin{vmatrix} 0 & \sqrt[p]{u} & 0 \dots & 0 \\ 0 & 0 & \sqrt[p]{u} \dots & 0 \\ \vdots & \ddots & \ddots & \\ 0 & \dots & 0 & \sqrt[p]{u} \\ \sqrt[p]{u}^{1-p} & \dots & 0 & 0 \end{vmatrix}$$

Then $g^\gamma, \gamma \in \Gamma$ and $\zeta_p I_p$ generate a finite Γ-stable subgroup of $GL_p(\mathcal{O}_E)$.

Hence for relative extensions $E/F, L \neq F$ and some units u it may happen that neither $F(\zeta_p, \sqrt[p]{u}) \subset FE_{ab}$ nor $F(\sqrt[p]{u})/F$ is unramified, when $L = F(\zeta_p)$.

However, it is still possible to extend, in part, the result of the Main Theorem in [2] to the case of relative algebraic number fields extensions, that are composites of fields having coprime discriminants.

References

1. V. A. Abrashkin, Galois moduli of period p group schemes over a ring of Witt vectors, Math. USSR Izvestiya,1988 vol. 31 pp. 1–46
2. H.-J. Bartels, D.A. Malinin, Finite Galois stable subgroups of GL_n, in: Noncommutative Algebra and Geometry, Edited by C. de Concini, F. van Oystaeyen, N. Vavilov and A. Yakovlev, Lecture Notes In Pure And Applied Mathematics, vol 243, 2006, pp. 1–22
3. H.-J. Bartels, Zur Galoiskohomologie definiter arithmetischer Gruppen, J. reine angew. Math., 1978, vol 298, pp. 89–97
4. H.-J. Bartels, Y. Kitaoka, Endliche arithmetische Untergruppen der GL_n, J. reine angew. Math. , 1980, vol 313, pp. 151–156
5. G. Cardona, Representations of G_k-groups and twists of the genus two curve $y^2 = x^5 - x$, J. Algebra , vol 303, 2006, pp. 707–721

6. Cassels J. W. S., Rational quadratic forms. London Mathematical Society Monographs, 13., Academic Press, London-New York, 1978.
7. S. D. Cohen, The distribution of the Galois groups of integral polynomials, Ill. J. Math., 1979, vol 23 , pp. 135–152
8. P. G. Lejeune Dirichlet, Vorlesungen ueber Zahlentheorie. (German) Herausgegeben und mit Zusaetzen versehen von R. Dedekind. Vierte, umgearbeitete und vermehrte Auflage, Chelsea Publishing Co., New York, 1968
9. J.-M. Fontaine, Il n'y a pas de variété abélienne sur $\mathbb{Z}$, Invent. math., 1985, vol 81 , pp. 515–538
10. P.X. Gallagher, The large sieve and probabilistic galois theory In: Proc. Symp. pure Math., 1973. V. XXIV. P. 91–101.
11. F.R. Gantmakher, The theory of matrices, 4th ed., "Nauka",addr Moscow, 1988; English transl. of 1st ed., Vols 1, 2, Chelsea, New York, 1959
12. G. H. Hardy, E. M. Wright, An introduction to the theory of numbers. The fourth edition., Oxford University Press, Oxford, 1975.
13. H.W. Knohloch, Zum Hilbertschen Irreduzibilit, Abh. Math. Sem. Hamburg, 1955. B. 19, S. 176–190.
14. H.W. Knobloch , Die Seltenheit der reduzibien Polynome, Jber. Deutch. Math. Verein., 1956. Abt.1. S. 12–19.
15. D.A.Malinin, Galois stability for integral representations of finite groups, Algebra i analiz, vol 12 , 2000, pp. 106–145; English transl. in St. Petersburg Math. J. 12, N 3
16. D.A.Malinin, Integral representations of p-groups over local fields, Dokl. Akad. Nauk SSSR, vol 309, N 5, 1989, pp. 1060–1063; English transl. in Sov. Math. Dokl. 40 (1990), N 3
17. D.A.Malinin, Integral representations over local fields for p-groups of a given class of nilpotency , St. Petersburg Math.J., 1999, vol 10, N 1, pp. 45–52
18. D.A.Malinin, Integral representations of finite groups with Galois action, Dokl. Russ. Akad. Nauk, 1996, vol 349, pp. 303–305
19. D. A. Malinin, On Existence of finite Galois stable groups over integers in unramified extensions of number fields, Publ. Math. Debrecen, 2002, vol 60, 1-2, pp. 179–191
20. J.Rohlfs, Arithmetische definierte Gruppen mit Galois-operation, Invent. Math., 1978, vol 48, pp. 185–205
21. J.Ritter and A.Weiss, Galois action on integral representations, J. London Math. Soc. (2), 1992 , vol 46, pp. 411–431
22. J.Ritter and A.Weiss, Regulators and Galois stability, Math. Nachr., 1992, vol 158, pp. 27–41
23. I. Shafarevich, I.:, Algebraic number fields, Proceedings of the International Congress of Mathematics, Stockholm 1962. Amer. Math. Soc. Translations, vol 31, pp. 25-39 , 1963
24. J. Tate, p–Divisible Groups, in: Conf. Local Fields (Dreibergen),addr Springer Verlag, Berlin and New York, 1967, pp. 158–183
25. B.L. Van der Waerden , Die Seltenheit der reguziblen Gleichungen mit Affekt, Math. Ann., 1934. B. 109, S. 13–16.

Number of points of non-absolutely irreducible hypersurfaces

Robert Rolland

Institut de Mathématiques de Luminy
case 930, F13288 Marseille cedex 9, France
E-mail : robert.rolland@acrypta.fr

Let F_q be the finite field with q elements and $n \geq 2$ an integer. We provide a bound on the number of $\mathbb{F}_q$-rational points of the hypersurfaces in the affine space and in the projective space of dimension n, defined on $\mathbb{F}_q$ which are irreducible over $\mathbb{F}_q$ but non-absolutely irreducible.

Keywords: finite field, algebraic set, variety, hypersurface, code, weight

1. Introduction and Notation

Let p be a prime, $q = p^t$ and $\mathbb{F}_q$ the finite field with q elements. We denote by $\overline{\mathbb{F}}_q$ the algebraic closure of $\mathbb{F}_q$. In the paper n is an integer ≥ 2 and d an integer such that

$$2 \leq d < n(q-1),$$

and $\mathbb{P}^n(\mathbb{F}_q)$ is the projective space of dimension n over $\mathbb{F}_q$; write $\mathbb{P}^n = \mathbb{P}^n(\overline{\mathbb{F}}_q)$.

Let us denote by $\mathcal{RP}(q,d,n)$ the space of reduced polynomials, that is, polynomials with partial degrees $\leq q-1$ in n variables, coefficients in $\mathbb{F}_q$ and total degree $\leq d$. The Generalized Reed-Muller code of order d over $\mathbb{F}_{q^n}$ is defined, using the elements $P \in \mathcal{RP}(q,d,n)$, by taking for codewords $\big(P(x)\big)_{x\in\mathbb{F}_{q^n}}$. Let $Z_q(P)$ be the set of zeros of the polynomial P in $\mathbb{F}_q$, and $\#Z_q(P)$ the number of these zeros. The weight $W(P)$ of the codeword associated to P satisfies the relation

$$W(P) = q^n - \#Z_q(P).$$

In particular, the minimum distance is attained by the non-zero polynomials over $\mathbb{F}_{q^n}$ which have the maximum number of zeros. It is known (Kasami and al. [3]) that this maximum number of zeros is $q^n - (q-b)q^{(n-a-1)}$ where a and b are the quotient and the remainder in the euclidian division of d by

$q-1$. The polynomials attaining this bound are products of d polynomials of degree 1 (Delsarte, Goethals, McWilliams [2]).

Theorem 1.1 (Delsarte,Goethals,McWilliams). *The maximum number of $\mathbb{F}_q$-rational points, for an algebraic set V of degree d in the affine space of dimension n which is not the whole space $\mathbb{F}_q^n$ is attained if and only if*

$$V = \left(\bigcup_{i=1}^{a}\left(\bigcup_{j=1}^{q-1} V_{i,j}\right)\right)\left(\bigcup_{j=1}^{b} W_j\right),$$

where the $V_{i,j}$ and W_j are d distinct hyperplanes defined on $\mathbb{F}_q$ such that for each fixed i the $V_{i,j}$ are $q-1$ parallel hyperplanes, the W_j are b parallel hyperplanes and the $a+1$ distinct linear forms directing these hyperplanes are linearly independant.

Cherdieu and Rolland in [1] gave, under some restrictions on d and q, the second codeword weight, that is the weight just above the minimum distance, which is not to be confused with the second order distance, and have proved that this weight is also attained by some products of linear functions. Sboui in [5] has weakened the restrictions on d and q. We improve here the estimates of the number of zeros of a reduced polynomial in n variables, of total degree at most d and coefficients in the finite field $\mathbb{F}_q$, which is irreducible but non-absolutely irreducible.

The result of Kasami and al. was extended by Sørensen in [7] to the projective generalized Reed-Muller codes, over the projective space $\mathbb{P}^n(\mathbb{F}_q)$, introduced by Lachaud in [4]. Here we extend the result of Delsarte and al. of [2] and we give estimates for the projective case.

2. Irreducible but non-absolutely irreducible polynomials

First, we give a key lemma which can be found in [6]:

Lemma 2.1. *Let P be a non-zero irreducible but not absolutely irreducible polynomial over the finite field $\mathbb{F}_q$, in n variables and of degree d. Then one can find a finite extension $\mathbb{F}_{q'}$ such that there exists a unique polynomial Q absolutely irreducible over the finite field $\mathbb{F}_{q'}$, in n variables and of degree d', satisfying the relation*

$$P = \prod_{\sigma \in G} Q^{\sigma},$$

where $G = Gal(\mathbb{F}_{q'}/\mathbb{F}_q)$ is the Galois group of $\mathbb{F}_{q'}$ over $\mathbb{F}_q$ and

$$Deg(P) = [\mathbb{F}_{q'} : \mathbb{F}_q] Deg(Q).$$

Let us remark that in the previous lemma, if P is homogeneous, then Q is also homogeneous. Hence we obtain the following lemma:

Lemma 2.2. *Let V be a projective algebraic set of dimension $n-1$ and degree d in the projective space $\mathbb{P}^n$, defined on the finite field $\mathbb{F}_q$, irreducible over $\mathbb{F}_q$ but non-absolutely irreducible. Then there exists a finite extension $\mathbb{F}_{q'}$ such that there exists a unique absolutely irreducible projective variety W defined over $\mathbb{F}_{q'}$ of degree d' such that*

$$V = \bigcup_{\sigma \in G} W^\sigma,$$

where $G = Gal(\mathbb{F}_{q'}/\mathbb{F}_q)$ is the Galois group of $\mathbb{F}_{q'}$ over $\mathbb{F}_q$ and

$$d = [\mathbb{F}_{q'} : \mathbb{F}_q]d'.$$

2.1. *The affine case*

Now, let us give the main theorem for the affine case.

Theorem 2.1. *Let $P \in \mathcal{RP}(q,d,n)$ be an irreducible but non-absolutely irreducible polynomial of degree d. Let us set a and b such that $d = a(q-1)+b$ and $0 \leq b < q-1$. Then the number $\#Z_q(P)$ of zeros of P over $\mathbb{F}_q$ satisfies*

$$\#Z_q(P) < q^n - 2q^{n-\lfloor \frac{d}{2(q-1)} \rfloor - 1}.$$

Also, if $a = 0$ this estimate can be improved to the following:

$$\#Z_q(P) < \frac{d}{2} q^{n-1}.$$

Proof. Using the lemma 2.1 we get:

$$Z_q(P) = \bigcup_{\sigma \in G} Z_q(Q^\sigma).$$

However all the conjugate polynomials Q^σ have the same zeros in $\mathbb{F}_q$. Hence $Z_q(P) = Z_q(Q)$.

Let us denote by s the dimension $[\mathbb{F}_{q'} : \mathbb{F}_q]$ of the vector space $\mathbb{F}_{q'}$ over the field $\mathbb{F}_q$. We know that:

$$d = Deg(P) = sDeg(Q) = sd'.$$

If $(w_1, \cdots, w_s)$ is a basis of $\mathbb{F}_{q'}$ over $\mathbb{F}_q$

$$Q(X) = \sum_{j=1}^{s} h_j(X) w_j,$$

where $h_j \in \mathcal{RP}(q, d', n)$ and are not all zero. Hence,

$$Z_q(P) = \bigcap_{j=1}^{s} Z_q(h_j).$$

All the non-zero h_j cannot be the same products of degree one polynomials (in this case, Q would be proportional to a polynomial over $\mathbb{F}_q$), so that, by the result of Delsarte, Goethals, McWilliams [2], $\#Z_q(P)$ cannot attain the maximum number of zeros given by the formula of Kasami, Lin, Peterson [3]

$$\#Z_q(P) < q^n - (q - b')q^{n-a'-1},$$

where $d' = a'(q-1) + b'$ and $0 \leq b' < q - 1$. But a' is the integer part of $d'/q - 1$, namely,

$$a' = \left\lfloor \frac{d'}{q-1} \right\rfloor = \left\lfloor \frac{d}{s(q-1)} \right\rfloor.$$

In any case

$$\#Z_q(P) < q^n - (q - (q-2))q^{n-\left\lfloor \frac{d}{s(q-1)} \right\rfloor - 1}$$

so that

$$\#Z_q(P) < q^n - 2q^{n-\left\lfloor \frac{d}{2(q-1)} \right\rfloor - 1}.$$

Now, if $a = 0$ then $a' = 0$ and we can improve the previous estimate. In this case we know that $b' = d' = d/s$, so that:

$$\#Z_q(P) < q^n - (q - d/s)q^{n-1},$$

$$\#Z_q(P) < \frac{d}{s}q^{n-1} \leq \frac{d}{2}q^{n-1}. \qquad \square$$

2.2. *The projective case*

Let us consider now a projective algebraic set $V(P)$ defined on $\mathbb{F}_q$ by one reduced homogeneous polynomial $P(x_0, x_1, \cdots, x_n)$ of degree d with coefficients in $\mathbb{F}_q$. We denote by $V_q(P)$ the set of the $\mathbb{F}_q$-rational points of $V(P)$. We suppose that P is irreducible over $\mathbb{F}_q$ but non-absolutely irreducible and we search for an estimate of the cardinality $\#V_q(P)$.

Lemma 2.3. *Let P be a homogeneous polynomial in $n+1$ variables of total degree d, with coefficients in $\mathbb{F}_q$, which does not vanish on the whole projective space $\mathbb{P}^n(\mathbb{F}_q)$. Then the following hold.*

(1) The number of $\mathbb{F}_q$-rational points $\#V_q(P)$ of the projective algebraic set defined by P satisfies the following:

$$\#V_q(P) \leq \frac{q^{n+1}-1}{q-1} - (q-b)q^{n-a-1}, \tag{1}$$

where

$$d-1 = a(q-1)+b \quad 0 \leq b < q-1.$$

(2) The bound in (1) is attained. The algebraic sets attaining this bound are exactely the following: there exists a hyperplane H defined on $\mathbb{F}_q$ such that P vanishes on H, and P restricted to the affine space $\mathbb{P}^n \setminus H$ is a maximal affine algebraic set as described in Theorem 1.1. As a consequence P is a product of d homogeneous polynomials of degree 1.

Proof. The point (1) is proved by Sørensen in [7]. However, in order to prove at the same time the new point (2) let us rewrite the proof. Suppose first that $V(P)$ contains a hyperplane H. We can suppose that this hyperplane is given by $x_0 = 0$, so that $P = x_0 P_1$, where P_1 is an homogeneous polynomial of degree $d-1$. The complement of H is an affine space

$$A = \{x \in \mathbb{P}^n \ | x_0 = 1\}.$$

Let $\widetilde{P_1}$ be the polynomial in n variables obtained from P_1 by setting $x_0 = 1$. This polynomial is defined on A and does not vanish on the whole affine space A. Hence,

$$\#V_q(P) \leq \#H_q + \#Z_q(\widetilde{P_1}),$$

where H_q is the set of $\mathbb{F}_q$-rational points of H. Now, using the result of Kasami and al. ([3]), we obtain:

$$\#V_q(P) \leq \frac{q^n-1}{q-1} + q^n - (q-b)q^{n-a-1},$$

$$\#V_q(P) \leq \frac{q^{n+1}-1}{q-1} - (q-b)q^{n-a-1}.$$

The bound is attained if and anly if the polynomial $\widetilde{P_1}$ verifies the conditions of maximality given in Theoreme 1.1.

The theorem is clearly true for $n = 1$. Let us suppose that its true for $n-1$. Now, suppose that $V(P)$ does not contain a hyperplane. As $\mathbb{P}^n$ can be covered by affine subsets of dimension n, it exists an affine subset A where P is not always zero. Let us write:

$$\mathbb{P}^n = A \cup H,$$

where H is the hyperplane at infinity. Hence,

$$\#V_q(P) = \#(V_q(P) \cap A) + \#(V_q(P) \cap H),$$

and using the bound of Kasami, Lin , Peterson and the induction hypothesis, we get

$$\#V_q(P) \leq q^n - (q-\beta)q^{n-\alpha-1} + \frac{q^n-1}{q-1} - (q-b)q^{n-a-1},$$

where $d = \alpha(q-1)+\beta$. Then:

$$\#V_q(P) < \frac{q^{n+1}-1}{q-1} - (q-b)q^{n-a-1}.$$

So that (1) is satisfied, and the bound cannot be attained by such an algebraic set. In conclusion, the bound is attained only by the algebraic sets described in (2). □

Theorem 2.2. *Let $V(P)$ be a projective algebraic set of degree d in $\mathbb{P}^n$ defined over $\mathbb{F}_q$ by the homogeneous polynomial P. We suppose that P is irreducible but non-absolutely irreducible over $\mathbb{F}_q$. Let a and b be two integers such that*

$$d-1 = a(q-1)+b \text{ and } 0 \leq b < q-1.$$

The number of $\mathbb{F}_q$-rational points of $V(P)$ satisfies the following:

$$\#V_q(P) < \frac{q^{n+1}-1}{q-1} - 2q^{n-\lfloor \frac{d-2}{2(q-1)} \rfloor - 1}.$$

Moreover if $a = 0$ then this estimate can be impoved by the following one:

$$\#V_q(P) < \frac{q^n-1}{q-1} + \left(\frac{d}{2}-1\right) q^{n-1}.$$

Proof. Using the notations and the results of lemma 2.1 we get:

$$V(P) = \bigcup_{\sigma \in G} V(Q^\sigma),$$

where $V(Q^\sigma)$ is the projective hypersurface defined by the absolutely irreducible homogeneous polynomial Q^σ of degree $d' = d/s$ and coefficients in the degree s extension $F_{q'}$ of F_q. Following the proof of Theorem 2.1 and using Lemma 2.3,

$$\#V_q(P) < \frac{q^{n+1}-1}{q-1} - (q-b')q^{n-a'-1},$$

where

$$d' - 1 = a'(q-1) + b' \quad 0 \leq b' < q - 1.$$

The integer a' satisfies

$$a' = \left\lfloor \frac{d'-1}{q-1} \right\rfloor = \left\lfloor \frac{d-s}{s(q-1)} \right\rfloor \leq \left\lfloor \frac{d-2}{2(q-1)} \right\rfloor.$$

In any case,

$$\#V_q(P) < \frac{q^{n+1}-1}{q-1} - 2q^{n-\lfloor \frac{d-2}{2(q-1)} \rfloor - 1}.$$

If $a = 0$ then $a' = 0$ and

$$\#V_q(P) < \frac{q^{n+1}-1}{q-1} - (q - d' + 1)q^{n-1} = \frac{q^n - 1}{q-1} + \left(\frac{d}{s} - 1\right) q^{n-1},$$

$$\#V_q(P) < \frac{q^n - 1}{q-1} + \left(\frac{d}{2} - 1\right) q^{n-1}.$$

$$\#V_q(P) < \frac{q^{n+1}-1}{q-1} - 2q^{n-\lfloor \frac{d-2}{2(q-1)} \rfloor - 1}.$$ □

References

1. J.-P. Cherdieu and R. Rolland. On the number of points of some hypersurfaces in $\mathbb{F}_q^n$. *Finite Field and their Applications*, 2:214–224, 1996.
2. P. Delsarte, J.M. Goethals, and F.J. MacWilliams. On generalized reed-muller codes and their relatives. *Information and Control*, 16:403–442, 1970.
3. T. Kasami, S. Lin, and W. Peterson. New generalizations of the reed-muller codes part i: primitive codes. *IEEE Transactions on Information Theory*, IT-14(2):189–199, March 1968.
4. G. Lachaud. Projective reed-muller codes. In *Coding Theory and Applications*, number 311 in Lecture Notes in Computer Science, pages 125–129. Springer-Verlag, 1988.
5. A. Sboui. Second highest number of points of hypersurfaces in $\mathbb{F}_q^n$. *Finite Fields and Their Applications*, 13(3):444–449, July 2007.
6. A.B. Sørensen. A note on algorithms deciding rationality and absolutely irreducibility based on the number of rational solutions. *RISC-Linz Series*, 91-37.0, August 1991.
7. A.B. Sørensen. Projective reed-muller codes. *Transactions on Information Theory*, IT-37(6):1567–1576, 1991.

Neuberg cubics over finite fields

N. J. Wildberger

School of Mathematics and Statistics
UNSW Sydney 2052 Australia
E-mail : n.wildberger@unsw.edu.au

The framework of universal geometry allows us to consider metrical properties of affine views of elliptic curves, even over finite fields. We show how the Neuberg cubic of triangle geometry extends to the finite field situation and provides interesting potential invariants for elliptic curves, focussing on an explicit example over $\mathbb{F}_{23}$. We also prove that tangent conics for a Weierstrass cubic are identical or disjoint.

1. Metrical views of cubics

This paper looks at the connection between modern Euclidean triangle geometry and the arithmetic of elliptic curves over finite fields using the framework of universal geometry (see [8]), a metrical view of algebraic geometry based on the algebraic notions of *quadrance* and *spread* rather than *distance* and *angle*. A good part of triangle geometry appears to extend to finite fields. In particular, the Neuberg cubic provides a rich organizational structure for many triangle centers and associated lines through the group law and related Desmic (linking) structure, even in a finite field. It and other triangle cubics have the potential to be useful geometrical tools for understanding elliptic curves. See [1], [2], [3], [4], [5] and [6] for background on triangle cubics.

For triangle geometers, finite fields hold potential applications to cryptography and also provide a laboratory for exploration that in many ways is more pleasant than the decimal numbers. A price to be paid, however, is that the usual tri-linear coordinate framework needs to be replaced by Cartesian or barycentric coordinates.

We begin with a brief review of the relevant notions from rational trigonometry, which allows the set-up of metrical algebraic geometry. Then we discuss the Neuberg cubic of a triangle and related centers, illustrated

in a particular example over $\mathbb{F}_{23}$. We also prove that for affine cubics in Weierstrass form the tangent conics are all disjoint provided -3 is not a square in the field.

It should be noted that there is also a projective version of universal geometry (see [9]), but here we stick to the affine situation.

2. Laws of Rational Trigonometry

Fix a finite field $\mathbb{F}$ not of characteristic two, whose elements are called **numbers**. A **point** A is an ordered pair $[x, y]$ of numbers, that is an element of $\mathbb{F}^2$. The **quadrance** $Q(A_1, A_2)$ between points $A_1 \equiv [x_1, y_1]$ and $A_2 \equiv [x_2, y_2]$ is the number

$$Q(A_1, A_2) \equiv (x_2 - x_1)^2 + (y_2 - y_1)^2 .$$

A **line** l is an ordered proportion $\langle a : b : c \rangle$, where a and b are not both zero. This represents the equation $ax + by + c = 0$. Such a line is **null** precisely when

$$a^2 + b^2 = 0.$$

Null lines occur precisely when -1 is a square. For distinct points $A_1 = [x_1, y_1]$ and $A_2 = [x_2, y_2]$ the line passing through them both is

$$A_1A_2 = \langle y_1 - y_2 : x_2 - x_1 : x_1y_2 - x_2y_1 \rangle .$$

Two lines $l_1 \equiv \langle a_1 : b_1 : c_1 \rangle$ and $l_2 \equiv \langle a_2 : b_2 : c_2 \rangle$ are **perpendicular** precisely when

$$a_1a_2 + b_1b_2 = 0.$$

For any fixed line l and any point A, there is a unique line n passing through A and perpendicular to l, called the **altitude** from A to l. If l is a non-null line then the altitude n meets l at a unique point F, called the **foot** of the altitude. In this case we may define the **reflection of** A **in** l to be the point $\sigma_l(A)$ such that F is the midpoint of the side $\overline{A\sigma_l(A)}$. If m is another line, then the **reflection of** m **in** l is the line $\Sigma_l(m)$ with the property that the reflection in l of any point A on m lies on $\Sigma_l(m)$.

The **spread** $s(l_1, l_2)$ between non-null lines $l_1 \equiv \langle a_1 : b_1 : c_1 \rangle$ and $l_2 \equiv \langle a_2 : b_2 : c_2 \rangle$ is the number

$$s(l_1, l_2) \equiv \frac{(a_1b_2 - a_2b_1)^2}{(a_1^2 + b_1^2)(a_2^2 + b_2^2)} = 1 - \frac{(a_1a_2 + b_1b_2)^2}{(a_1^2 + b_1^2)(a_2^2 + b_2^2)}.$$

This number $s = s(l_1, l_2)$ is 0 precisely when the lines are parallel, and 1 precisely when the lines are perpendicular. It has the property that $s(1 - s)$

is a square in the field, and every such **spread number** s can be shown to be the spread between some two lines.

The spread between lines may alternatively be expressed as a ratio of quadrances: if l_1 and l_2 intersect at a point A, choose any other point B on l_1, and let C on l_2 be the foot of the altitude line from B to l_2, then

$$s(l_1, l_2) = \frac{Q(B, C)}{Q(A, B)}.$$

Reflection in a line preserves quadrance between points and spread between lines. Given three distinct points A_1, A_2 and A_3, we use the notation

$$Q_1 \equiv Q(A_2, A_3) \qquad Q_2 \equiv Q(A_1, A_3) \qquad Q_3 \equiv Q(A_1, A_2)$$

and

$$s_1 \equiv s(A_1A_2, A_1A_3) \qquad s_2 \equiv s(A_2A_1, A_2A_3) \qquad s_3 \equiv s(A_3A_1, A_3A_2).$$

A **triangle** $\overline{A_1A_2A_3}$ is a set of three non-collinear points, and is **non-null** precisely when its three lines A_1A_2, A_2A_3 and A_1A_3 are non-null. Here are the five main laws of rational trigonometry, which may be viewed as purely algebraic identities involving only rational functions.

Triple quad formula The points A_1, A_2 and A_3 are collinear precisely when

$$(Q_1 + Q_2 + Q_3)^2 = 2\left(Q_1^2 + Q_2^2 + Q_3^2\right).$$

Pythagoras' theorem The lines A_1A_3 and A_2A_3 are perpendicular precisely when

$$Q_1 + Q_2 = Q_3.$$

Spread law For a non-null triangle $\overline{A_1A_2A_3}$

$$\frac{s_1}{Q_1} = \frac{s_2}{Q_2} = \frac{s_3}{Q_3}.$$

Cross law For a non-null triangle $\overline{A_1A_2A_3}$ define the **cross** $c_3 \equiv 1 - s_3$. Then

$$(Q_1 + Q_2 - Q_3)^2 = 4Q_1Q_2c_3.$$

Triple spread formula For a non-null triangle $\overline{A_1A_2A_3}$

$$(s_1 + s_2 + s_3)^2 = 2\left(s_1^2 + s_2^2 + s_3^2\right) + 4s_1s_2s_3.$$

See [8] for proofs, and many more facts about geometry in such a purely algebraic setting.

3. Neuberg cubics

Many interesting points, lines, circles, parabolas, hyperbolas and cubics have been associated to a triangle in the plane, such as the centroid G, orthocenter O, circumcenter C, incenter I, Euler line e (which passes through O, G and C), nine-point circle and so on. Perhaps the most remarkable of these is the *Neuberg cubic*, which in the earlier literature was called the 32 point cubic, but these days is known to pass through many more triangle centers (see [4], [5], [6]). Most such objects depend crucially on a metrical structure on the affine plane.

Figure 1 shows the Neuberg cubic for the triangle $\overline{A_1A_2A_3}$ with vertices $A_1 = [0, 0]$, $A_2 = [1, 0]$ and $A_3 = [3/4, 3/4]$. For this triangle the Euler line, which passes through the orthocenter O and the circumcenter C, is horizontal. Various incenters I_i are shown, as well as reflections of the vertices in the sides. These are just a few of the many points on the Neuberg cubic. Note that the tangents to the four incenters are also horizontal, and it turns out that the asymptote of the cubic is also.

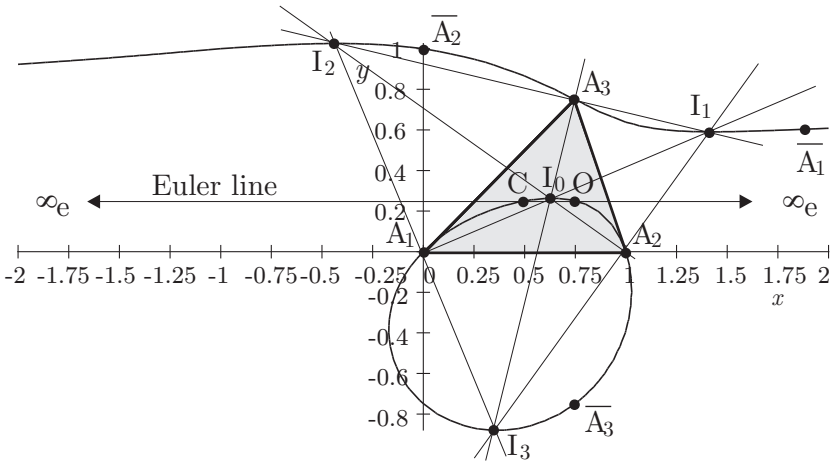


Fig. 1. Neuberg cubic for $A_1 = [0, 0]$, $A_2 = [1, 0]$ and $A_3 = [3/4, 3/4]$

With the completely algebraic language of rational trigonometry, we consider such a picture for *triangles in finite fields*. For this it will be convenient to alter the usual point of view somewhat, elevating the quadrangle $\overline{I_0I_1I_2I_3}$ of *incenters* to a primary position. This ensures that the reference triangle $\overline{A_1A_2A_3}$ actually has vertex bisectors.

For a triangle $\overline{I_1I_2I_3}$ the altitudes from each point to the opposite side intersect at the orthocenter, which we here denote I_0. In terms of Cartesian coordinates $I_j = [x_j,y_j]$,

$$x_0 \equiv \frac{\begin{pmatrix} x_1x_2y_2 - x_1x_3y_3 + x_2x_3y_3 - x_3x_2y_2 + x_3x_1y_1 - x_2x_1y_1 \\ +y_1y_2^2 - y_1y_3^2 + y_2y_3^2 - y_3y_2^2 + y_3y_1^2 - y_2y_1^2 \end{pmatrix}}{x_1y_2 - x_1y_3 + x_2y_3 - x_3y_2 + x_3y_1 - x_2y_1}$$

and

$$y_0 \equiv \frac{\begin{pmatrix} x_1y_1y_2 - x_1y_1y_3 + x_2y_2y_3 - x_3y_3y_2 + x_3y_3y_1 - x_2y_2y_1 \\ +x_1^2x_2 - x_1^2x_3 + x_2^2x_3 - x_3^2x_2 + x_3^2x_1 - x_2^2x_1 \end{pmatrix}}{x_1y_2 - x_1y_3 + x_2y_3 - x_3y_2 + x_3y_1 - x_2y_1}.$$

In terms of barycentric coordinates, if the quadrances of the triangle $\overline{I_1I_2I_3}$ are R_1, R_2 and R_3 and

$$\begin{aligned} \mathcal{A} &= (R_1 + R_2 + R_3)^2 - 2\left(R_1^2 + R_2^2 + R_3^3\right) \\ &= 4\left(x_1y_2 - x_1y_3 + x_2y_3 - x_3y_2 + x_3y_1 - x_2y_1\right)^2 \end{aligned}$$

is the **quadrea** of the triangle (sixteen times the square of the area in the decimal number situation), then $I_0 = \beta_1 I_1 + \beta_2 I_2 + \beta_3 I_3$ where

$$\begin{aligned} \beta_1 &\equiv (R_3 + R_1 - R_2)(R_1 + R_2 - R_3)/\mathcal{A} \\ \beta_2 &\equiv (R_1 + R_2 - R_3)(R_2 + R_3 - R_1)/\mathcal{A} \\ \beta_3 &\equiv (R_2 + R_3 - R_1)(R_3 + R_1 - R_2)/\mathcal{A}. \end{aligned}$$

If $\overline{I_1I_2I_3}$ is non-null, which we henceforth assume, then the feet of its altitudes exist and we call them respectively A_1, A_2 and A_3. Thus for example I_1, I_0 and A_1 are collinear points which lie on a line perpendicular to I_2I_3. In the quadrangle $\overline{I_1I_2I_3I_4}$ we have a complete symmetry between the four points I_0, I_1, I_2 and I_3. So we could have started with any three of these points, and the orthocenter of such a triangle would have been the fourth point, with the **orthic triangle** $\overline{A_1A_2A_3}$ obtained always the same.

Theorem 3.1 (Orthic triangle). *The lines I_0I_1 and I_2I_3 are bisectors of the vertex of $\overline{A_1A_2A_3}$ at A_1, in the sense that*

$$\begin{aligned} s(I_0I_1, A_1A_2) &= s(I_0I_1, A_1A_3) \\ s(I_2I_3, A_1A_2) &= s(I_2I_3, A_1A_3). \end{aligned}$$

The proof uses a computer, and one can further verify that the former spread is

$$\frac{\left(x_2x_3 - x_1x_3 - x_1x_2 - y_1y_2 - y_1y_3 + y_2y_3 + x_1^2 + y_1^2\right)^2}{R_2R_3}$$

while the latter spread is

$$\frac{(x_1y_2 - x_1y_3 + x_2y_3 - x_3y_2 + x_3y_1 - x_2y_1 + x_3y_2)^2}{R_2R_3}.$$

These two spreads sum to 1, as they must since I_0I_1 and I_2I_3 are perpendicular. So the triangle $\overline{A_1A_2A_3}$ has a special property: each of its vertices has bisectors. Over the decimal numbers, every triangle has vertex bisectors, but in [8] it is shown that in general this amounts to the condition that the spreads of the triangle are *squares*.

For a point P, let P_1, P_2 and P_3 denote its reflections in the sides A_2A_3, A_1A_3 and A_1A_2 respectively, and define the **Neuberg cubic** N_c of $\overline{A_1A_2A_3}$ to be the locus of points P such that P_1, P_2 and P_3 are perspective with A_1, A_2 and A_3 respectively: in other words that P_1A_1, P_2A_2 and P_3A_3 are concurrent lines.

4. An example over $\mathbb{F}_{23}$

We work in the prime field $\mathbb{F}_{23}$, in which the squares are $1, 4, 9, 16, 2, 13, 3, 18, 12, 8$ and 6. Note that -1 is not a square, but that $3 = 7^2$ is a square. The latter fact implies that equilateral triangles exist in $\mathbb{F}_{23}^2$. Let

$$I_1 = [6, 4] \qquad I_2 = [22, 22] \qquad I_3 = [21, 12].$$

These points have been chosen so that the orthocenter of $\overline{I_1I_2I_3}$ is $I_0 = [0, 0]$. The feet of the altitudes are

$$A_1 = [13, 1] \qquad A_2 = [5, 5] \qquad A_3 = [2, 11].$$

The lines of $\overline{A_1A_2A_3}$ are

$$A_1A_2 = \langle 3 : 6 : 1\rangle \qquad A_2A_3 = \langle 6 : 3 : 1\rangle \qquad A_1A_2 = \langle 12 : 4 : 1\rangle$$

and the spreads of the triangle $\overline{A_1A_2A_3}$ are

$$s_1 = 12 \qquad s_2 = 16 \qquad s_3 = 6.$$

Note that as expected these numbers are squares, and one can check that

$$s(A_1I_0, A_1A_2) = s(A_1I_0, A_1A_3) = 5$$
$$s(A_2I_0, A_2A_1) = s(A_2I_0, A_2A_3) = -6$$
$$s(A_3I_0, A_3A_1) = s(A_3I_0, A_3A_2) = -7.$$

The connection between for example the spread $s = 5$ and the spread of its 'double' $r = 12$ is given by the *second spread polynomial*,

$$r = S_2(s) = 4s(1 - s)$$

which in chaos theory is known as the *logistic map*. The spread polynomials have many remarkable properties that hold also over finite fields, see [8].

We need the following formula for a reflection.

Theorem 4.1 (Reflection of a point in a line). *If $l \equiv \langle a : b : c \rangle$ is a non-null line and $A \equiv [x, y]$, then*

$$\sigma_l(A) = \left[\frac{\left(b^2 - a^2\right) x - 2aby - 2ac}{a^2 + b^2}, \frac{-2abx + \left(a^2 - b^2\right) y - 2bc}{a^2 + b^2} \right].$$

Using this, the reflections of $P = [x, y]$ in the sides of $\overline{A_1A_2A_3}$ are

$$\begin{aligned} P_1 &= [4x + 13y + 12, 13x + 19y + 6] \\ P_2 &= [13x + 4y + 1, 4x + 10y + 8] \\ P_3 &= [19x + 13y + 6, 13x + 4y + 12]. \end{aligned}$$

The lines P_1A_1, P_2A_2 and P_3A_3 are then

$$\begin{aligned} &\langle 13x + 19y + 5 : 19x + 10y + 1 : 19x + 19y + 3 \rangle \\ &\langle 4x + 10y + 3; 10x + 19y + 4 : 22x + 16y + 11 \rangle \\ &\langle 13x + 4y + 1 : 4x + 10y + 19 : 22x + 20y + 19 \rangle \end{aligned}$$

and these are concurrent precisely when

$$\begin{vmatrix} 13x + 19y + 5 & 19x + 10y + 1 & 19x + 19y + 3 \\ 4x + 10y + 3 & 10x + 19y + 4 & 22x + 16y + 11 \\ 13x + 4y + 1 & 4x + 10y + 19 & 22x + 20y + 19 \end{vmatrix} = 0.$$

Expanding gives the Neuberg cubic $\overline{A_1A_2A_3}$: an affine curve over $\mathbb{F}_{23}$ with equation

$$y^3 + x^2y + 22y^2 + 7xy + 9x^2 + 13y = 0. \tag{1}$$

The tangent line to a point $[a, b]$ on the curve has equation

$$x(18a + 7b + 2ab) + y\left(7a + 21b + a^2 + 3b^2 + 13\right) + 3b + 7ab + 9a^2 + 22b^2 = 0.$$

There is another revealing way to obtain the Neuberg cubic.

Theorem 4.2 (Reflection of a line in a line). *The reflection in the non-null line $l \equiv \langle a : b : c \rangle$ sends $\langle a_1 : b_1 : c_1 \rangle$ to*

$$\left\langle \left(a^2 - b^2\right) a_1 + 2abb_1 : 2aba_1 - \left(a^2 - b^2\right) b_1 : 2aca_1 + 2bcb_1 - \left(a^2 + b^2\right) c_1 \right\rangle.$$

In a triangle with vertex bisectors, the reflection of a line through a given vertex in either of the vertex bisectors at that vertex is the same.

Theorem 4.3 (Isogonal conjugates). *If a triangle* $\overline{A_1A_2A_3}$ *has vertex bisectors at each vertex, then for any point* P *the reflections of* A_1P, A_2P *and* A_3P *in the vertex bisectors at* A_1, A_2 *and* A_3 *respectively are concurrent.*

The point of concurrence of these lines is P^*, the **isogonal conjugate** of $P = [x, y]$. The proof again is a calculation using coordinates. We may use the reflection of a line in a line theorem to establish a precise formula for P^* in the special case of our example reference triangle $\overline{A_1A_2A_3}$:

$$P^* = \left[\frac{2x + 22xy + 2x^2 + 17y^2}{4x + 20y + 5x^2 + 5y^2 + 21}, \frac{2y + 15xy + x^2 + 2y^2}{4x + 20y + 5x^2 + 5y^2 + 21}\right].$$

Over the decimal numbers, the Neuberg cubic is *also* the locus of those $P = [x, y]$ such that PP^* is parallel to the Euler line. We can verify this also in our finite example, since this condition amounts to

$$y = \frac{2y + 15xy + x^2 + 2y^2}{4x + 20y + 5x^2 + 5y^2 + 21}$$

which in turn is equivalent to the equation (1) of N_c. It follows that if P lies on N_c, then so does P^*—this is a useful way to obtain new points from old ones. In fact the Euler line is parallel to the tangent to N_c at the infinite point ∞_e.

The cubic (1) is nonsingular, has 27 points lying on it, and its projective extension has one more point at infinity, namely $\infty_e = [1 : 0 : 0]$. Here are all the points, the notation will be explained more fully below:

$[0, 0] = I_0$	$[0, 8] = E_3'$	$[0, 16] = S'$	$[2, 11] = A_3$
$[3, 13] = S = \overline{A_3}$	$[4, 5]$	$[5, 5] = A_2$	$[5, 14] = \infty_e^*$
$[6, 4] = I_1$	$[7, 1]$	$[7, 2] = E_2'$	$[7, 21] = O$
$[8, 10] = \overline{A_1} = E_2$	$[13, 1] = A_1$	$[13, 7]$	$[13, 16] = F'$
$[14, 9]$	$[16, 11]$	$[17, 7] = E_1'$	$[17, 8]$
$[17, 9] = \overline{A_2} = E_1$	$[18, 21] = C = E_3$	$[19, 13] = F$	$[21, 2]$
$[21, 10]$	$[21, 12] = I_3$	$[22, 22] = I_2$	$[1 : 0 : 0] = \infty_e$

To define the group structure on the cubic, define $X \star Y$ to be the (third) intersection of the line XY with the (projective) cubic, so that for example

$$I_0 \star I_0 = \infty_e.$$

We choose the base point of the group structure to be the point I_0. Then define

$$X \cdot Y = (X \star Y) \star I_0$$

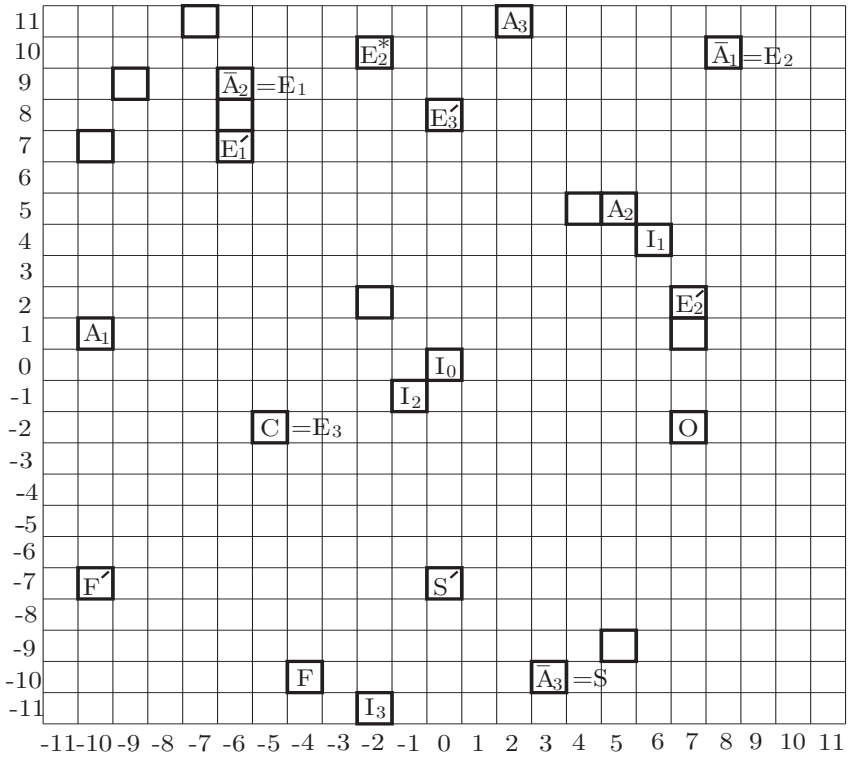


Fig. 2. The Neuberg cubic for $A_1 = [13, 1]$, $A_2 = [5, 5,]$ and $A_3 = [2, 11]$

with inverse

$$X^{-1} = X \star (I_0 \star I_0) = X \star \infty_e = X^*.$$

So we are writing the group multiplicatively, and note that I_0 is not a flex, so that X, Y and Z collinear is equivalent to $X \cdot Y \cdot Z = \infty_e$, not $X \cdot Y \cdot Z = I_0$. Furthermore X is of order two when $X \star X = \infty_e$ and X is not I_0, which happens when X is I_1, I_2 or I_3, and the four incenters form a Klein 4-group.

The triangle $\overline{A_1A_2A_3}$ can be recovered from the Neuberg cubic by first finding the asymptote (tangent to the point at infinity), then finding the four points $\overline{I_0I_1I_2I_3}$ on the cubic with a tangent parallel to this asymptote, and then taking the orthic triangle of any three of them. This can all be done algebraically using the group law, since $I_j \star I_j = \infty_e$ and $A_1 = I_2 \star I_3$ etc.

5. Points on the Neuberg cubic

Not only is the Neuberg cubic defined metrically, but it also has many points on it that are metrical in nature. More than a hundred are known, we will illustrate some of these for our example. The Neuberg cubic N_c of $\overline{A_1A_2A_3}$ first of all passes through A_1, A_2 and A_3. It also passes through the reflections of these points in the sides of the triangle, in this case

$$\overline{A_1} = [8, 10] \qquad \overline{A_2} = [17, 9] \qquad \overline{A_3} = [3, 13]\,.$$

It also passes through the four incenters I_0, I_1, I_2 and I_3 of $\overline{A_1A_2A_3}$. In a general field there is no notion of 'interior point', so these four incenters should be regarded symmetrically. The Neuberg cubic passes through the orthocenter $O = [7, 21]$ and the circumcenter $C = [18, 21]$ of $\overline{A_1A_2A_3}$, and these are isogonal conjugates, that is

$$O^* = C.$$

The line OC is the Euler line e of $\overline{A_1A_2A_3}$ and it has equation $y = 21$, so that it is horizontal and passes through the infinite point $\infty_e = [1 : 0 : 0]$. Note that

$$\infty_e^* = [5, 14]\,.$$

Since $3 = 7^2$ is a square, on any side of $\overline{A_1A_2A_3}$ we may create two equilateral triangles, for example on the side containing $A_1 = [13, 1]$ and $A_3 = [2, 11]$ we can choose a third point

$$[13 + 2, 1 + 11]/2 \pm \frac{7}{2}[1 - 11, 2 - 13]$$

namely

$$E_2 = [8, 10] \qquad \text{or} \qquad E_2' = [7, 2]\,.$$

Thus $\overline{A_1A_3E_2}$ and $\overline{A_1A_3E_2'}$ are equilateral triangles with

$$Q(A_1, A_3) = Q(A_1, E_2) = Q(A_3, E_2) = Q(A_1, E_2') = Q(A_3, E_2') = 14.$$

As opposed to the case over the decimal numbers, there seems to be no obvious notion of these triangles being either 'exterior' or 'interior' to $\overline{A_1A_2A_3}$. Using all three sides gives the six points

$$\begin{array}{lll} E_1 = [17, 9] & E_2 = [8, 10] & E_3 = [18, 21] \\ E_1' = [13, 7] & E_2' = [7, 2] & E_3' = [0, 8] \end{array}$$

and their isogonal conjugates

$$\begin{array}{lll} E_1^* = [14, 9] & E_2^* = [21, 10] & E_3^* = [7, 21] \\ E_1'^* = [17, 7] & E_2'^* = [21, 2] & E_3'^* = [17, 8]\,. \end{array}$$

All twelve of these points lie on the Neuberg cubic. Yet the centroids of the six equilateral triangles thus formed are

$$\begin{array}{lll} G_1 = [8, 16] & G_2 = [0, 15] & G_3 = [12, 9] \\ G_1' = [22, 0] & G_2' = [15, 20] & G_3' = [6, 20] \end{array}$$

and you may check that

$$\begin{aligned} Q(G_1, G_2) = Q(G_2, G_3) = Q(G_1, G_3) = 19 \\ Q(G_1', G_2') = Q(G_2', G_3') = Q(G_1', G_3') = 12 \end{aligned}$$

so that Napolean's theorem that the centroids of both 'external' and 'internal' equilateral triangles themselves form an equilateral triangle seems to hold. It seems curious that the six points E_i and E_j' are thereby divided naturally into two groups.

The Fermat points of a triangle may be defined over the decimal numbers as the perspectors of the 'external and internal equilateral triangles', and with the above interpretation, these points exist also in this field. There is another approach to their definition. The vertex bisectors at A_1 of $\overline{A_1A_2A_3}$ intersect A_2A_3 at the points $X_1 = [20, 21]$ and $Y_1 = [12, 14]$. The circle through these points with center the midpoint of $\overline{X_1Y_1}$ has equation $(x-16)^2 + (y-6)^2 = 11$ and is called an **Apollonius circle** of $\overline{A_1A_2A_3}$. There is also such a circle starting with A_2, with equation $(x-1)^2 + (y-14)^2 = 5$ and one starting with A_3, with equation $(x-4)^2 + (y-17)^2 = 17$. These three Appollonius circles intersect at two points, called the **isodynamic points** of $\overline{A_1A_2A_3}$, given by

$$S = [3, 13] \qquad \text{and} \qquad S' = [0, 16].$$

The Neuberg cubic passes through the two isodynamic points. The centres of the three Appollonius circles are collinear, and lie on the **Lemoine line** with equation $16x + 7y + 1 = 0$.

The isogonal conjugates of the isodynamic points are the **Fermat points**

$$F = S^* = [19, 13] \qquad \text{and} \qquad F' = (S')^* = [13, 16].$$

It may be checked that F is also the centre of perspectivity between $\overline{A_1A_2A_3}$ and $\overline{E_1E_2E_3}$, while F' is the centre of perspectivity between $\overline{A_1A_2A_3}$ and $\overline{E_1'E_2'E_3'}$. It may be remarked that in the decimal number plane, the Fermat points also have an interpretation in terms of minimizing the sum of the distances to the vertices of the triangle, but this kind of statement cannot be expected to have a simple analog in universal geometry.

The **Brocard line** with equation $10x + 10y + 1 = 0$ passes through the circumcenter $C = [18, 21]$, the **symmedian point** $K = G^* = [10, 6]$ and the two isodynamic points S and S'. It is perpendicular to the Lemoine line.

6. Quadrangles and Desmic structure

Elliptic curves naturally give rise to interesting configurations of 12 points and 16 lines, called by John Conway *Desmic (or linking) structure,* where each line passes through three points and each point lies on four lines (see [7]). To describe this situation, begin with a triangle ABC and two generic points P and Q. Then define

$$\begin{array}{lll} A' = (BP)\,(CQ) & B' = (CP)\,(AQ) & C' = (AP)\,(BQ) \\ A'' = (BQ)\,(CP) & B'' = (CQ)\,(AP) & C'' = (AQ)\,(BP) \end{array}$$

This insures that $\overline{ABC}$ and $\overline{A'B'C'}$ are perspective from some perspector D'', that $\overline{A'B'C'}$ and $\overline{A''B''C''}$ are perspective from some perspector D and that $\overline{A''B''C''}$ and $\overline{ABC}$ are perspective from some perspector D'. Furthermore the points D, D' and D'' are collinear.

Put another way, two triangles which are doubly perspective are triply perspective (essentially a consequence of Pappus' theorem). The various

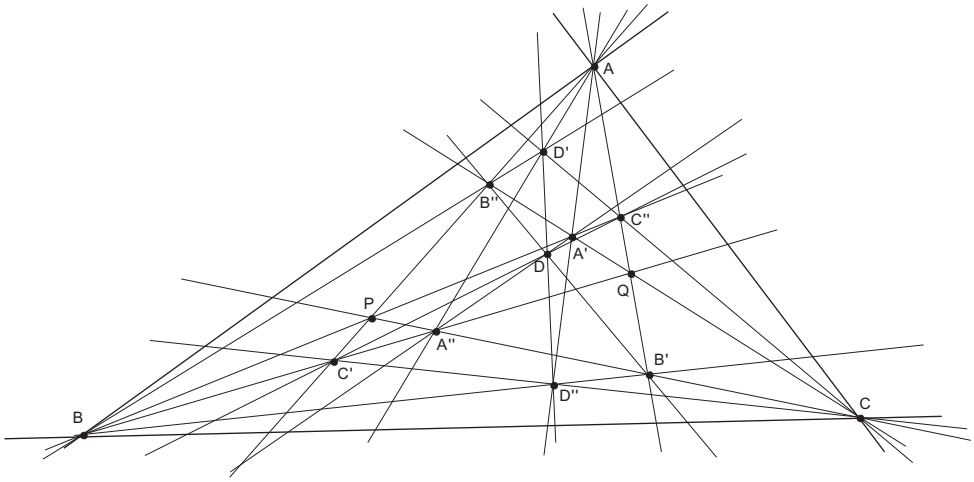


Fig. 3. Desmic $16 - 12$ structure

collinearities can be recorded in terms of the array

A	B	C	D
A'	B'	C'	D'
A''	B''	C''	D''

A triple of points from the array is collinear precisely when a) each is from a different row, and b) if no D is involved each is from a different column, and c) if a D is involved, then the other two are both from the same column. An example of the former would be A, B'' and C', or C, B' and A''. An example of the latter would be D', B and B'' or D, D' and D''. There are then exactly 16 such collinearities among these 12 points.

Given a point P on a cubic, there are in general four points X_1, X_2, X_3 and X_4 on the cubic, other than P, whose tangents pass through P. Call these four points a **quadrangle** of the Neuberg cubic, and more specifically the **quadrangle to** P. To illustrate this, we write $X_1, X_2, X_3, X_4 : P$.

Here are some quadrangles for our cubic:

$$A_1, A_2, A_3, \infty_e : \infty_e^*$$
$$I_1, I_2, I_3, I_o : \infty_e$$
$$\overline{A_1}, \overline{A_2}, \overline{A_3}, C : [0, 8]$$
$$\overline{A_1}^*, \overline{A_2}^*, \overline{A_3}^*, O : [17, 8]$$

Given three collinear points on a cubic, the associated quadrangles form a Desmic structure. Here are some examples for our cubic

A_1	A_2	A_3	∞_e
I_1	I_2	I_3	I_0
I_1	I_2	I_3	I_0

A_1	A_2	A_3	∞_e
E_1	E_2	E_3	$[3, 13]$
E_1^*	E_2^*	E_3^*	$[19, 13]$

A_1	A_2	A_3	∞_e
E_1'	E_2'	E_3'	$[0, 16]$
$E_1'^*$	$E_2'^*$	$E_3'^*$	$[13, 16]$.

We see that having recognized an (affine) cubic curve as a Neuberg cubic of a triangle, we have lots of natural and deep geometry that connects to the group multiplication. A natural question is: given an elliptic curve can

we find an affine view of it which is a Neuberg cubic? And if so, how can we classify such views, and use them to understand elliptic curves?

Such an approach ought to be especially useful in the convenient laboratory provided by finite fields.

7. Tangent conics to an affine cubic

Here is a quite different use of affine coordinates in the study of an elliptic curve. For more information and examples involving tangent conics, see [8]. Different metrical interpretations of tangent conics thus allows one to distinguish points on an affine curve from the nature of the tangent conic. Figure 4 gives a view of some tangent conics to $[x_0, y_0]$ for the curve $y^2 = x^3 - x$ over the decimal numbers. Tangent conics to points on the 'egg' are ellipses, while others are either hyperbolas opening horizontally, a pair of lines, or hyperbolas opening vertically depending respectively on whether x_0 is less than, equal to, or greater than $\sqrt{\frac{2}{3}\sqrt{3}+1}$. Rather remarkably, these tangent conics nowhere intersect.

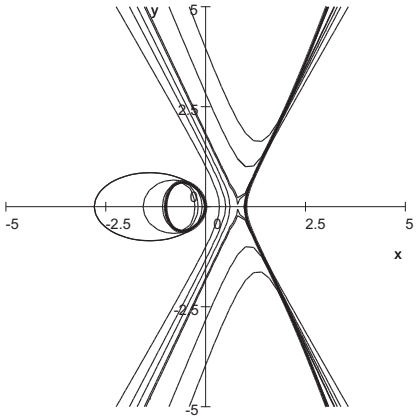


Fig. 4. Tangent conics to $y^2 = x^3 - x$

Theorem 7.1. *Over a field in which* -3 *is not a square, any two tangent conics of the affine curve* $y^2 = ax^3 + bx + c$ *are either identical or disjoint.*

Proof. Recall that the tangent conic to an affine curve is the part of the Taylor expansion of degree two or less (see [8, Chapter 19]). More specifically, to find the tangent conic to $y^2 = ax^3 + bx + c$ at a point $A = [x_0, y_0]$

on it, first translate the curve by $-A$, yielding the equation

$$(y+y_0)^2 = a(x+x_0)^3 + b(x+x_0) + c$$

or, after simplification using the fact that A lies on the original curve,

$$y^2 + 2yy_0 - ax^3 - 3ax^2x_0 + x\left(-b - 3ax_0^2\right) = 0.$$

Then take the quadratic part:

$$y^2 + 2yy_0 - 3ax^2x_0 + x\left(-b - 3ax_0^2\right) = 0$$

and translate this conic back by A, yielding

$$(y-y_0)^2 + 2(y-y_0)y_0 - 3a(x-x_0)^2x_0 + (x-x_0)\left(-b - 3ax_0^2\right) = 0$$

or after simplification

$$y^2 - 3ax^2x_0 + x\left(3ax_0^2 - b\right) - ax_0^3 - c = 0.$$

To find the intersection between two such tangent conics

$$y^2 - 3ax^2x_0 + x\left(3ax_0^2 - b\right) - ax_0^3 - c = 0$$
$$y^2 - 3ax^2x_1 + x\left(3ax_1^2 - b\right) - ax_1^3 - c = 0$$

take the difference between the two equations, which factors as

$$a(x_1 - x_0)\left(3x^2 - 3x(x_0+x_1) + x_0^2 + x_0x_1 + x_1^2\right).$$

If $x_0 = x_1$ then the tangent conics coincide. Otherwise we get an intersection when the second quadratic factor has a zero. But its discriminant is

$$9(x_0+x_1)^2 - 4 \times 3\left(x_0^2 + x_0x_1 + x_1^2\right) = (-3)(x_1 - x_0)^2$$

and so if -3 is not a square then there is no solution and so the tangent conics are disjoint. □

Note that the equation of the tangent conic

$$y^2 - 3ax^2x_0 + x\left(3ax_0^2 - b\right) - ax_0^3 - c = 0.$$

can be rewritten as

$$y^2 - \left(ax^3 + bx + c\right) + a(x-x_0)^3 = 0.$$

Figure 5 gives a view of some tangent conics for the curve $y^2 = x^3 + x$. The tangent conic to $[0,0]$ is a parabola, and otherwise they are either hyperbolas opening horizontally, a pair of lines, or hyperbolas opening vertically depending respectively on whether x_0 is less than, equal to, or greater than $\frac{1}{3}\sqrt{3}\sqrt{2\sqrt{3}-3}$.

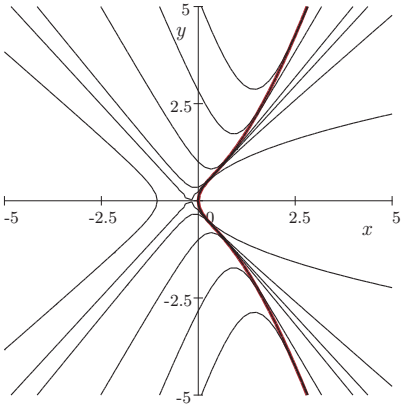


Figure 5: Tangent conics to $y^2 = x^3 + x$

Figure 6 shows the nodal cubic $y^2 = x^3 + x^2$ with tangent conic at $[x_0, y_0]$ given by

$$y^2 + 3xx_0^2 + x^2(-3x_0 - 1) - x_0^3 = 0.$$

We get a parabola when $x_0 = -1/3$, ellipses for x_0 less than that, hyperbolas opening horizontally till $x_0 = 0$, when we get the pair of lines $y = \pm x$, then hyperbolas opening upwards.

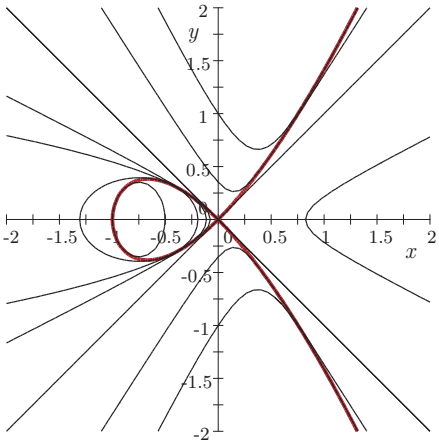


Figure 6: Tangent conics to $y^2 = x^3 + x^2$

References

1. H. M. Cundy and C. F. Parry, Geometrical properties of some Euler and circular cubics, Part 1, *J. Geom.* 66 (1999), 72-103.

2. H. M. Cundy and C. F. Parry, Geometrical properties of some Euler and circular cubics, Part 2, *J. Geom.* **68** (2000), 58-75.
3. H. M. Cundy and C. F. Parry, Some cubic curves associated with a triangle, *J. Geom.* 53 (1995), 41-66.
4. J-P. Ehrmann and B. Gibert, *Special Isocubics in the Triangle Plane*, available from http://perso.wanadoo.fr/bernard.gibert/downloads.html.
5. C. Kimberling, *Triangle Centers and Central Triangles*, vol **129** of Congressus Numerantium, Utilitas Mathematica Publishing Inc, Winnipeg Manitoba, 1998.
6. C. Kimberling, *Encyclopedia of Triangle Centers*, available at http://faculty.evansville.edu/ck6/encyclopedia/ETC.html.
7. W. Stothers, 'Grassmann cubics and Desmic structures', Forum Geometricorum **6** (2006) 117-138.
8. N. J. Wildberger, *Divine Proportions: Rational Trigonometry to Universal Geometry*, Wild Egg Books (http://wildegg.com), Sydney, 2005.
9. N. J. Wildberger, 'Affine and projective universal geometry', arXiv:math.MG/0612499v1, 18 Dec., 2006.

Partitions of Vector Spaces over Finite Fields

Yevhen Zelenyuk

School of Mathematics,
University of the Witwatersrand,
Private Bag 3, Wits 2050, South Africa
E-Mail: Yevhen.Zelenyuk@wits.ac.za
http://www.maths.wits.ac.za/yzelenyuk.html

Let V be an infinite vector space over a finite field F and let $F \neq \mathbb{F}_2$. We construct a partition $\{A_m : m < \omega\}$ of V such that for every infinite affine subspace W of V and $m < \omega$, one has $W \cap A_m \neq \emptyset$.

Keywords: Vector space, finite field, partition

Let V be an infinite vector space over a finite field F. In 1971, R. Graham and B. Rothschild [1] (see also [2, Section 2.4]) proved the following result.

Theorem 1. *Whenever V is partitioned into finitely many cells, there are arbitrarily large finite affine subspaces contained in one cell.*

In case $F = \mathbb{F}_2$, a stronger statement holds:

Theorem 2. *Let $F = \mathbb{F}_2$. Then whenever V is partitioned into finitely many cells, there is an infinite affine subspace contained in one cell.*

Theorem 2 is an immediate consequence of the Finite Sums Theorem [3] (see also [4, Section 5.2]) which can be stated as follows:

Theorem 3. *Whenever V is partitioned into finitely many cells, there is a one-to-one sequence $(x_n)_{n<\omega}$ in V with $FS((x_n)_{n<\omega})$ contained in one cell, where*

$$FS((x_n)_{n<\omega}) = \{\sum_{n\in H} x_n : \emptyset \neq H \in [\omega]^{<\omega}\}$$

and $[X]^{<\omega}$ is the set of finite subsets of the set X.

Now let $F \neq \mathbb{F}_2$ and suppose that V is countable. Then there is a 2-partition of V such that every infinite affine subspace meets each cell of the partition.

Example 4. Represent V as $\bigoplus_\omega F$. Take any partition $\{F_0, F_1\}$ of $F \setminus \{0\}$ into two nonempty subsets and define the partition $\{A_0, A_1\}$ of $V \setminus \{0\}$ by

$$A_i = \{x : \text{ the last nonzero coordinate of } x \text{ belongs to } F_i\}.$$

We claim that, for every infinite affine subspace W of V and $i < 2$, one has $W \cap A_i \neq \emptyset$. Indeed, let $g \in W$ and $U = W - g$. Then U is an infinite vector subspace of V and $W = g + U$. Pick any $x \in U$ such that the number of the last nonzero coordinate of x is greater than that of g. Choose $\varepsilon \in F \setminus \{0\}$ such that the last nonzero coordinate of εx belongs to F_i. Then $g + \varepsilon x \in W \cap A_i$.

The next example shows that there is even an ω-partition of V such that every infinite affine subspace meets each cell of the partition.

Example 5. Let $V = \bigoplus_\omega F$. For every $x \in V \setminus \{0\}$, consider the sequence of nonzero coordinates of x and define $\nu(x)$ to be the number of pairs of distinct neighbouring elements in this sequence. Define the partition $\{A_m : m < \omega\}$ of $V \setminus \{0\}$ by

$$A_m = \{x \in A : \nu(x) \equiv 2^m \mod 2^{m+1}\}.$$

Equivalently, A_m consists of all $x \in V \setminus \{0\}$ such that the index of the leftmost nonzero digit in the binary expansion of $\nu(x)$ is m.

Let U be an infinite vector subspace of V, $g \in V$ and $W = g + U$. We claim that $W \cap A_m \neq \emptyset$ for each $m < \omega$. To see this, let $k = 2^{m+1} - 1$ and choose a sequence $(x_i)_{i \leq k}$ in U such that

(1) the number of the last nonzero coordinate of x_0 is greater than that of g, and

(2) for each $i < k$, the number of the last nonzero coordinate of x_i is less than the number of the first nonzero coordinate of x_{i+1}.

Put $\varepsilon_0 = 1$ and, by induction on $i = 1, \ldots, k$, choose $\varepsilon_i \in F \setminus \{0\}$ such that the first nonzero coordinate of $\varepsilon_i x_i$ is the same as the last nonzero coordinate of $\varepsilon_{i-1} x_{i-1}$. Without loss of generality one may suppose that all $\varepsilon_i = 1$. Then

$$\nu(g + x_0 + x_1 + \cdots + x_n) = \nu(g + x_0) + \nu(x_1) + \cdots + \nu(x_k).$$

Pick any $\varepsilon \in F \setminus \{0, 1\}$ and for each $i \leq k$, put

$$g_i = g + x_0 + \varepsilon x_1 + \cdots + \varepsilon^{i-1} x_{i-1} + \varepsilon^i x_i + \varepsilon^i x_{i+1} + \cdots + \varepsilon^i x_k,$$

in particular,

$$g_0 = g + x_0 + x_1 + \cdots + x_k.$$

Then

$$\begin{aligned}\nu(g_i) &= \nu(g + x_0) + (\nu(x_1) + 1) + \cdots + (\nu(x_i) + 1) + \nu(x_{i+1}) + \cdots + \nu(x_k)\\ &= \nu(g + x_0) + \nu(x_1) + \cdots + \nu(x_k) + i\\ &= \nu(g_0) + i.\end{aligned}$$

It follows that there is $j \leq k$ such that

$$\nu(g_j) \equiv 2^m \mod 2^{m+1},$$

and so $g_j \in A_m$. Clearly $g_j \in W$.

Both these partitions have been known for a long time. The problem was that they did not work in case V is uncountable.

In [5], it is proved that every group of the form $\bigoplus_{\alpha<\kappa} G_\alpha$, where κ is an infinite cardinal and for each $\alpha < \kappa$, G_α is a finite group of odd order, admits an ω-partition such that every coset modulo infinite subgroup meets each cell of the partition.

In this paper we show the following.

Theorem 6. *Let V be an arbitrary infinite vector space over a finite field F and let $F \neq \mathbb{F}_2$. Then there is a partition $\{A_m : m < \omega\}$ of V such that for every infinite affine subspace W of V and $m < \omega$, one has $W \cap A_m \neq \emptyset$.*

The proof of Theorem 6 is a modification of that of [5].

The crucial idea of the construction is contained in the next lemma.

Lemma. *Let $(c_n)_{n<\omega}$ be an increasing sequence in $\mathbb{N}$ such that $c_0 = 1$ and $c_{n+1} - c_n \to \infty$. Then there is a function $\varphi : \mathbb{N} \to [\mathbb{N}]^{<\omega}$ with the following properties:*

(1) if $a \in [c_n, c_{n+1})$, then $\varphi(a) = \{a_0, \ldots, a_{n-1}\}$, where $a_k \in [c_k, c_{k+1})$ for all $k \leq n-1$ (in particular, if $a \in [c_0, c_1)$, then $\varphi(a) = \emptyset$), and

(2) for every $d \in \mathbb{N}$, there is $c \in \mathbb{N}$ such that whenever $a, b \in \mathbb{N}$, $0 < |a - b| \leq d$, $u \in \varphi(a)$, $v \in \varphi(b)$ and $u, v \geq c$, one has $|u - v| > d$.

Proof. Without loss of generality one may suppose that $c_{n+1} - c_n \geq 5$ for all $n < \omega$. For each $n < \omega$, pick an integer $d_n \geq 2$ such that $d_n^2 + d_n - 1 \leq c_{n+1} - c_n$ and $d_n \to \infty$. To define φ, let $k < \omega$ and let $a \geq c_{k+1}$. Choose the largest integer $l \geq 0$ for which $c_{k+1} + l d_k^2 \leq a$, then choose the largest integer

$i \geq 0$ for which $c_{k+1} + ld_k^2 + id_k \leq a$, and then put $j = a - c_{k+1} - ld_k^2 - id_k$. Thereby we write a in the form

$$a = c_{k+1} + ld_k^2 + id_k + j,$$

where l is a non-negative integer and $i, j \in \{0, \ldots, a_k - 1\}$. Put

$$a_k = c_k + jd_k + i.$$

Since $a_k \in [c_k, c_k + d_k^2)$, so defined function φ satisfies condition (1). To check (2), let $d \in \mathbb{N}$ be given. Choose $n_0 < \omega$ such that $d_n \geq d + 2$ for all $n \geq n_0$ and put $c = c_{n_0}$. Now let $a, b \in \mathbb{N}$, $0 < |a - b| \leq d$, $u \in \varphi(a)$, $v \in \varphi(b)$ and $u, v \geq c$. Since

$$\begin{aligned} c_{k+1} - a_k &= c_{k+1} - c_k - jd_k - i \\ &\geq d_k^2 + d_k - 1 - jd_k - i \\ &\geq d_k, \end{aligned}$$

one may suppose that $u = a_k$ and $v = b_k$ for some $k \geq n_0$. Let

$$a = c_{k+1} + ld_k^2 + id_k + j,$$

$$b = c_{k+1} + l'd_k^2 + i'd_k + j'.$$

Then

$$a_k = jd_k + i,$$

$$b_k = j'd_k + i'.$$

Thus, we have that

$$a - b = [(l - l')d_k + (i - i')]d_k + (j - j'),$$

$$a_k - b_k = (j - j')d_k + (i - i').$$

Notice that $|i - i'| < d_k$ and $|j - j'| < d_k$. Then it follows from the second equality that if $|j - j'| > 1$, then $|a_k - b_k| > d_k$. Therefore one may suppose that $|j - j'| \leq 1$. But then it follows from the first equality that

$$(l - l')d_k + (i - i') = 0.$$

This gives us $l = l'$ and $i = i'$. Consequently, $|j - j'| = 1$, and we obtain that $|a_k - b_k| = d_k$. □

We are now in a position to prove Theorem 6.

Proof of Theorem 6. Let $V = \bigoplus_\kappa F$, where $\kappa = |V|$, and let $\varphi : \mathbb{N} \to [\mathbb{N}]^{<\omega}$ be a function guaranteed by the Lemma. For every $x \in V \setminus \{0\}$, define the subset $\mathrm{supp}_\varphi(x) \subseteq \mathrm{supp}(x)$ (as usual, $\mathrm{supp}(x) = \{\alpha < \kappa : x(\alpha) \neq 0\}$) and the nonnegative integer $\nu(x)$ as follows.

Let $\mathrm{supp}(x) = \{\alpha_0, \dots, \alpha_{l-1}\}$, where $\alpha_0 < \dots < \alpha_{l-1}$, and let $l \in [c_n, c_{n+1})$. Then $\varphi(l) = \{l_0, \dots, l_{n-1}\}$ for some $l_k \in [c_k, c_{k+1})$, $k \leq n-1$. Put

$$\mathrm{supp}_\varphi(x) = \{\alpha_{l_0}, \dots, \alpha_{l_{n-1}}\}$$

and define $\nu(x)$ to be the number of pairs of distinct neighbouring elements in the sequence

$$x(\alpha_{l_0}), \dots, x(\alpha_{l_{n-1}}).$$

(If $l \in [c_0, c_1)$, then $\varphi(l) = \emptyset$, and then $\mathrm{supp}_\varphi(x) = \emptyset$ and $\nu(x) = 0$.)

We define the partition $\{A_m : m < \omega\}$ of $V \setminus \{0\}$ by

$$A_m = \{x \in A : \nu(x) \equiv 2^m \mod 2^{m+1}\}.$$

The proof that the partition so defined is as required involves the following notion.

Let $(x_n)_{n<\omega}$ be a sequence in V. A sequence $(y_n)_{n<\omega}$ is called a *sum subsystem* of $(x_n)_{n<\omega}$ if there is a sequence $(H_n)_{n<\omega}$ in $[\omega]^{<\omega} \setminus \emptyset$ such that for every $n < \omega$, $\max H_n < \min H_{n+1}$ and $y_n = \sum_{i \in H_n} x_i$.

In these terms the Finite Sums Theorem says that whenever V is partitioned into finitely many cells, there is a sum subsystem $(y_n)_{n<\omega}$ of $(x_n)_{n<\omega}$ with $\mathrm{FS}((y_n)_{n<\omega})$ contained in one cell [4, Corollary 5.15].

Now let U be an infinite vector subspace of V, $g \in V$, $W = g + U$. We shall show that $W \cap A_m \neq \emptyset$ for each $m < \omega$.

Put $k = 2^{m+1} - 1$. Given a sequence $(y_i)_{i \leq k}$ in U, let

$$g_0 = g + y_0 + y_1 + \dots + y_k.$$

It suffices to construct a sequence $(y_i)_{i \leq k}$ such that

(1) $\mathrm{supp}_\varphi(g_0) \cap \mathrm{supp}(y_i) \neq \emptyset$,

(2) $\max(\mathrm{supp}_\varphi(g_0) \cap \mathrm{supp}(g)) < \min(\mathrm{supp}_\varphi(g_0) \cap \mathrm{supp}(y_0))$

if $\mathrm{supp}_\varphi(g_0) \cap \mathrm{supp}(g) \neq \emptyset$, and

(3) $\max(\mathrm{supp}_\varphi(g_0) \cap \mathrm{supp}(y_i)) < \min(\mathrm{supp}_\varphi(g_0) \cap \mathrm{supp}(y_{i+1}))$ for all $i < k$.

Then one could repeat the argument from Example 5 to finish the proof.

Choose first a sequence $(x_n)_{n<\omega}$ in $U \setminus \{0\}$ such that

$$\mathrm{supp}(g) \cap \mathrm{supp}(x_i) = \emptyset \quad \text{and} \quad \mathrm{supp}(x_i) \cap \mathrm{supp}(x_j) = \emptyset$$

for all $i < j < \omega$. Clearly, one may suppose that the sequence $(\min\operatorname{supp}(x_n))_{n<\omega}$ is increasing. Let

$$\gamma = \sup\{\min\operatorname{supp}(x_n) : n < \omega\}.$$

Then every $x \in V$ can be uniquely written in the form $x = x' + x''$, where $x', x'' \in V$,

$$\operatorname{supp}(x') = \operatorname{supp}(x) \cap \gamma \quad \text{and} \quad \operatorname{supp}(x'') = \operatorname{supp}(x) \setminus \operatorname{supp}(x').$$

Choose inductively a sum subsystem $(y_n)_{n<k}$ of $(x_n)_{n<\omega}$ such that

(a) $\max\operatorname{supp}(g') < \min\operatorname{supp}(y'_0)$ and $\max\operatorname{supp}(y'_i) < \min\operatorname{supp}(y'_{i+1})$ for all $i < k-1$, and

(b) each of the intervals $(|\operatorname{supp}(g')|, |\operatorname{supp}(g' + y'_0)|]$ and

$$(|\operatorname{supp}(g' + y'_0 + \cdots + y'_i)|, |\operatorname{supp}(g' + y'_0 + \cdots + y'_i + y'_{i+1})|],$$

where $i < k-1$, contains some interval $[c_n, c_{n+1})$.

Let $h = g + y_0 + \cdots + y_{k-1}$, $y_{k-1} = \sum_{n\in H} x_n$ and $n_0 = \max H + 1$. We claim that there is $y_k \in \operatorname{FS}((x_n)_{n_0\le n<\omega})$ such that

(i) $\max\operatorname{supp}(h') < \min\operatorname{supp}(y_k)$,

(ii) the interval $(|\operatorname{supp}(h')|, |\operatorname{supp}(h+y_k)|]$ contains some $[c_n, c_{n+1}]$, and

(iii) $\operatorname{supp}_\varphi(h + y_k) \cap \operatorname{supp}(h'') = \emptyset$.

Then the sequence $(y_i)_{i\le k}$ will satisfy conditions (1)-(3).

Assume the contrary. Then by the Finite Sums Theorem, there exist $\delta \in \operatorname{supp}(h'')$ and a sum subsystem $(z_n)_{n<\omega}$ of $(x_n)_{n_0\le n<\omega}$ such that

$$\max\operatorname{supp}(h') < \min\operatorname{supp}(z_0), \text{ and}$$

$$\delta \in \operatorname{supp}_\varphi(h + z) \text{ for all } z \in \operatorname{FS}((z_n)_{n<\omega}).$$

Put $d = |\operatorname{supp}(h + z_0)|$ and let c be a constant guaranteed by the Lemma. Pick $z \in \operatorname{FS}((z_n)_{1\le n<\omega})$ such that

$$\max\operatorname{supp}(h' + z'_0) < \min\operatorname{supp}(z), \text{ and}$$

$$|\operatorname{supp}(h' + z')| \ge c.$$

Let $a = |\operatorname{supp}(h + z_0 + z)|$ and $b = |\operatorname{supp}(h + z)|$. Then

$$0 < a - b = |\operatorname{supp}(z_0)| \le |\operatorname{supp}(h + z_0)| = d.$$

Let $u = |\operatorname{supp}(h + z_0 + z) \cap \delta|$, $v = |\operatorname{supp}(h + z) \cap \delta|$, $w = |\operatorname{supp}(h + z_0) \cap \delta|$ and $t = |\operatorname{supp}(h) \cap \delta|$. Then $u = v + w - t$, and so

$$u - v = w - t,$$

which is a contradiction, since $u \in \varphi(a)$, $v \in \varphi(b)$, $u, v \ge c$ and $w - t \le w \le d$. □

References

1. R. Graham and B. Rothschild, *Ramsey's Theorem for n-parameter sets*, Trans. Amer. Math. Soc **159** (1971), 257-292.
2. R. Graham, B. Rothschild, and J. Spencer, *Ramsey Theory*, Wiley, New York, 1990.
3. N. Hindman, *Finite sums from sequences within cells of a partition of* $\mathbb{N}$, J. Combin. Theory Ser. A **17** (1974), 1-11.
4. N. Hindman and D. Strauss, *Algebra in the Stone-Čech compactification*, De Gruyter, Berlin, 1998.
5. Y. Zelenyuk, *Partitions and sums with inverses in Abelian groups*, J. Combin. Theory Ser. A, accepted.

AUTHOR INDEX