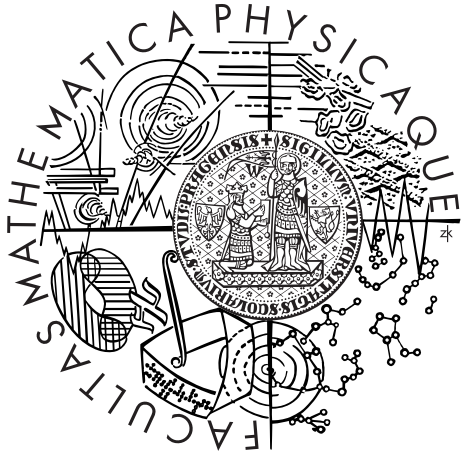
Charles University in Prague
Faculty of Mathematics and Physics

MASTER THESIS



Jakub Skalický

Effective elliptic curves arithmetics over finite fields

Department of Algebra

Supervisor of the master thesis: Jan Krhovják
Study programme: Mathematics
Specialization: MMIB

Prague 2012

I would like to thank all people, who have supported me during my studies and while writing my thesis. First and foremost, they are my family and close friends. I must express gratitude to CEPIA Technologies as well – they have provided me with all support I needed, including the financial one. Thank you.

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Brno 30th July 2012 Jakub Skalický

Název práce: Efektivní aritmetika eliptických křivek nad konečnými tělesy

Autor: Jakub Skalický

Katedra: Katedra algebry

Vedoucí diplomové práce: Mgr. Jan Krhovják, CEPIA Technologies

Abstrakt: Práce se zabývá aritmetikou eliptických křivek nad konečnými tělesy a způsoby, jak tyto výpočty zefektivnit. V první části jsou pomocí pojmů a vět z algebraické geometrie definovány eliptické křivky a odvozeny jejich základní vlastnosti včetně základních algoritmů na počítání s body křivky. Ve druhé kapitole je vidět, jak lze výpočty zrychlit pomocí techniky time-memory tradeoff, tj. přidání redundance a konečně ve třetí zavádíme zcela nový tvar křivek, který je pro dané účely velmi efektivní.

Klíčová slova: eliptické křivky nad konečnými tělesy, efektivní aritmetika eliptických křivek, ECDSA, Edwardsovy křivky

Title: Effective Elliptic Curves Arithmetics Over Finite Fields

Author: Jakub Skalický

Department: Department of Algebra

Supervisor: Jan Krhovják, CEPIA Technologies

Abstract: The thesis deals with arithmetics of elliptic curves over finite fields and methods to improve those calculations. In the first part, algebraic geometry helps to define elliptic curves and derive their basic properties including the group law. The second chapter seeks ways to speed up these calculations by means of time-memory tradeoff, i.e. adding redundancy. At last, the third part introduces a wholly new curve form, which is particularly effective for such purposes.

Keywords: elliptic curves over finite fields, effective elliptic curves arithmetics, ECDSA, Edwards curves

Contents

Preface

In my Master's Thesis I will focus on the problem of effective arithmetics on elliptic curves. What might seem negligible when computing with small numbers, becomes an almost unsuperable setback when taking big numbers into account. In practise, considering recent cryptographic elliptic curve standards built over $GF(p)$ with p a few hundred bits long, even a slight theoretical increase in computing efficiency might result in a great improvement in practical implementations. However, like in many other fields, there is no universally applicable best solution and one must carefully choose the one that fits best.

The history of elliptic curves began in 18th century, when Leonhard Euler showed that it is possible to integrate algebraically the following equation

$$\frac{dx}{\sqrt{\alpha + \beta x + \gamma x^2 + \delta x^3 + \epsilon x^4}} + \frac{dy}{\sqrt{\alpha + \beta y + \gamma y^2 + \delta y^3 + \epsilon y^4}} = 0.$$

Followed by Abel, who himself added a vast amount of theory, Weierstrass and other famous mathematicians, elliptic curves[1] started to reveal themselves. Probably the single most famous exploit using elliptic curves was Wiles' proof of Fermat's Last Theorem, in its final correct form published in October 1994 [34, 35].[2] Nevertheless, from a recent cryptographer's point of view, the most attention is focused on curves' group structure providing good background for various public-key cryptosystems. Other important practical uses include primality proving [2] and integer factorization [25], for instance.

In 1976, Diffie and Hellman [10] first introduced public-key cryptography. A decade later, in 1985, ElGamal [14] described how to take advantage of the discrete logarithm problem while constructing a public-key signature scheme. His approach was in a slightly different way implemented in Digital Signature Algorithm, part of the US Digital Signature Standard [27].

At about the same time as ElGamal, Koblitz [24] and Miller [26] independently proposed taking advantage of discrete logarithm problem in the group of points of an elliptic curve defined over a finite field. The first step towards what we nowadays know as Elliptic Curve Digital Signature Algorithm (ECDSA) was taken in 1992 by S. Vanstone [33] in a public comment on NIST's DSA proposal. Nevertheless, it had taken several more years before it became a standard: it was accepted as ISO 14888-3 in 1998, as ANSI X9.62 in 1999 and both IEEE 1363-2000 and FIPS 186-2 in 2000 [20].

The primary advantage of ECDSA over DSA is the absence of a subexponential-time algorithm that would find discrete logarithms in groups on elliptic curves. Therefore, to maintain the same level of security, it is possible to use smaller key lenghts, which saves bandwidth and renders implementations more effective. These features particularly appeal in cases when the environment is limited in memory or computing power, such as smart cards, PC cards or wireless devices.

To be more precise, security level of 80 bits (meaning that the potential attacker would have to make at least 2^{80} signature generations to find the secret

[1] In the early era referred to rather as *elliptic functions*.

[2] For a brief look at the astonishing history of solving this well-known mathematical problem, see e.g. [22].

key) requires a DSA public key to be at least 1024 bits long, whereas ECDSA public key comprises only 160 bits.

These apparent advantages of ECSDA over DSA are balanced with a relatively more strenuous implementation, which might be the reason why DSA still prevails. To construct an elliptic curve cryptosystem, one must take some basic steps:

(i) select an underlying field F_q,

(ii) select a representation of elements of F_q,

(iii) implement arithmetics in F_q,

(iv) select a suitable elliptic curve E over F_q and

(v) implement the elliptic curve operations in E.

This thesis is focused solely on the fifth point, which indeed might be a good indication of how much theory there is. For instance, as of March, 2012, the Explicit-Formulas Database [5] contains 581 explicit formulas!

1. Elliptic curves in general

The first goal must be to establish a well-based definition of elliptic curves. To achieve this, we must take a number of steps.

1.1 Affine and projective varieties

By K we will always mean a *perfect field*, i.e. a field, in which Frobenius endomorphism is an automorphism. Its algebraic closure will be denoted as $\bar{K}$ and field of constants as $\tilde{K}$ (i.e. $\tilde{K} = \{x \in K : x \text{ is algebraic over } K\}$).

Definition 1.1.1. The *affine n-space* over a field K is $\mathbb{A}^n = \mathbb{A}^n(\bar{K}) = \bar{K}^n = \big\{(x_1, \ldots, x_n) : x_i \in \bar{K} \text{ for all } 1 \leq i \leq n\big\}$. The set of K-rational points of $\mathbb{A}^n$ is $\mathbb{A}^n(K) = \{(x_1, \ldots, x_n) : x_i \in K \text{ for all } 1 \leq i \leq n\}$.

Definition 1.1.2. Let $\bar{K}[X] = \bar{K}[X_1, \ldots, X_n]$ be a polynomial ring in n variables and let $I \subset \bar{K}[X]$ be an ideal. An *(affine) algebraic set* is any set of the form $V(I) = \{P \in \mathbb{A}^n : f(P) = 0 \text{ for all } f \in I\}$. If V is an algebraic set, the *ideal of* V is given by $I(V) = \big\{f \in \bar{K}[X] : f(P) = 0 \text{ for all } P \in V\big\}$.

An algebraic set V is defined over K (we denote this by V/K), if its ideal $I(V)$ can be generated by polynomials in $K[X]$.

Finally, an algebraic set V is called an *(affine) algebraic variety* if $I(V)$ is a prime ideal in $\bar{K}[X]$.

The next thing to do is to define the dimension of V.

Definition 1.1.3. Let V be a variety, then the *affine coordinate ring of* V is defined by

$$\bar{K}[V] = \frac{\bar{K}[X]}{I(V)}.$$

Its quotient field, denoted $\bar{K}(V)$, is called the *function field of* V and its transcendence degree over $\bar{K}$ is called the *dimension of* V.

Remark 1.1.4. For our purposes we need to investigate only a limited number of cases. In general, if $V \subset \mathbb{A}^n$ is given by a single non-constant irreducible polynomial equation $f(X_1, \ldots, X_n) = 0$, then $\dim(V) = n - 1$. The proof is simple: in order to V be a variety, $I(V)$ must be a prime ideal, which in (multivariate) polynomial rings is equivalent to $f(X_1, \ldots, X_n)$ being irreducible. At the same time, it constitutes an equation of n variables and 0, therefore the transcendence degree over $\bar{K}$ is $n - 1$.

We now switch from affine varieties to projective ones, so that we can finalize the definition of a curve.

Definition 1.1.5. The *projective n-space* over K (denoted $\mathbb{P}^n$ or $\mathbb{P}^n(\bar{K})$) is the set of all $(n+1)$-tuples $(x_0, \ldots, x_n) \in \mathbb{A}^{n+1}$, such that not all $x_i = 0$, modulo the equivalence relation given by $(x_0, \ldots, x_n) \sim (y_0, \ldots, y_n) \Leftrightarrow \exists \lambda \in \bar{K}^*$ such that $x_i = \lambda y_i$ for all $0 \leq i \leq n$. An equivalence class $\{(\lambda x_0, \ldots, \lambda x_n)\}$ is denoted by $[x_0 : \ldots : x_n]$ and $x_0, \ldots, x_n$ are called *homogeneous coordinates* for the corresponding point in $\mathbb{P}^n$. The set of K-rational points in $\mathbb{P}^n$ is the set $\mathbb{P}^n(K) = \{[x_0, \ldots, x_n] \in \mathbb{P}^n : \text{all } x_i \in K\}$.

Definition 1.1.6. A polynomial $f \in \bar{K}[X] = \bar{K}[X_0, \ldots, X_n]$ is *homogeneous of degree* d, if

$$f(\lambda X_0, \ldots, \lambda X_n) = \lambda^d f(X_0, \ldots, X_n)$$

for all $\lambda \in \bar{K}$. An ideal $I \subset \bar{K}[X]$ is *homogeneous* if it is generated by homogeneous polynomials.

Although it is not well-defined to evaluate a general homogeneous polynomial f at any point $P \in \mathbb{P}^n$ (this would be possible if and only if $d = 0$), it does make sense to ask whether f vanishes at a certain projective point. From this question a definition similar to affine algebraic variety arises.

Definition 1.1.7 (Projective Variety)**.** Let $I \subset \bar{K}[X_0, \ldots, X_n]$ be a homogenenous ideal. A *(projective) algebraic set* is any set of the form

$$V(I) = \{P \in \mathbb{P}^n : f(P) = 0 \text{ for all homogeneous } f \in I\}.$$

If V is a projective algebraic set, the *(homogeneous) ideal of* V, denoted $I(V)$, is the ideal in $\bar{K}[X]$ generated by

$$\left\{f \in \bar{K}[X] : f \text{ is homogeneous and } f(P) = 0 \text{ for all } P \in V\right\}.$$

Such a V is defined over K, denoted by V/K, if its ideal $I(V)$ can be generated by homogeneous polynomials in $K[X]$. Finally, a projective algebraic set V is called a *(projective) variety* if its homogeneous ideal $I(V)$ is a prime ideal in $\bar{K}[X]$.

Remark 1.1.8. We shall connect affine and projective varieties somehow, so that we are able to define important properties on projective varieties by their affine counterparts. To do so, we must at first explain the behaviour of projective points. Let

$$\begin{aligned} \phi_i : \mathbb{A}^n &\to \mathbb{P}^n \\ (y_1, \ldots, y_n) &\to (y_1, \ldots, y_{i-1}, 1, y_{y+1}, \ldots, y_n) \end{aligned}$$

and let $U_i = \{[x_0 : \ldots : x_n] \in \mathbb{P}^n : x_i \neq 0\}$, then (since $x_i \neq 0$, all quantities x_j/x_i are well-defined) there is a natural bijection

$$\begin{aligned} \phi_i^{-1} : U_i &\to \mathbb{A}^n \\ [x_0 : \ldots : x_n] &\to \left(\frac{x_0}{x_i}, \frac{x_1}{x_i}, \cdots, \frac{x_{i-1}}{x_i}, \frac{x_{i+1}}{x_i}, \cdots, \frac{x_n}{x_i}\right). \end{aligned}$$

Let V be a projective algebraic set with a homogeneous ideal $I(V) \subset \bar{K}[X]$. The sets $U_0, \ldots, U_n$ cover all $\mathbb{P}^n$, so $V \cap U_0, \ldots, V \cap U_n$ cover the whole V. At the same time, fix i and let $V \cap \mathbb{A}^n = \phi_i^{-1}(V \cap U_i)$, then $V \cap \mathbb{A}^n$ is an affine algebraic set with ideal $I(V \cap \mathbb{A}^n) \subset \bar{K}[Y]$ given by

$$I(V \cap \mathbb{A}^n) = \{f(Y_1, \ldots, Y_{i-1}, 1, Y_i, \ldots, Y_n) : f(X_0, \ldots, X_n) \in I(V)\}.$$

It follows that any subset $V \cap U_0, \ldots, V \cap U_n$ is via the appropriate ϕ_i^{-1} an affine variety. The replacement of $f(X_0, \ldots, X_n)$ by $f(Y_0, \ldots, Y_{i-1}, 1, Y_i, \ldots, Y_n)$

is called the *dehomogenization with respect to* X_i. Of course, a backward process exists as well: for any $f(Y) \in \bar{K}[Y]$ let

$$f^*(X_0, \ldots, X_n) = X_i^d \cdot f\left(\frac{X_0}{X_i}, \frac{X_1}{X_i}, \cdots, \frac{X_{i-1}}{X_i}, \frac{X_{i+1}}{X_i}, \cdots, \frac{X_n}{X_i}\right),$$

where $d = \deg f$ is the smallest integer for which f^* is a polynomial. We call f^* the *homogenization of* f *with respect to* X_i.

Definition 1.1.9. Let V be an affine algebraic set with ideal $I(V)$. The *projective closure of* V, denoted $\bar{V}$, is generated by $\{f^*(X) : f \in I(V)\}$.

Now we have all the tools necessary to state and prove the relationship between both affine and projective varieties.

Proposition 1.1.10. *(i) Let V be an affine variety. Then $\bar{V}$ is a projective variety and $V = \bar{V} \cap \mathbb{A}^n$.*

(ii) Let V be a projective variety. Then $V \cap \mathbb{A}^n$ is an affine variety and either $V \cap \mathbb{A}^n = \emptyset$ or $V = \overline{V \cap \mathbb{A}^n}$.

(iii) If an affine (or projective, respectively) variety V is defined over K, then $\bar{K}$ (or $V \cap \mathbb{A}^n$, respectively) is also defined over K.

Proof. The third point is clear from the definitions of projective closure and affine (respectively, projective) algebraic sets. The first two points follow from [16, I.2.3, p. 11]: V is covered by open sets $V \cap U_i$, $0 \leq i \leq n$, which are homeomorphic to affine varieties via the homogenization of f with respect to X_i (which is exactly the process how to get the projective closure of an affine variety). □

With proposition 1.1.10 in hand, we are able to define the crucial properties of projective varieties in terms of the affine subvariety $V \cap \mathbb{A}^n$. Recall that we identify $\mathbb{A}^n$ with $U_i \subset \mathbb{P}^n$ via ϕ_i^{-1}.

Definition 1.1.11. Let V/K be a projective variety and choose $\mathbb{A}^n \subset \mathbb{P}^n$ so that $V \cap \mathbb{A}^n \neq \emptyset$. The *dimension of* V is the dimension of $V \cap \mathbb{A}^n$. The *function field of* V, denoted $K(V)$, is the function field of $V \cap \mathbb{A}^n$ and similarly for the $\bar{K}(V)$.

To finish this section, we present a definition which will cast some light on future terms.

Definition 1.1.12. A *curve* is a projective variety of dimension one.

It is important to notice that proposition 1.1.10 can be interpreted as follows: any affine variety can be identified with a unique projective variety. Therefore it is possible to abuse notation a little bit and write affine or homogeneous equations as desired. For example, saying "let V be a projective variety" and providing some non-homogeneous equations only means that V is the projective closure of the indicated affine variety W. All points of $V \setminus W$ are called the *points at infinity on* V.

1.2 Algebraic function field

The goal of this section would be to establish Weierstrass equation as a defining equation of algebraic function fields. We present the reader with a couple of the most important definitions of which we will later take advantage.

Definition 1.2.1. Let $K \subseteq F$ be two fields such that $\exists x \in F$, x transcendental over K and $[F : K(x)] < \infty$. Then (K, F) is called an *algebraic function field*. Instead of (K, F) we usually write F/K or F alone.

Definition 1.2.2. A *place* P is any set, for which there exists a valuation ring $O \in F/K$ such that $P = O \setminus O^*$. The corresponding O is denoted by O_P and the set of all places in F/K by $\mathbb{P} = \mathbb{P}_{F/K}$.

Definition 1.2.3. Let F/K be an algebraic function field such that every $x \in F \setminus K$ is transcendent over K (i.e. $K = \tilde{K}$). Let $\mathbb{P} = \mathbb{P}_{F/K}$ be the basis of a free abelian group. Then this group is denoted by $\mathrm{Div}(F/K)$, its elements are called *divisors* and are of the form $\sum_{P \in \mathbb{P}} a_P P$, $a_P \in \mathbb{Z}$ and $a_P \neq 0$ only in finitely many cases.

Any sum of the form $\sum_{P \in \mathbb{P}} v_P(x)P$, $x \in F^*$, is a divisor ($v_P(x) \neq 0$ only for finitely many $P \in \mathbb{P}$, cf. [16, II.6.1, p. 131]). This divisor is called *principle* and denoted by (x).

Definition 1.2.4. Let $A \in \mathrm{Div}(F/K)$. The *Riemann-Roch space* $\mathcal{L}(A)$ is defined as $\{x \in F^* : (x) \geq A\} \cup \{0\}$. Its dimension is denoted by $l(A)$.

By Riemann theorem we know that there exists $\gamma > 0$ such that $\forall A \in \mathrm{Div}(F/K)$: $\deg(A) - l(A) < \gamma$. This is the basis for a definition crucial for elliptic curves:

Definition 1.2.5. The smallest $\gamma \geq 0$ fulfilling $\deg(A) - l(A) < \gamma$ for all $A \in \mathrm{Div}(F/K)$ is called the *genus* and denoted by g.

Proposition 1.2.6 (Riemann-Roch corollary)**.** *Let F/K be an algebraic function field of genus g and let $K = \tilde{K}$. If for any $A \in \mathrm{Div}(F/K)$: $l(A) \geq 2g - 1$, then $l(A) = \deg(A) + g - 1$.*

The following definition brings us at last close to definition of elliptic curves.

Definition 1.2.7. Let K be a field. The *long Weierstrass equation* is an equation of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

where $a_i \in K$, $1 \leq i \leq 6$, $i \neq 5$.

Remark 1.2.8. Notice that, by defining $g(x, y) = y^2 + a_1xy + a_3y$ and $f(x) = x^3 + a_2x^2 + a_4x + a_6$ the Weierstrass equation can be rewritten as $w(x, y) = g(x, y) - f(x)$. Henceforward, we shall only use $w(x, y)$ as an abbreviation for an arbitrary Weierstrass equation.

Our goal is to prove that (under certain circumstances) this is an equation defining an elliptic curve. To do so, we must prove that for any function field F/K there exists a Weierstrass equation and vice versa.

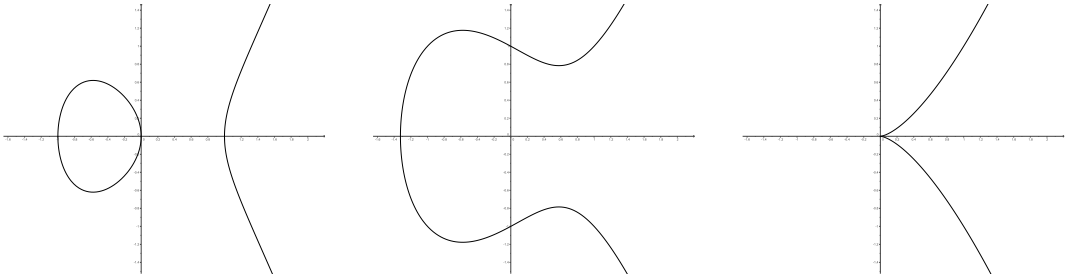


Figure 1.1: A gallery of Weierstrass curves. Left: $y^2 = x^3 - x$, center: $y^2 = x^3 - x + 1$, right: $y^2 = x^3$. Note that in the last case the curve is not smooth at $(0, 0)$, making a *cusp*. See section 1.3 for further explanation.

Proposition 1.2.9. *Let F/K be an algebraic function field, let $K = \tilde{K}$ and $\mathbb{P} = \mathbb{P}_{F/K}$. Let $g = 1$ and $P \in \mathbb{P}$, $\deg(P) = 1$. Then there exist $x, y \in F$ such that $F = K(x, y)$, $x \in \mathcal{L}(2P) \setminus \mathcal{L}(P)$, $y \in \mathcal{L}(3P) \setminus \mathcal{L}(2P)$, $[F : K(x)] = 2$, $[F : K(y)] = 3$. Moreover, there exist $a_i \in K$, $i \in \{1, 2, 3, 4, 6\}$, a_i being coefficents of the long Weierstrass equation.*

Proof. As follows from Riemann-Roch corollary, $\mathcal{L}(0) = K$, therefore $l(0) = 1$. On the other hand, for $k \geq 1$: $l(kP) \geq 1 = 2g - 1$ and again by Riemann-Roch corollary we get that $l(kP) = \deg(kP) = k$. Hence, it is possible to choose arbitrarily some $x \in \mathcal{L}(2P) \setminus \mathcal{L}(P)$ and $y \in \mathcal{L}(3P) \setminus \mathcal{L}(2P)$. Moreover, the condition $x \in \mathcal{L}(2P) \setminus \mathcal{L}(P)$ means that $(x)_- = 2P$ (and similarly $(y)_- = 3P$). We have $v_P(x) = -2$, $v_P(y) = -3$, $v_P(x^2) = -4$, $v_P(xy) = -5$ and $v_P(x^3) = v_P(y^2) = -6$.

Let $Z = \{1, x, y, x^2, xy, x^3, y^2\}$, then for any divisor $Q \neq P$: $v_Q(z) \geq 0$ for any $z \in Z$. In other words, $Z \subseteq \mathcal{L}(6P)$. As has already been proven, $l(6P) = 6$, therefore there exist $u_1, u_2, u_3 \in K$ and $v_1, v_2, v_3, v_4 \in K$ such that

$$u_1 y^2 + u_2 xy + u_3 y = v_1 x^3 + v_2 x^2 + v_3 x + v_4 \tag{1.1}$$

with at least some u_i or v_i non-zero. Both $Z \setminus \{x^3\}$ and $Z \setminus \{y^2\}$ are basis of $\mathcal{L}(6P)$ (there is always exactly one element of $\mathcal{L}(iP) \setminus \mathcal{L}((i-1)P)$, $1 \leq i \leq 6$). It follows that both $u_1 \neq 0$ and $v_1 \neq 0$. Multiplying (1.1) by $u_1^3 v_1^2$ yields

$$u_1^4 v_1^2 y^2 + u_1^3 v_1^2 u_2 xy + u_1^3 v_1^2 u_3 y = u_1^3 v_1^3 x^3 + u_1^3 v_1^2 v_2 x^2 + u_1^3 v_1^2 v_3 x + u_1^3 v_1^2 v_4.$$

Substitution $y' = u_1^{-2} v_1^{-1} y$ and $x' = u_1^{-1} v_1^{-1} x$ leads to

$$(y')^2 + u_2 x' y' + u_1 v_1 u_3 y' = (x')^3 + u_1 v_2 (x')^2 + u_1^2 v_1 v_3 x' + u_1^3 v_1^2 v_4,$$

which clearly shows how to define a_i.

The last thing to do is to prove $F = K(x, y)$. At first, since $[F : K(z)] = n \cdot \deg(P)$ whenever $(z)_- = nP$, we have that $[F : K(x)] = 2$ and $[F : K(y)] = 3$. Now $2 = [F : K(x)] = [F : K(x, y)][K(x, y) : K(x)]$, so $[F : K(x, y)]$ divides 2. Similarly for y, $[F : K(x, y)]$ divides 3. The only possible option is that $[F : K(x, y)] = 1$ and $F = K(x, y)$ as desired. □

To reverse the whole situation and see whether any Weierstrass equation defines an algebraic function field, we present the reader with the following proposition.

Proposition 1.2.10. *Let F/K be a field extension such that $F = K(x, y)$, x and y transcendent over K and Weierstrass equation holds: $w(x, y) = 0$. Then F/K is an algebraic function field with $\tilde{K} = K$, $[F : K(x)] = 2$, $[F : K(y)] = 3$. Any polynomial $u \in K[x_1, x_2]$ such that $u(x, y) = 0$ is a multiple of w.*

Proof. $w(x, y) = 0$ provides an alternative approach to F: it says that $x \in F$ is algebraic over $K(y)$ and therefore it is possible to rewrite F as $K(x, y) = K(y)[x]$. Minimal polynomial of x over $K(y)$ must divide $w(x, y)$, so $\deg m_{x|K(y)} = [F : K(y)]$ divides $\deg_x w(x, y) = 3$. Similarly we show that $[F : K(x)]$ divides 2. To prove that they are equal, we must elliminate the cases $F = K(x)$ and $F = K(y)$.

Let $F = K(x)$ and let $v = v_\infty$ be a valuation defined by $v(a/b) = \deg b - \deg a$ for $a, b \in K[x]$ non-zero.[1] We have $g(x, y) = f(x)$ and $v(f(x)) = -3$. Let $v(y) \geq -1$, then $v(g(x, y)) \geq -2$; in case $v(y) \leq -2$, $v(g(x, y)) \leq -4$. Both assumptions contradict $v(f(x)) = -3$, therefore $F \neq K(x)$ and $[F : K(x)] = 2$.

We apply the same process on the case $F = K(y)$. Again we take $v = v_\infty$, this time with y being the variable (i.e. $v(y) = -1$). If $v(x) \geq 0$, then $v(f(x)) \geq 0$, but $v(g(x, y)) = -2$. Therefore $v(x) \leq -1$ and $v(f(x)) = 3v(x)$. At the same time, $v(g(x, y)) \geq -1 + v(x)$. But since $-3r < -1 - r$ for any $r \geq 1$, both conditions cannot be satisfied simultaneously and again $[F : K(y)] = 3$.

Now let $\tilde{K} = \{x \in F : x \text{ algebraic over } K\}$. Then $x, y \in F$ are transcendent over $\tilde{K}$, hence $F = \tilde{K}(x, y)$ and by the first part of this proof $[F : \tilde{K}(x)] = 2$, $[F : \tilde{K}(y)] = 3$. Take e.g. $[F : K(x)]$, it equals $2 = [F : K(x)] = [F : \tilde{K}(x)][\tilde{K}(x) : K(x)] = [F : \tilde{K}(x)][\tilde{K} : K]$ (algebraic and transcendent extensions have the same degree) and because $F \neq \tilde{K}(x)$, it follows that $\tilde{K} = K$.

Let $u(x, y) = 0$ for some $u \in K[x_1, x_2]$. Then u can be expressed as $u = aw + t$, where $\deg_y(t) \leq 1$ (every occurence of y in power greater than 2 can be "hidden" in aw) and $t(x, y) = 0$. If $\deg_y(t) = 0$, $t(x) = 0$ and x would be algebraic over K, which is a contradiction. Therefore $\deg_y(t) = 1$, say $t(x, y) = t_1(x)y + t_0(x)$. If $t_1(x) = 0$, again x would be algebraic over K. So $t_1(x) \neq 0$ and $y = -t_0(x)/t_1(x)$, $y \in K(x)$ and $F = K(x)$, which is a contradiction. It follows that $t = 0$ and u is a multiple of w. □

To connect this section with the previous one, we shall prove that $w(x, y)$ is irreducible as a bivariate polynomial.

Proposition 1.2.11. *Let F/K be a field extension, $F = K(x, y)$, x, y transcendent over K and $w(x, y) = 0$. Then $w(x_1, x_2) \in K[x_1, x_2]$ is irreducible.*

Proof. We denote $K[x, y] \subseteq F$ the subring of F and $K[x_1, x_2]$ the ring of bivariate polynomials with coefficients in K. Moreover, let $K[f] = K[x_1, x_2]/(f)$ for some $f \in K[x_1, x_2]$. If f is irreducible, the quotient field of $K[f]$ is called a function field and denoted by $K(f)$. Let

$$\begin{aligned} \varphi : K[x_1, x_2] &\to K[x, y] \\ x_1 &\to x \\ x_2 &\to y \\ \forall s \in K : s &\to s \end{aligned}$$

[1] This is the valuation of algebraic function field F/K of genus 0, connected to P_∞.

be a ring homomorphism, which is clearly surjective. For any $u \in K[x_1, x_2]$, $\varphi(u) = 0$ if and only if $u(x, y) = 0$ in F. By proposition 1.2.10 $u(x, y) = 0$ if and only if u is a multiple of w, or in other words, $u \in (w)$. It follows that φ is injective as well and indeed an isomorphism.

Hence $K[w] \cong K[x, y]$. The quotient field of $K[x, y]$ is $K(x, y) = F$, therefore $K(w) \cong F$. To complete the proof, it suffices to notice that $K[x_1, x_2]/(w) \cong K[x, y] \subseteq F$ is an integral domain and that factorization by ideals renders an integral domain if and only if the ideal is prime, which in case of polynomials occurs of and only if the polynomial (here w) is irreducible. □

Remark 1.2.12. Let C be the set of solutions of $w(x, y) = 0$ in $\mathbb{A}^2$. We would like to prove that C is a curve. Firstly, as we have just seen, $w(x, y)$ is irreducible, hence the ideal $I(C)$ is prime and C is by definition an affine variety. In view of remark 1.1.4, $w(x, y) = 0$ ensures that $\dim C = 1$. Proposition 1.1.10 (i) states that dimensions of the corresponding affine and projective varieties are the same. To sum up, with a slight abuse of notation (see above), C fulfills the conditions of definition 1.1.12 and is a curve. We call it a *Weierstrass curve*.

1.3 Singularity of Weierstrass curves

We now know that every algebraic function field of genus one[2] might be obtained as a set of solutions of a Weierstrass equation $w(x, y) = 0$. Conversely, every Weierstrass equation $w(x, y)$ provides an algebraic function field. Unfortunately, some cases are not favourable and do not generate function field of genus one. On the other hand, the genus is not an arbitrary one, it is very limited. But before we prove this limitation, we shall need a lemma.

Lemma 1.3.1. *Let $F = K(x, y)$, x, y transcendental over K and $w(x, y) = 0$. Then for every $Q \in \mathbb{P}_{F/K}$ either $v_Q(x) \geq 0$ and $v_Q(y) \geq 0$, or $v_Q(x) < 0$ and $v_Q(y) < 0$. Such a place $P = P_\infty \in \mathbb{P}_{F/K}$ exists only one,* $\deg(P) = 1$, $(x)_- = 2P$, $(y)_- = 3P$.

Proof. As in remark 1.2.8 we express $w(x, y)$ as $g(x, y) - f(x)$. Let $g = g(x, y) \in F$ and $f = f(x) \in F$. For $Q \in \mathbb{P}_{F/K}$ let $v_Q(x) \geq 0$ and $v_Q(y) < 0$. Then $v_Q(f) \geq 0 > 2v_Q(y) = v_Q(g)$. If $v_Q(x) < 0$ and $v_Q(y) \geq 0$, $v(f) = 3v_Q(x) < v_Q(x) \leq v_Q(g)$. Both assumptions lead to a contradiction, hence both valuations must be either negative or non-negative at the same time. Let P denote the place for which $v_P(x) < 0$ and $v_P(y) < 0$. If $v_P(x) \leq v_P(y)$, then $v_P(f) = 3v_P(x) < v_P(x) + v_P(y) \leq v_P(g)$. We get that $v_P(x) > v_P(y)$.

Moreover, $v_P(f) = 3v_P(x) = 2v_P(y) = v_P(g)$, so $v_P(x)$ is a multiple of 2 and $v_P(y)$ a multiple of 3. We know that $2 = [F : K(x)] = \deg((x)_-)$, where $(x)_- = \sum_{Q \in \mathbb{P}_{F/K}} a_Q Q$. If $a_P \neq 0$, then $a_P = -v_P(x) < 0$ and $3v_P(x) = 2v_P(y)$ and a_P is even. This is only possible if $a_Q = 0$ for every $Q \neq P$ ($2 = \sum a_Q \deg(Q)$). As a consequence we get $\deg(P) = 1$ and $(x)_- = 2P$. Similarly we get $(y)_- = 3P$ and the proof is complete. □

Proposition 1.3.2. *Let $F = K(x, y)$, x, y transcendental over K and $w(x, y) = 0$. Then F/K is an algebraic function field of genus* 0 *or* 1.

[2] I.e. an elliptic function field, see later.

Proof. Every $k \geq 2$ can be expressed as $k = 2i + 3j$, $i, j \geq 0$. Therefore, with help of the previous lemma $(x^i y^j)_- = i(x)_- + j(y)_- = kP$. This means that $\mathcal{L}(kP) \setminus \mathcal{L}((k-1)P) \neq \emptyset$ for every $k \geq 2$, hence $l(kP) \geq k$. By Riemann-Roch corollary (1.2.6) $g - 1 = \deg(kP) - l(kP) \leq 0$ for k sufficiently big. It follows that $g \leq 1$ (and $g \geq 0$ by its definition). □

Definition 1.3.3. Let F/K be an algebraic function field given by a Weierstrass equation. We say F/K is singular if and only if $g = 0$.

So far we have been working with general Weierstrass equation. It is however possible to alter coefficients and still maintain the function field F/K.

Proposition 1.3.4. *Let F/K be given by $w(x, y) = 0$. Let $u \in F^*$ and $s, t \in K$ such that $x = u^2\bar{x} + r$ and $y = u^3\bar{y} + u^2 s\bar{x} + t$ for some $\bar{x}, \bar{y}$. Then there exists a Weierstrass equation $\bar{w}$ such that $\bar{w}(\bar{x}, \bar{y}) = 0$ gives F/K as well and its coefficients fulfill:*

$$\begin{aligned}
u\bar{a_1} &= a_1 + 2s, \\
u^2\bar{a_2} &= a_2 - a_1 s - s^2 + 3r, \\
u^3\bar{a_3} &= a_3 + a_1 r + 2t, \\
u^4\bar{a_4} &= a_4 - a_3 s + 2a_2 r - a_1 rs - a_1 t - 2st + 3r^2, \\
u^6\bar{a_6} &= a_6 + a_4 r + a_2 r^2 + r^3 - a_3 t - a_1 rt - t^2.
\end{aligned}$$

Proof. The place $P = P_\infty(w)$ is uniquely determined by $w(x, y) = 0$ (lemma 1.3.1). Let $P = P_\infty(\bar{w})$ for $\bar{w}(\bar{x}, \bar{y}) = 0$. Again by lemma 1.3.1, $x, \bar{x} \in \mathcal{L}(2P) \setminus \mathcal{L}(P)$ and $y, \bar{y} \in \mathcal{L}(3P) \setminus \mathcal{L}(2P)$. These are precisely the conditions in which we proved existence of Weierstrass equation in the first place. Therefore such $\bar{w}$ exists, let $\bar{a_i}$ be its coefficients, $1 \leq i \leq 6$, $i \neq 5$.

Basis of $\mathcal{L}(2P)$ is $\{1, \bar{x}\}$ and of $\mathcal{L}(3P)$ correspondingly $\{1, \bar{x}, \bar{y}\}$. Hence, $x = u\bar{x} + r$ and $y = v\bar{y} + s\bar{x} + t$ for some $r, s, t \in K$ and $u, v \in K^*$. Substitution into $w(x, y) = 0$ yields $w(u\bar{x} + r, v\bar{y} + s\bar{x} + t) = 0$. It must be a multiple of $\bar{w}$ and because their degrees equal, the multiplication coefficient must be linear and both u^3 and v^2. Set $\gamma = v/u$, then since $u^3 = v^2$, $\gamma^2 = u$ and $\gamma^3 = v$. Let $\bar{u} = \gamma$ and $\bar{s} = s\gamma^{-2}$, we get that $x = \bar{u}^2\bar{x} + r$ and $y = \bar{u}^3\bar{y} + \bar{u}^2\bar{s}\bar{x} + t$. Uniqueness of such a transformation follows from the uniqueness of expression of x and y to a basis of $\mathcal{L}(2P)$ and $\mathcal{L}(3P)$.

We have proven that $\bar{w}(\bar{x}, \bar{y}) = u^{-6} w(u^2\bar{x} + r, u^3\bar{y} + u^2 s\bar{x} + t)$ and the rest of proof is only a straightforward calculation. □

It is easy to that if $\bar{w}$ can be obtained from w, the process works the other way round as well. Therefore we call two Weierstrass equations w and $\bar{w}$ *equivalent*, if one can be obtained from the other. The transformation itself is called an *admissible change of variables*. Before we use it to transform long Weierstrass equation into simpler forms, we define a couple of quantities.

Definition 1.3.5. Let $w(x, y) = 0$ be a Weierstrass equation. Define

$$\begin{aligned}
b_2 &= a_1^2 + 4a_2, \\
b_4 &= 2a_4 + a_1a_3, \\
b_6 &= a_3^2 + 4a_6, \\
b_8 &= a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2, \\
\Delta &= -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6, \\
c_4 &= b_2^2 - 24b_4, \\
c_6 &= -b_2^3 + 36b_2b_4 - 216b_6, \\
j &= \frac{c_4^3}{\Delta}.
\end{aligned}$$

Furthermore, Δ is called the *discriminant* and j the *j-invariant*.

Remark 1.3.6. It does not require much work to verify the following equations:

$$4b_8 = b_2b_6 - b_4^2 \text{ and } 12^3\Delta = c_4^3 - c_6^2.$$

What is more important, is the calculation of these quantities after an admissible change of variables. It is not difficult at all, it is rather tedious substitution of a_i's by $\bar{a_i}$'s according to proposition 1.3.4. We shall only state the results:

$$\begin{aligned}
u^2\bar{b_2} &= b_2 + 12r, \\
u^4\bar{b_4} &= b_4 + rb_2 + 6r^2, \\
u^6\bar{b_6} &= b_6 + 2rb_4 + r^2b_2 + 4r^3, \\
u^8\bar{b_8} &= b_8 + 3rb_6 + 3r^2b_4 + r^3b_2 + 3r^4, \\
u^4\bar{c_4} &= c_4, \\
u^6\bar{c_6} &= c_6, \\
u^{12}\bar{\Delta} &= \Delta, \\
\bar{j} &= j.
\end{aligned}$$

As we can see, j-invariant fully deserves its name: two equivalent Weierstrass equations have the same j-invariant. Moreover, over $\bar{K}$ even the converse is true [30, III.1.4 (b)].

Using admissible change of variables it is possible to transform long Weierstrass equation into shorter ones depending on the field characteristics. We omit the case $\mathrm{char}(K) = 3$.

Proposition 1.3.7. *Let C/K be a curve given by a long Weierstrass equation. Then there exist $u \in K^*$ and $r, s, t \in K$ that the substitution*

$$x = u^2\bar{x} + r \qquad y = u^3\bar{y} + u^2s\bar{x} + t$$

leads to a so-called short Weierstrass equation *of the form:*

(i) $y^2 = x^3 + a_4x + a_6$ *if* $\mathrm{char}(K) \neq 2, 3$,

(ii) $y^2 + a_1xy = x^3 + a_2x^2 + a_6$ *or* $y^2 + a_3y = x^3 + a_4x + a_6$ *if* $\mathrm{char}(K) = 2$.

Proof. Let $\mathrm{char}(K) \neq 2, 3$. At first set

$$x = \bar{x} \qquad y = \bar{y} - (a_1/s)\bar{x} - a_3/2,$$

that ensures $\bar{a_1} = \bar{a_3} = 0$ and transforms Weierstrass equation into $\bar{y}^2 = \bar{x}^3 + \bar{a_2}\bar{x}^2 + \bar{x_4}\bar{x} + \bar{a_6}$. Now set

$$\hat{x} = \bar{x} - \bar{a_2}/3 \qquad \hat{y} = \bar{y},$$

then we obtain $\hat{a_2} = 0$ as desired.

When $\mathrm{char}(K) = 2$ and $a_1 \neq 0$, the substitution

$$x = a_1^2\bar{x} + a_3/a_1 \qquad y = a_1^3\bar{y} + (a_1^2 a_4 + a_3^2)/a_1^3.$$

leads to the form $y^2 + xy = x^3 + a_2x^2 + a_6$. On the other hand, if $a_1 = 0$, then the form $y^2 + a_3y = x^3 + a_4x + a_6$ can be achieved by setting

$$x = \bar{x} + a_2 \qquad y = \bar{y}.$$

□

Remark 1.3.8. For further use, it is practical to express Δ in terms of every short Weierstrass equation: again straightforward calculations lead to

- $y^2 = x^3 + a_4x + a_6$ implies $\Delta = -16(4a_4^3 + 27a_6^2)$,
- $y^2 + xy = x^3 + a_2x^2 + a_6$ implies $\Delta = a_6$,
- $y^2 + a_3y = x^3 + a_4x + a_6$ implies $\Delta = a_3^4$.

Our goal will now be to prove that two equivalent Weierstrass equations define the same F/K, establish the notion of singularity on a curve and connect it with genus.

Definition 1.3.9. Let C be a plain algebraic curve defined by the polynomial equation $h(x, y) = 0$. Then $P = (x_0, y_0) \in C$ is a *singular point* of C if and only if

$$\frac{\partial h}{\partial x}(x_0, y_0) = 0 \text{ and } \frac{\partial h}{\partial y}(x_0, y_0) = 0.$$

C is non-singular (smooth) if and only if it has no singular points.

The following proposition offers much more user-friendly tool to determine whether a curve is singular or not.

Proposition 1.3.10. [30, III.1.4 (a)] *Let C be a curve given by a Weierstrass equation $w(x, y) = 0$. Then C is non-singular if and only if $\Delta \neq 0$.*

Proof. At first we show that the point at infinity is never singular. To achieve this, we look at the curve in $\mathbb{P}^2$ with homogeneous equation

$$W(X, Y, Z) = Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3 = 0$$

and at the point $\mathcal{O} = [0, 1, 0]$. Since $\partial W/\partial Z(\mathcal{O}) = 1 \neq 0$, by definition is $\mathcal{O}$ not a singular point of C. Next suppose that C is singular at $P_0 = (x_0, y_0)$. The substitution

$$x = \bar{x} + x_0 \qquad y = \bar{y} + y_0$$

leaves both Δ and c_4 invariant ($u = 1$, hence $u^{-4}c_4 = c_4$ and $u^{-12}\Delta = \Delta$), so we may assume that C is singular at $(0, 0)$. Then

$$a_6 = w(0,0) = 0 \qquad a_4 = (\partial w/\partial x)(0,0) = 0 \qquad a_3 = (\partial w/\partial y)(0,0) = 0,$$

so C can be expressed as

$$C: \ w(x,y) = y^2 + a_1xy - a_2x^2 - x^3 = 0.$$

A straightforward computation of Δ brings the desired result: $b_2 = a_1^2 + 4a_2$, $b_4 = 0$, $b_6 = 0$, $b_8 = 0$, i.e. $\Delta = 0$.

Conversely, we shall prove that if C is smooth, then $\Delta \neq 0$. At first we assume that $\mathrm{char}(K) \neq 2$. Then we have a Weierstrass equation

$$C: \ y^2 = 4x^3 + b_2x^2 + 2b_4x + b_6,$$

(the substitution being $\bar{y} = 1/2(y - a_1x - a_3)$), hence, C is singular if and only if there is a point $(x_0, y_0) \in C$ satisfying

$$2y_0 = 12x_0^2 + 2b_2x_0 + 2b_4 = 0.$$

We see that any such point must be of the form $(x_0, 0)$. Substituting it into the Weierstrass equation we get that x_0 must be a double root of $4x^3 + b_2x^2 + 2b_4x + b_6 = 0$. This cubic polynomial has a double root if and only if its discriminant (which is equal to 16Δ) vanishes, which completes the proof.

Now let $\mathrm{char}(K) = 2$. We want to prove that if $\Delta = 0$, the curve is singular. By proposition 1.3.7, the equation can be transformed into one of the two simpler forms. At first, let us assume $w(x, y) = y^2 + xy + x^3 + a_2x^2 + a_6$. From remark 1.3.8 we know that in this case $\Delta = a_6$. By assumption, $\Delta = 0$, hence $a_6 = 0$. Partial derivation yields

$$\frac{\partial w}{\partial y}(x,y) = x \qquad \frac{\partial w}{\partial x}(x,y) = y + x^2$$

so for both partial derivations to be zero the point $(0, 0)$ must be on the curve. But since $a_6 = 0$, $(0, 0) \in C$ and C is singular. The second short Weierstrass equation $y^2 + a_3y = x^3 + a_4x + a_6$ implies $\Delta = a_3^4$, therefore our goal is to show that if $a_3 = 0$, the curve is singular. Again we compute

$$\frac{\partial w}{\partial y}(x,y) = a_3 \qquad \frac{\partial w}{\partial x}(x,y) = x^2 + a_4$$

and since $a_3 = 0$, the x-coordinate of a potential singular point (x_0, y_0) shall satisfy $x_0^2 = a_4$. Let it be this way, then $y_0^2 = x_0^3 + x_0^2x_0 + a_6 = a_6$, which has a solution as well. We have proved that (x_0, y_0) is a singular point on C and therefore C is singular. □

To restate all the aforementioned, we present this corollary.

Corollary 1.3.11. *Let C be a curve given by a Weierstrass equation,* $\mathrm{char}(K) \neq 3$. *Then C is equivalent to one of the following equations:*

- *if* $\mathrm{char}(K) \neq 2$, *then C is equivalent to $\bar{C}$ given by $y^2 = x^3 + a_4x + a_6$. $\bar{C}$ is singular if and only if $4a_4^3 + 27a_6^2 = 0$,*
- *if* $\mathrm{char}(K) = 2$ *and $a_1 \neq 0$, C is equivalent to $\bar{C}$ given by $y^2 + xy = x^3 + a_2x^2 + a_6$. $\bar{C}$ is singular if and only if $a_6 = 0$.*
- *if* $\mathrm{char}(K) = 2$ *and $a_1 = 0$, C is equivalent to $\bar{C}$ given by $y^2 + a_3y = x^3 + a_4x + a_6$. $\bar{C}$ is singular if and only if $a_3 = 0$.*

Notions of singularity of a curve and singularity of an algebraic function field are closely related, or in better words, are the same.

Proposition 1.3.12. *Let C be the curve given by a Weierstrass equation $w(x, y) = 0$. Let F/K be the algebraic function field given by the same $w(x, y) = 0$ (i.e. $F \cong K(w)$) with genus g. Then C is singular if and only if $g = 0$.*

Proof. According to [30, Section III] or [11], the conditions in corollary 1.3.11 are exactly those for which $g = 0$. □

The previous proposition at last allows for a well-based definition of an elliptic curve.

Definition 1.3.13 (Elliptic Curve)**.** An elliptic curve E is a smooth algebraic curve of genus one, on which there is a specified point $\mathcal{O}$.

From the definition we immediately see that every elliptic curve can be expressed as a non-singular plane cubic; conversely, every smooth Weierstrass plane cubic curve is an elliptic curve. Also it is possible to say that E defines an algebraic function field of genus one, we call it an *elliptic function field* and denote it by E/K.

1.4 Group Law

In this section, the ultimate target is to define group structure on an elliptic curve. We will prove that such a group exists and is unique. Moreover, we will state exact algorithms to perform group operations and present a graphical explanation of the whole process.

There are two approaches to this problem. The first one brings the definition out of blue and then requires to perform a vast number of formula validation to prove correctness. The other one takes yet another detour to mathematical background, but then without tedious calculations allows for a well-based definition as well as deep insight into the matter. This is the one we choose.

We shall start by the following lemma.

Lemma 1.4.1. *Let F/K be an algebraic function field, $P \in \mathbb{P}_{F/K}$ and let $E \subseteq F$ be fields such that F is an algebraic extension of E. Then $E \subseteq \mathcal{O}_P$ can not be true.*

Proof. To achieve a contradiction, let us suppose that the statement is true. Therefore $v_P(c) \geq 0$ for any $c \in E$. Let $z \in F$ be such that $v_P(z) < 0$. Then $v_P(cz^i) \geq iv_P(z)$ for any $c \in E$ and $i \geq 0$. Since z is algebraic over E, $z^n = \sum_{0 \leq i < n} c_i z^i$ for some $n \geq 2$ and $c_i \in E$. For every $i < n$, $v_P(c_i z^i) \geq (n-1)v_P(z) > nv_P(z) = v_P(z^n)$, which is a contradiction. □

Proposition 1.4.2. *Let $P \in \mathbb{P}_{F/K}$ and $x \in F$ be transcendental over K. If $v_P(x) \geq 0$, there exist a unique monic irreducible polynomial $a \in K[x]$ such that $(a) = P \cap K[x]$. If $\deg(a) > 1$, then $\deg(P) > 1$.*

Proof. Since $v_P(x) \geq 0$, it follows that $v_P(b) \geq 0$ for every $b \in K[x]$. Moreover, $P \cap K[x]$ is a prime ideal in $K[x]$ (P is a prime ideal in $\mathcal{O}_P$). If $P \cap K[x] = 0$, then $v_P(b) = 0$ for every non-zero $b \in K[x]$ and hence $v_P(c) = 0$ for every $c \in K^*(x)$. This is only possible if $K(x) \subseteq \mathcal{O}_P$, but this is forbidden by the previous lemma. We have proven that $P \cap K[x] = (a)$ for some irreducible $a \in K[x]$.

Let us remind the definition $\deg(P) = [\mathcal{O}_P/P : (K+P)/P]$. Therefore $\deg(P) = 1$ if and only if $K + P = \mathcal{O}_P$. By this assumption, $\mathcal{O}_P \ni x = t + \alpha$, $t \in P$ and $\alpha \in K$. Then $t = x - \alpha \in P \cap K[x]$ is a multiple of a. □

These two preparatory statements have brought us nearer to our goal, which is the following theorem.

Theorem 1.4.3. *Let F/K be an algebraic function field such that $F = K(x, y)$, x, y transcendental over K. Let $f \in K[x_1, x_2]$ be an irreducible polynomial giving the isomorphism $F \cong K(f)$. Let $v_P(x) \geq 0$ and $v_P(y) \geq 0$. Then there exist unique monic irreducible polynomials $p \in K[x]$ and $q \in K[y]$, such that $P \cap K[x] = (p)$, $P \cap K[y] = (q)$. If $\deg(P) = 1$, then $p = x - \alpha$, $q = y - \beta$ for some $\alpha, \beta \in K$ such that $f(\alpha, \beta) = 0$.*

Conversely, if $f(\alpha, \beta) = 0$ for $\alpha, \beta \in K$ and f is not singular in (α, β), there exists unique $P \in \mathbb{P}_{F/K}$, denoted by $P_{(\alpha,\beta)}$, such that $x - \alpha \in P_{(\alpha,\beta)}$ and $y - \beta \in P_{(\alpha,\beta)}$. Moreover, $\deg\left(P_{(\alpha,\beta)}\right) = 1$.

Proof. Proposition 1.4.2 secures both existence and uniqueness of p and q. If $\deg(P) = 1$, by the same proposition $p = x - \alpha$ and $q = y - \beta$. It remains to prove $f(\alpha, \beta) = 0$. We express $f(x, y) = \sum_{i,j \geq 0} f_{ij}(x-\alpha)^i(y-\beta)^j$, which implies $f_{00} = f(\alpha, \beta)$. Since $f(x, y) = 0$ and every term $f_{ij}(x-\alpha)^i(y-\beta)^j \in P$ for $(i, j) \neq (0, 0)$, we get that $f_{00} = f(\alpha, \beta) \in P$. But $P \cap \mathcal{O}_P^* = \emptyset$, so $f(\alpha, \beta) \notin K^*$ and $f(\alpha, \beta) = 0$ as desired.

What is left is to prove that if the converse assumptions hold, then P is unique and $\deg(P) = 1$. As in the first part, we express $u \in K[x_1, x_2]$ as $\sum_{i,j \geq 0} u_{ij}(x - \alpha)^i(y - \beta)^j$. We have $u \in \mathcal{O}_P$ and $u - u_{00} \in P$. Therefore, if $u_{00} = u(\alpha, \beta) \neq 0$, then $u \in \mathcal{O}_P^*$. We have proved that $\mathcal{O}_P$ contains every $\frac{u(x,y)}{v(x,y)}$, $u, v \in K[x_1, x_2]$ and $v(\alpha, \beta) \neq 0$. We know that in F/K such a valuation ring exists and consists exactly of these elements. It is the image of $\mathcal{O}_{(\alpha,\beta)} \subseteq K[f]$, P is the image of its maximal ideal $\mathcal{M}_{(\alpha,\beta)}$. For $\varphi \in \mathcal{O}_{(\alpha,\beta)}$, $\varphi - \varphi_{(\alpha,\beta)} \in K$, so $\mathcal{O}_{(\alpha,\beta)} = \mathcal{M}_{(\alpha,\beta)} + K$ and $\mathcal{M}_{(\alpha,\beta)}$ is of degree 1. Uniqueness of P is clear. □

Corollary 1.4.4. *Let $F = K(x, y)$ be an elliptic function field given by $w(x, y) = 0$. Let $\mathbb{P}_{F/K}^{(1)} = \{P \in \mathbb{P}_{F/K} : \deg(P) = 1\}$. Then $\mathbb{P}_{F/K}^{(1)} = \{P_\infty\} \cup \{P_{(\alpha,\beta)} : w(\alpha, \beta) = 0 \text{ and } \alpha, \beta \in K\}$.*

Proof. According to theorem 1.4.3, $\mathbb{P}_{F/K}^{(1)}$ is the set of all $P \in \mathbb{P}_{F/K}$, such that $v_P(x) \geq 0$ and $v_P(y) \geq 0$. At the same time, proposition 1.3.1 ensures that the only place dissatisfying this condition is P_∞, which is of degree 1. □

This corollary shows the path which we shall take to get to the target. Every point on an elliptic curve corresponds to a place, which is of degree 1. The converse is true over $\bar{K} = K$: then exist $\alpha, \beta \in K$ such that $x - \alpha, y - \beta \in P$ according to 1.4.2. But with respect to 1.4.3, such $P = P_{(\alpha,\beta)}$.

We can therefore switch our attention from points on elliptic curve to places of degree 1. Neither they have a natural group structure available, that is why we take advantage of an isomorphism to $\mathrm{Pic}(F/K)$. To do so, we must state a few facts.

At first, for an arbitrary algebraic function field, we look at the following groups: $\mathrm{Princ}(F/K)$, $\deg^{-1}(0)$ and $\mathrm{Div}(F/K)$. The fist two are clearly subgroups of $\mathrm{Div}(F/K)$, but since $\deg((x)) = 0$ for every principal divisor (x), even $\mathrm{Princ}(F/K) \leq \deg^{-1}(0)$. For $A, B \in \mathrm{Div}(F/K)$, $A \sim B$ means $A - B \in \mathrm{Princ}(F/K)$.

Definition 1.4.5. Let F/K be an algebraic function field. The group $\deg^{-1}(0)/\mathrm{Princ}(F/K)$ is called the *Picard group* and denoted by $\mathrm{Pic}(F/K)$.

Proposition 1.4.6. *Let F/K be an elliptic function field and let $A \in \mathrm{Div}(F/K)$.*

(i) If $\deg(A) \geq 1$, then $l(A) = \deg(A)$.

(ii) If $\deg(A) = 1$, then there exists unique $P \in \mathbb{P}_{F/K}$, $\deg(P) = 1$, such that $A \sim P$.

(iii) If $\deg(A) = 0$, $P \in \mathbb{P}_{F/K}$, $\deg(P) = 1$, then there exists unique $Q \in \mathbb{P}_{F/K}$, $\deg(Q) = 1$, such that $A \sim Q - P$.

Proof. The first item follows immediately from Riemann-Roch corollary ($l(A) = \deg(A) + g - 1$ and $g = 1$). From basic properties of divisors we know that there exist $A' \sim A$, such that $A' \geq 0$. Moreover, it satisfies $\deg(A') = \deg(A) = 1$, so $A' = P$ for some $P \in \mathbb{P}_{F/K}$, $\deg(P) = 1$. If P_1 and P_2 were two such possible choices, $P_1 - P_2$ would be principal, ergo $P_1 - P_2 = (t)$ and $P_1 = (t)_+$ for some $t \in F$. This would imply $[F : K(t)] = 1$, $F = K(t)$ and $g = 0$. That is why P is unique. At last, if $\deg(A) = 0$ and $\deg(P) = 1$, then $\deg(P + A) = 1$ and by the second part there exists $Q \in \mathbb{P}_{F/K}$, $\deg(Q) = 1$, such that $P + A \sim Q$ and therefore $A \sim Q - P$. □

Corollary 1.4.7. *Let F/K be an elliptic function field and let $\mathbb{P}^{(1)} = \{Q \in \mathbb{P}_{F/K} : \deg(Q) = 1\}$. Then for any $P \in \mathbb{P}^{(1)}$ the mapping $Q \to Q - P$ is a bijection of $\mathbb{P}^{(1)}$ and $\mathrm{Pic}(F/K)$.*

This corollary at last explains where to get the desired group structure. In other words, elements of Picard group of F/K correspond to places of degree 1, which in turn correspond to points on elliptic curve. Thus we have a group, a set and a bijection, which can easily transfer the group structure. In this case, for every $P \in \mathbb{P}^{(1)}$, $Q_1 \boxplus Q_2 = Q_3 \Leftrightarrow [Q_1 - P] + [Q_2 - P] = [Q_3 - P]$, or $Q_1 \boxplus Q_2 = Q_3$ if and only if $Q_3 \in \mathbb{P}^{(1)}$ fulfills $Q_3 \sim Q_1 + Q_2 - P$.

Let E/K be an elliptic function field defined by a non-singular Weierstrass equation $w(x, y) = 0$. To uniquely determine the bijection between $\mathrm{Pic}(E/K)$ and $\mathbb{P}^{(1)}_{E/K}$, we fix P_∞ as the neutral element. It is the natural choice for it is the only point at infinity of E/K[3]. The newly acquired operation $\oplus$ thus satisfies $(P \oplus Q) + P_\infty \sim P + Q$ in $\mathrm{Pic}(E/K)$.

Definition 1.4.8. The whole just defined group is called an *elliptic group*, denoted by $E(K)$ and formally equal to $(\{P_\infty\} \cup \{P_\alpha : \alpha \in V(w) \cap \mathbb{A}^2(K)\}, \oplus, P_\infty)$.

From the definition and the one-to-one correspondence between P_α and $\alpha \in V(w) \cap \mathbb{A}^2(K)$ it is clear that we can easily switch from adding "places of degree one" to adding "points on curve" and back. The latter case (P_α identified with $\alpha \in V(w) \cap \mathbb{A}^2(K)$ and P_∞ with 0, in this case normally denoted by $\mathcal{O}$) is usually denoted by $w(K)$. Normally, $w(K)$ and $E(K)$ are identified with each other, there is no reasonable argument to distinguish between them. When performing the group operation in $E(K)$, two cases can occur.

Proposition 1.4.9. *Let $P, Q, R \in E(K)$. Then*

(i) $P \oplus Q = P_\infty \Leftrightarrow P + Q - 2P_\infty \in \mathrm{Princ}(E/K)$ *and*

(ii) $P \oplus Q \oplus R = P_\infty \Leftrightarrow P + Q + R - 3P_\infty \in \mathrm{Princ}(E/K)$.

Proof. As for the first case, by definition $P \oplus Q = P_\infty$ if and only if $2P_\infty \sim P + Q$, which is the desired condition. Additionaly, $P \oplus (Q \oplus R) + P_\infty \sim P + (Q \oplus R) \sim P - P_\infty + Q + R$. Therefore $(P \oplus Q \oplus R) + 2P_\infty \sim P + Q + R$ and the proof is complete. □

As we shall shortly hereafter learn, the first case describes adding two points with the same x-coordinate (or in other words, finding an inverse), whereas the latter one handles with adding two arbitrary points (other than in the former case). To deduce exact formulas, we shall use a few new terms and facts.

Definition 1.4.10. Let E/K be an elliptic function field given by $w(x, y) = 0$. A *line* is every $l \in E$, such that there exist $(0, 0) \neq (\lambda_1, \lambda_2) \in K^2$ and $\lambda_3 \in K$: $l = \lambda_1 x + \lambda_2 y + \lambda_3$. A point (α_1, α_2) is on l if and only if $l(\alpha_1, \alpha_2) = 0$. For $\alpha \in V(w)$ and l going through α, l is a tangent to w in α if and only if $\exists \gamma \in K^*$ such that $\lambda_j = \gamma \left(\frac{\partial w}{\partial x_j} \right)$ for both $j \in \{1, 2\}$.

In the following three statements, we make use of l in a little more imaginable form: $l = y - \lambda x - \mu$. The background of all three lemmas is that E/K is an elliptic function field given by a non-singular $w(x, y) = 0$. The goal is to present tools which will help us to determine more concretely the mechanism of addition in $E(K)$.

Lemma 1.4.11. *Let $\alpha \in V(w) \cap \mathbb{A}^2(K)$ and let v_α denote the valuation at P_α. Then for every line $l \in E$ exactly one of the following options is true:*

(i) $v_\alpha = 0$ *and l is not going through α;*

(ii) $v_\alpha = 1$ *and l is not a tangent to w in α;*

[3] See the end of section 1.1.

(iii) $v_\alpha \geq 2$ *and* l *is a tangent to* w *in* α.

Lemma 1.4.12. *Let* $l \in E$ *be a line going through* $\alpha = (\gamma, \lambda\gamma + \mu) \in V(w)$, $\gamma, \lambda, \mu \in K$ *and let* $\partial w/\partial y(\alpha) \neq 0$. *Then* γ *is a root of* $\tau(x) = w(x, \lambda x + \mu)$ *and its multiplicity equals* $v_\alpha(y - \lambda x - \mu)$.

Lemma 1.4.13. *Let all the conditions of the previous lemma be satisfied. Then the line* $l = y - \lambda x - \mu$ *is a tangent to* w *in* α *if and only if* γ *is a multiple root of* $\tau(x) = w(x, \lambda x + \mu)$.

To describe the two possible cases in adding non-zero elements of $E(K)$, we must add a proposition associated to each of them. The first one shall serve to clarify the situation when the sum of two places equals P_∞.

Proposition 1.4.14. *Let* $\alpha = (\alpha_1, \alpha_2) \in V(w) \cap \mathbb{A}^2(K)$. *Set* $\rho(y) = w(\alpha_1, y)$. *Then there exists a unique* $\beta_2 \in K$ *such that* $\rho(y) = (y - \alpha_2)(y - \beta_2)$, $\beta = (\alpha_1, \beta_2) \in V(w) \cap \mathbb{A}^2(K)$ *and* $P_\alpha \oplus P_\beta = P_\infty$. *The case* $\alpha = \beta$ *occurs if and only if* $x = \alpha_1$ *is a tangent to* w *in* α. *Moreover,* $\{\alpha, \beta\} = V(x - \alpha_1, w)$.

Proof. Let $l = x - \alpha_1$. Then $v_Q(l) \geq 0$ for every $Q \neq P_\infty$ (lemma 1.3.1). At the same time, $v_\infty(l) = v_\infty(x) = -2$, so $\deg((l)_+) = 2$. Since $l(\alpha) = \alpha_1 - \alpha_1 = 0$, with respect to lemma 1.4.11, $v_\alpha(l) \geq 1$. Therefore $(l)_+ = P_\alpha + P_\beta$ for some $\beta = (\beta_1, \beta_2) \in V(w) \cap \mathbb{A}^2(K)$, $\deg(P_\alpha) = \deg(P_\beta) = 1$ and $v_\beta(l) \geq 1$. This means that l goes through β, hence $\beta_1 = \alpha_1$. Additionaly, both α_2 and β_2 must be roots of $\rho(y) = w(\alpha_1, y) = 0$. Should $\alpha_2 = \beta_2$, then $v_\alpha(l) = 2$ and α_2 will be a double root of $\rho(y)$ by 1.4.12. It follows that in either case, $\rho(y) = (y - \alpha_2)(y - \beta_2)$.

To sum up, l intersects w in exactly two points, α and β, with $\alpha = \beta$ if and only if l is a tangent to w in α. At last, $(l) = P_\alpha + P_\beta - 2P_\infty$, which according to proposition 1.4.9 means nothing but $P_\alpha \oplus P_\beta = P_\infty$. □

In other words, this proposition gives us the tool to determine β such that $\alpha \oplus \beta = \mathcal{O}$ in $w(K)$. This equation can be restated as $\alpha = \ominus\beta$, which lays down an algorithm to find an inverse. Addition of two general points is described hereafter.

Proposition 1.4.15. *Let* $\alpha, \beta \in V(w) \cap \mathbb{A}^2(K)$, $\alpha = (\alpha_1, \alpha_2)$ *and* $\beta = (\beta_1, \beta_2)$. *Let* $P_\alpha \oplus P_\beta \neq P_\infty$. *If* $\alpha \neq \beta$, *there exist unique* $\lambda, \mu \in K$, $\lambda = \frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1}$, *such that* $l = y - \lambda x - \mu$ *goes through both* α *and* β. *Else (i.e.* $\alpha = \beta$*), there exist unique* $\lambda, \mu \in K$ *such that* $l = y - \lambda x - \mu$ *is a tangent to* w *in* α.

Let $\tau(x) = -w(x, \lambda x + \mu)$. *Then there exists a unique* $\gamma_1 \in K$, *such that* $\tau(x) = (x - \alpha_1)(x - \beta_1)(x - \gamma_1)$. *The point* $\gamma = (\gamma_1, \lambda\gamma_1 + \mu) \in V(w) \cap \mathbb{A}^2(K)$, $P_\alpha \oplus P_\beta \oplus P_\gamma = P_\infty$ *and* $\{\alpha, \beta, \gamma\} = V(w, l)$.

Proof. To exclude unfavourable cases, suppose that $\alpha \neq \beta$, but $\alpha_1 = \beta_1$. This is described by the previous proposition ($P_\alpha + P_\beta = P_\infty$) and does not concern us here. We can therefore assume $\alpha_1 \neq \beta_1$ and find l in an appropriate form. Next, let $\alpha = \beta$. Should $x - \alpha_1$ be a tangent to w in α, we will again find ourselves in the situation of the previous proposition. It follows that the tangent (which must exist according to lemma 1.4.13) has another form, namely that stated here.

Always $v_\infty(l) = v_\infty(y) = -3$ and $v_Q(l) \geq 0$ for any $Q \neq P_\infty$. Moreover, either $v_\alpha(l) \geq 1$ and $v_\beta(l) \geq 1$ (when $\alpha \neq \beta$), or $v_\alpha(l) \geq 2$ (when $\alpha = \beta$). Anyway,

$P_\alpha + P_\beta \leq (l)_+$ and $\deg((l)_+) = 3$. Therefore there exists unique Q such that $P_\alpha + P_\beta + Q = (l)_+$. Its degree must equal 1 and since $Q \neq P_\infty$, $Q = P_\gamma$ for some $\gamma \in V(w) \cap \mathbb{A}^2(K)$. Thus we have obtained a point $\gamma = (\gamma_1, \lambda\gamma_1 + \mu)$ for some uniquely determined $\gamma_1 \in K$.

It is now clear that $\alpha_1, \beta_1, \gamma_1$ are roots of $\tau(x) = -w(x, \lambda x + \mu) = 0$, every one's multiplicity being equal to valuation of l in the respective place. That is why it is possible to write $\tau(x) = (x - \alpha_1)(x - \beta_1)(x - \gamma_1)$. If there was $\delta = (\delta_1, \delta_2) \in V(w, l)$, δ_1 would have to be a root of $\tau(x)$ and $\{\alpha, \beta, \gamma\} = V(w, l)$.

The last statement yet to be proved, that $P_\alpha \oplus P_\beta \oplus P_\gamma = P_\infty$, follows immediately from $P_\alpha + P_\beta + P_\gamma - 3P_\infty = (l)$ and proposition 1.4.9. □

This situation is that $\alpha \oplus \beta \oplus \gamma = \mathcal{O}$ or $\alpha \oplus \beta = \ominus\gamma$. The algorithm to perform addition $\alpha \oplus \beta$, $0 \notin \{\alpha, \beta\}$ would thus be:

(i) determine whether $\alpha = \ominus\beta$. If so, $\alpha \oplus \beta = \mathcal{O}$;

(ii) else, find γ such that $\alpha \oplus \beta \oplus \gamma = \mathcal{O}$;

(iii) find $\ominus\gamma$.

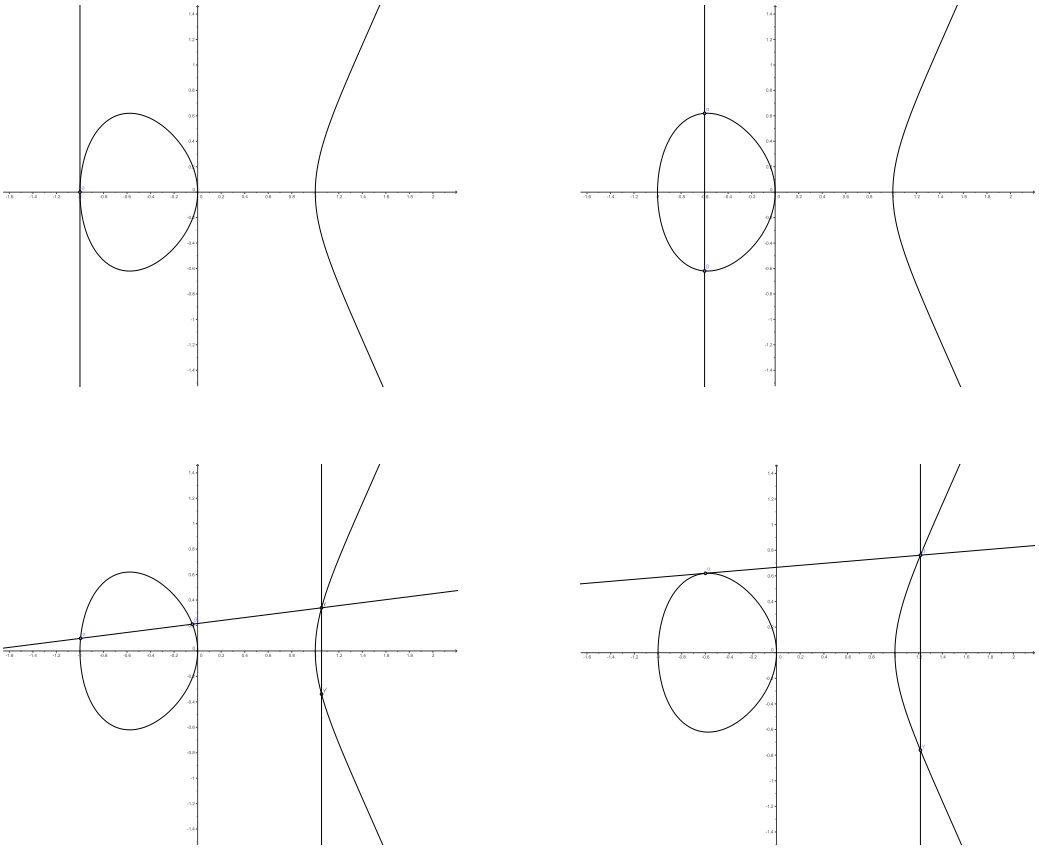


Figure 1.2: All cases that might occur when adding points on an elliptic curve.

Although this casts much light on arithmetics on elliptic curves, it still does not provide us with anything to actually perform computations. To revert this setback, we shall derive them from the long Weierstrass equation and propositions 1.4.14 and 1.4.15.

Let E/K be an elliptic function field defined by a non-singular long Weierstrass equation $w(x,y) = y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6$ and let $\alpha = (\alpha_1, \alpha_2) \in w(K) \setminus \{\mathcal{O}\}$. As in 1.4.14, $\rho(y) = w(\alpha_1, y) = y^2 + (a_1\alpha_1 + a_3)y + \ldots = (y - \alpha_2)(y - \beta_2)$. But such β_2 must fulfill $\alpha_2 + \beta_2 = -a_1\alpha_1 - a_3$. The formula to compute inverse is thus

$$\ominus(\alpha_1, \alpha_2) = (\alpha_1, -\alpha_2 - a_1\alpha_1 - a_3) \text{ for any } (\alpha_1, \alpha_2) \in w(K). \tag{1.2}$$

As a corollary we see that $\alpha \oplus \beta = \mathcal{O}$ if and only if $\alpha_1 = \beta_1$ and $\beta_2 = -\alpha_2 - a_1\alpha_1 - a_3$.

Now let $\alpha = (\alpha_1, \alpha_2)$ and $\beta = (\beta_1, \beta_2)$ be elements of $w(K) \setminus \{\mathcal{O}\}$ such that $\alpha \oplus \beta \neq 0$. If $\alpha \neq \beta$, let $\lambda = \frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1}$ (this is imminent from the fact that λ is the tangent of l). Otherwise, l is a tangent to w in α and from the derivations

$$\begin{aligned} \frac{\partial w}{\partial y} &= 2y + a_1x + a_3, \\ \frac{\partial w}{\partial x} &= a_1y - 3x^2 - 2a_2x - a_4, \end{aligned}$$

follows that

$$\lambda = \frac{3\alpha_1^2 + 2a_2\alpha_1 + a_4 - a_1\alpha_2}{2\alpha_2 + a_1\alpha_1 + a_3}.$$

At last, from the definition of l, $\mu = \alpha_2 - \lambda\alpha_1$. We have thus obtained all the coefficients necessary to compute $\gamma = (\gamma_1, \gamma_2)$. Since

$$\begin{aligned} \tau(x) &= -w(x, \lambda x + \mu) = \\ &= x^3 + (a_2 - \lambda^2 - a_1\lambda)x^2 + (a_4 - 2\lambda\mu - a_3\lambda - a_1\mu) + a_6 - \mu^2 - a_3\mu \end{aligned}$$

and $\alpha_1, \beta_1, \gamma_1$ are its roots, $\alpha_1 + \beta_1 + \gamma_1 = \lambda^2 + a_1\lambda - a_2$. Therefore

$$\begin{aligned} \gamma_1 &= \lambda^2 + a_1\lambda - a_2 - \alpha_1 - \beta_1, \\ \gamma_2 &= \lambda\gamma_1 + \mu = \lambda\gamma_1 + \alpha_2 - \lambda\alpha_1. \end{aligned}$$

To compute $\ominus\gamma$, by equation 1.2 it suffices to replace γ_2 by $-\gamma_2 - a_1\gamma_1 - a_3 = -\lambda\gamma_1 - \alpha_2 + \lambda\alpha_1 - a_1\gamma_1 - a_3 = \lambda(\alpha_1 - \gamma_1) - \alpha_2 - a_1\gamma_1 - a_3$.

These derived formulas deserve their own theorem.

Theorem 1.4.16 (General group law)**.** *Let $\alpha, \beta \in w(K) \setminus \{\mathcal{O}\}$, $\alpha = (\alpha_1, \alpha_2)$, $\beta = (\beta_1, \beta_2)$. Let $\gamma = \alpha \oplus \beta$. Then*

(i) $\gamma = \mathcal{O}$ if and only if $\alpha_1 = \beta_1$ and $\beta_2 + \alpha_2 + a_1\alpha_1 + a_3 = 0$ or

(ii) $\mathcal{O} \neq \gamma = (\gamma_1, \gamma_2)$, where

$$\begin{aligned} \gamma_1 &= \lambda^2 + a_1\lambda - a_2 - \alpha_1 - \beta_1, \\ \gamma_2 &= \lambda(\alpha_1 - \gamma_1) - \alpha_2 - a_1\gamma_1 - a_3, \\ \lambda &= \begin{cases} \frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1} & \text{for } \alpha_1 \neq \beta_1, \\ \frac{3\alpha_1^2 + 2a_2\alpha_1 + a_4 - a_1\alpha_2}{2\alpha_2 + a_1\alpha_1 + a_3} & \text{for } \alpha_1 = \beta_1. \end{cases} \end{aligned}$$

Naturally, since this is a general case functioning for long Weierstrass equations in arbitrary characteristics, we can adjust the formulas for short Weierstrass equations in $\mathrm{char}(K) = 2$ and $\mathrm{char}(K) = p > 3$.

Let us begin with the latter case. According to proposition 1.3.7, $w(x, y) = 0$ is equivalent to $y^2 = x^3 + a_4x + a_6$ for some $a_4, a_6 \in K$ satisfying $4a_4^3 + 27a_6^2 \neq 0$. It follows that $a_1 = a_2 = a_3 = 0$ and the formula can be significantly simplified.

Proposition 1.4.17 (Group law in char > 3)**.** *Let conditions of theorem 1.4.16 be satisfied and let γ denote the sum $\alpha \oplus \beta$. Moreover, let* $\mathrm{char}(K) = p > 3$*. Then*

(i) $\gamma = \mathcal{O}$ if and only if $\alpha_1 = \beta_1$ and $\alpha_2 = -\beta_2$ or

(ii) $\mathcal{O} \neq \gamma = (\gamma_1, \gamma_2)$, where

$$\begin{aligned}
\gamma_1 &= \lambda^2 - \alpha_1 - \beta_1, \\
\gamma_2 &= \lambda(\alpha_1 - \gamma_1) - \alpha_2, \\
\lambda &= \begin{cases} \frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1} & \text{for} \quad \alpha_1 \neq \beta_1, \\ \frac{3\alpha_1^2 + a_4}{2\alpha_2} & \text{for} \quad \alpha_1 = \beta_1. \end{cases}
\end{aligned}$$

These formulas are so simple that there can not be any obstacles implementing them, so the only commentary we add here is the time consumption. In case we add two same points (i.e. we perform *point doubling*), the total cost is $1I + 2M + 2S$, otherwise (*point addition*) $1I + 2M + 1S$. What do these symbols mean? Since we work only on elliptic curves and have not said anything about the underlying field, the most effective (and fair) method to compare different algorithms is to compute their cost in operations in the underlying field, whatever this might be. I will denote field inversion, M field multiplication and S field squaring.

In characteristics 2, according to proposition 1.3.7, we have two different equations defining an elliptic curve:

$$\begin{aligned}
y^2 + xy &= x^3 + a_2x^2 + a_6, \ a_6 \neq 0; \\
y^2 + a_3 &= x^3 + a_4x + a_6, \ a_3 \neq 0.
\end{aligned}$$

The latter one however suffers from a crucial cryptographic weakness: it is supersingular and hence highly susceptible to the so-called MOV attack [7, Chapter V]. Therefore for cryptographic purposes it is useless and we shall turn our attention to the former one, which will constitute the "only" form of an elliptic curve over field of characteristics 2. Naturally, the group law works over supersingular elliptic curves as well and the formulas could be easily derived, but for the sake of brevity (and significance) we shall omit them. Consequently, proposition analogous to 1.4.17 is as follows.

Proposition 1.4.18 (Group law in char $= 2$)**.** *Let conditions of theorem 1.4.16 be satisfied and let γ denote the sum $\alpha \oplus \beta$. Moreover, let* $\mathrm{char}(K) = 2$*. Then*

(i) $\gamma = \mathcal{O}$ if and only if $\alpha_1 = \beta_1$ and $\alpha_2 = -\beta_2$ or

(ii) $\mathcal{O} \neq \gamma = (\gamma_1, \gamma_2)$, where

$$\begin{aligned}
\gamma_1 &= \lambda^2 + \lambda + \alpha_1 + \beta_1 + a_2, \\
\gamma_2 &= \lambda(\alpha_1 + \gamma_1) + \alpha_2 + \gamma_1, \\
\lambda &= \begin{cases} \frac{\beta_2 + \alpha_2}{\beta_1 + \alpha_1} & \text{for} \quad \alpha_1 \neq \beta_1, \\ \frac{\alpha_2}{\alpha_1} + \alpha_1 & \text{for} \quad \alpha_1 = \beta_1. \end{cases}
\end{aligned}$$

Again the time analysis is very simple, each case costs $1I + 2M + 1S$.

2. Redundant point representations

As we have seen at the end of the previous chapter, point addition in affine coordinates requires inverting elements of the underlying field, which is a costly operation. Accordingly, our effort now turns to avoidance of field inversions. The most natural way to achieve it is to leave affine coordinates and take a look at the projective ones. In section 1.1 we have established one-to-one correspondence of affine and projective varieties as well as tools to switch between them, so our task is only to apply this knowledge to the group law.

To start with, we shall homogenize the Weierstrass equation with respect to a new variable Z (note that the standard notation is (x, y) for affine points and $[X : Y : Z]$ for projective ones). We can do the whole process for $\mathrm{char}(K) > 3$ and then turn back and briefly introduce the case $\mathrm{char}(K) = 2$ as well. According to remark 1.1.8 and proposition 1.3.7,

$$w^*(X, Y, Z) = Z^3 w\left(\frac{X}{Z}, \frac{Y}{Z}\right) = Y^2Z - X^3 - a_4XZ^2 - a_6Z^3.$$

Let $P = (x, y) \in V(w) \cap \mathbb{A}^2(K)$ be an affine point. To obtain its projective coordinates $[X : Y : Z]$, we simply apply ϕ_3 defined in remark 1.1.8, i.e. $\phi_3((x, y)) = (x, y, 1)$. This injection naturally does not cost any expensive field operation and its time consumption may hence be neglected. However, the backward process is a different story. If $[X : Y : Z]$ is an arbitrary projective point, to transform it into an affine one we apply ϕ_3^{-1} on it: $\phi_3^{-1}((X, Y, Z)) = \left(\frac{X}{Z}, \frac{Y}{Z}\right)$. This means to find Z^{-1} and multiply both X and Y by it. This costs $1I + 2M$, but on the plus side, it has to be performed only once per whole computation.

At this place, it is convenient to explain special cases. If $Z = 0$, then $w^*(X, Y, Z)$ implies also $X = 0$ and $Y \in K \setminus \{0\}$. Since projective points can be scaled arbitrarily, this point is denoted by $[0 : 1 : 0]$. Transforming it back to affine coordinates, it becomes the point at infinity $\mathcal{O}$. It follows that $\mathcal{O}$ is exactly the only point for which $Z = 0$.

2.1 Standard projective coordinates

This all having said, we can turn our attention to an effective way how to modify the group law to accomodate new coordinates. At first, have a look at the doubling formulas. We have seen that

$$x' = \left(\frac{3x^2 + a_4}{2y}\right)^2 - 2x,$$
$$y' = \left(\frac{3x^2 + a_4}{2y}\right)(x - x') - y.$$

If we substitute $x = X/Z$, $y = Y/Z$, $x' = X'/Z'$ and $y' = Y'/Z'$, we get that

$$\frac{X'}{Z'} = \left(\frac{3\left(X/Z\right)^2 + a_4}{2Y/Z}\right)^2 - 2X/Z = \frac{\frac{W^2}{Z^4}}{\frac{4Y^2}{Z^2}} - \frac{2X}{Z} = \frac{W^2}{4Y^2Z^2} - \frac{2X}{Z},$$
$$\frac{Y'}{Z'} = \left(\frac{\frac{W}{Z^2}}{\frac{2Y}{Z}}\right)\left(\frac{X}{Z} - \frac{X'}{Z'}\right) - \frac{Y}{Z} = \frac{W}{2YZ}\left(\frac{X}{Z} - \frac{X'}{Z'}\right) - \frac{Y}{Z},$$

where $W = 3X^2 + a_4Z^2$. If we substitute X'/Z' in the second equation by the formula in the first one, we immediately see that the largest denominator we can get from expanding the parenthesis is equal to $8Y^3Z^3$. This is nothing but our Z'. Therefore,

$$\begin{aligned} Z' &= 8Y^3Z^3, \\ X' &= 2W^2YZ - 16XY^3Z^2, \\ Y' &= 4XY^2ZW - W^3 + 8XY^2Z, \end{aligned}$$

which can be efficiently implemented as in algorithm 1. The cost of such doubling is $8M + 5S$ or $8M + 3S$, the latter case being for $a_4 = -3$.

Algorithm 1 Standard projective coordinates: point doubling in $\mathrm{char}(K) > 3$

Input: $P = (X, Y, Z)$ a projective point on E
Output: $P' = (X', Y', Z') : P' = 2P$

if $Y = 0$ or $Z = 0$ **then**	
return $\mathcal{O}$	
end if	
if $a_4 = -3$ **then**	
$W = 3 \cdot (X + Z) \cdot (X - Z)$	$1M$
else	
$W = a_4 \cdot Z^2 + 3 \cdot X^2$	$1M + 2S$
end if	
$S = Y \cdot Z$	$1M$
$B = X \cdot Y \cdot S$	$2M$
$H = W^2 - 8 \cdot B$	$1S$
$X' = 2 \cdot H \cdot S$	$1M$
$Y' = W \cdot (4 \cdot B - H) - 8 \cdot Y^2 \cdot S^2$	$2M + 2S$
$Z' = 8 \cdot S^3$	$1M$
return (X', Y', Z')	total: $8M + 3S$ or $8M + 5S$

General point addition requires similar analysis. We shall shorten it and state only the results:

$$\frac{X_3}{Z_3} = \left(\frac{Y_2Z_1 - Y_1Z_2}{X_2Z_1 - X_1Z_2}\right)^2 - \frac{X_1}{Z_1} - \frac{X_2}{Z_2},$$
$$\frac{Y_3}{Z_3} = \left(\frac{Y_2Z_1 - Y_1Z_2}{X_2Z_1 - X_1Z_2}\right)\left(\frac{X_1}{Z_1} - \frac{X_3}{Z_3}\right) - \frac{Y_1}{Z_1}.$$

Again substituting X_3/Z_3 into the second equation shows that $Z_3 = (X_2Z_1 - X_1Z_2)^3 Z_1 Z_2$ yields the desired result, as shown in algorithm 2. Its cost is $12M + 2S$.

Algorithm 2 Standard projective coordinates: point addition in $\mathrm{char}(K) > 3$

Input: $P_1 = (X_1, Y_1, Z_1), P_2 = (X_2, Y_2, Z_2)$ projective points on E
Output: $P_3 = (X_3, Y_3, Z_3) : P_3 = P_1 \oplus P_2$

if $Z_1 = 0$ **then**	
return P_2	
else if $Z_2 = 0$ **then**	
return P_1	
end if	
$U_1 = Y_2 \cdot Z_1$	$1M$
$U_2 = Y_1 \cdot Z_2$	$1M$
$V_1 = X_2 \cdot Z_1$	$1M$
$V_2 = X_1 \cdot Z_2$	$1M$
if $V_1 = V_2$ **then**	
if $U_1 \neq U_2$ **then**	
return $\mathcal{O}$	
else	
return POINT_DOUBLE(X_1, Y_1, Z_1)	$(8M + 5S)$
end if	
end if	
$U = U_1 - U_2$	
$V = V_1 - V_2$	
$W = Z_1 \cdot Z_2$	$1M$
$A = U^2 \cdot W - V^3 - 2 \cdot V^2 \cdot V_2$	$3M + 2S$
$X_3 = V \cdot A$	$1M$
$Y_3 = U \cdot (V^2 \cdot V_2 - A) - V^3 \cdot U_2$	$2M$
$Z_3 = V^3 \cdot W$	$1M$
return (X_3, Y_3, Z_3)	total: $12M + 2S$

The same procedure applies to the case $\mathrm{char}(K) = 2$ as well. For sake of brevity, we do not repeat it in its whole extent. The resultant formulas are:

- **(point doubling)**

$$\begin{aligned} Z' &= (XZ)^3, \\ X' &= XZ(X^2 + YZ)^2 + X^2Z^2(X^2 + YZ) + a_2X^3Z^3, \\ Y' &= X^3Z(X^2 + YZ) + (X^2 + YZ)^3 + XZ(X^3 + YZ)^2 + \\ &\quad a_2X^2Z^2(X^2 + YZ) + X^3YZ^2 + X^4Z^2, \end{aligned}$$

  which can be algorithmized as follows:

$$\begin{array}{lll} A = X^2 & B = A + Y \cdot Z & C = X \cdot Z \\ D = B + C & E = C^2 & F = B \cdot D + a_2 \cdot E \\ X' = C \cdot F & Y' = D \cdot F + A^2 \cdot C & Z' = C \cdot E \end{array}$$

  at the cost $8M + 3S$.

- **(point addition)**

$$\begin{aligned} A &= Y_2Z_1 + Y_1Z_2, \\ B &= X_2Z_1 + X_1Z_2, \\ Z_3 &= B^3Z_1Z_2, \\ X_3 &= ABZ_1Z_2(A + B) + B^3(B + a_2Z_1Z_2), \\ Y_3 &= A^3Z_1Z_2 + X_2AB^2Z_1 + AB^2Z_1Z_2(a_2 + 1) + V^3Z_1(X_2 + Y_2) + \\ &\quad + a_2V^3Z_1Z_2, \end{aligned}$$

  which can be implemented in the following way:

$$\begin{array}{l} A_1 = Y_1 \cdot Z_2; \quad B_1 = X_1 \cdot Z_2; \quad A = A_1 + Y_2 \cdot Z_1; \quad B = B_1 + X_2 \cdot Z_1; \\ C = A + B; \quad D = B^2; \quad E = Z_1 \cdot Z_2; \quad F = B \cdot D; \\ G = (A \cdot C + a_2 \cdot D) \cdot E + F; \quad X_3 = B \cdot G; \\ Y_3 = D \cdot (A \cdot B_1 + B \cdot A_1) + C \cdot G; \quad Z_3 = E \cdot F; \end{array}$$

  costing $15M + 1S$.

It seems that this would be the end of this section, but the converse is true. One interesting case still waits to be resolved. Consider, for example, that we have the task to add two affine points. As we have stated at the beginning of this chapter, transformation from affine to projective coordinates simply requires setting $Z = 1$. But what does this step do with the formulas? How many multiplications render themselves trivial? It follows that fixing $Z = 1$ in the doubling formulas or $Z_1 = 1$ in the addition (or even $Z_1 = Z_2 = 1$) lead to much faster formulas. Of course, we can not use them more than once, for the new Z-coordinate does not stay equal to 1, but it truly is a speed-up.

A comprehensive table containing time consumptions for all the aforementioned cases is given in table 2.1.

operation	$\mathrm{char}(K) > 3$	$\mathrm{char}(K) = 2$
doubling	$8M + 5S$	$8M + 3S$
doubling, $Z = 1$	$3M + 5S$	$6M + 3S$
addition	$12M + 2S$	$15M + 1S$
addition, $Z_1 = 1$	$9M + 2S$	$12M + 1S$
addition, $Z_1 = Z_2 = 1$	$5M + 2S$	$8M + 1S$

Table 2.1: Operations in projective coordinates

operation	$\mathrm{char}(K) > 3$	$\mathrm{char}(K) = 2$
doubling	$1M + 8S$	$5M + 5S$
doubling, $Z = 1$	$1M + 5S$	$2M + 2S$
addition	$8M + 6S$	$15M + 5S$
addition, $Z_1 = 1$	$7M + 4S$	$11M + 3S$
addition, $Z_1 = Z_2 = 1$	$4M + 2S$	n/a

Table 2.2: Operations in Jacobian coordinates

2.2 Jacobian coordinates

Standard projective coordinates elliminate the need to invert field elements, however speedups can be more significant. It makes sense to exploit $[X : Y : Z]$ in a different way: by setting $x = X/Z^2$ and $y = Y/Z^3$ (so-called *weighed projective coordinates* or more commonly *Jacobian coordinates*) we get that

$$Y^2 = X^3 + a_4 X Z^4 + a_6 Z^6.$$

This means that the point at infinity, previously at $[0 : 1 : 0]$ now becomes $[1 : 1 : 0]$. More precisely, it should be $(\gamma^2, \gamma^3, 0)$ for some $\gamma \in K^*$, but since we never actually operate on these coordinates, any triplet with $Z = 0$ (especially $[1 : 1 : 0]$) would do. Conversion from affine to Jacobian coordinates is trivial, the other way round requires $1I + 3M + 1S$:

$$A = 1/Z_1 \qquad B = A^2$$
$$X_3 = B \cdot X_1 \quad Y_3 = A \cdot B \cdot Y_1 \quad Z_3 = 1$$

Again, the whole method of deriving formulas is nothing but expressing X'/Z' and Y'/Z' in terms of input variables and then algorithmizing the resultant polynomial. For sake of brevity, we shall only state the appropriate results. To do so, a table (2.2) seems to be the best method.

Note that n/a does not mean that it is not possible to compute a sum of two points with both Z-coordinates equal to zero, but that there is no algorithm specifically adapted for such an option.

2.3 Chudnovsky coordinates

In 1986, Chudnovsky brothers [9] noticed that in Jacobian coordinates one can save some operation in a particular case. Take for example that one performs an

addition of two points $(X_1 : Y_1 : Z_1) + (X_2 : Y_2 : Z_2)$. During the run of the algorithm, the values Z_2^2 and Z_2^3 are computed, requiring $1M + 1S$. Consider the case that one wants to perform an addition $(X_2 : Y_2 : Z_2) + (X_3 : Y_3 : Z_3)$ for some $(X_3 : Y_3 : Z_3)$. By caching Z_2^2 and Z_2^3 from the previous computation, one can naturally subtract $1M + 1S$ from the operation cost.

There are two basic approaches to this solution. The first one is which the Chudnovsky brothers pursued, namely adding redundancy. The new so-called *Chudnovsky coordinates* therefore are quintuples (X, Y, Z, Z^2, Z^3). It follows that the speedup by $1M + 1S$ is balanced by two more coordinates. Unfortunately, the speedup is only during addition of two distinct points. In point doubling, there is no need to compute Z_2^3 and $1M$ required is wholly unjustified. It therefore seems reasonable to determine Z_2^2 and Z_2^3 only when necessary and mix Chudnovsky coordinates with Jacobian otherwise.

This option can be however described more clearly. Instead of using new coordinate system (which itself is not preferable in all cases), we shall rather introduce a notion of *readdition*. It comprises the situation just explained and its solution is different only in philosophic terms. More concretely, readdition means that we "somewhere" cache the appropriate values only when needed.

Algorithm 3 Chudnovsky Coordinates: Point Doubling

Input: $P = (X, Y, Z, Z^2, Z^3)$
Output: $2P = (X', Y', Z', Z'^2, Z'^3)$

if $Y = 0$ **then**	
return $\mathcal{O}$	
else	
$S \leftarrow 4 \cdot X \cdot Y^2$	$1M + 1S$
if $a \neq -3$ **then**	
$M \leftarrow 3X^2 + a \cdot (Z^2)^2$	$1M + 3S$
else	
$M \leftarrow 3(X + Z^2) \cdot (X - Z^2)$	$1M + 1S$
end if	
$X' \leftarrow M^2 - 2S$	$1S$
$Y' \leftarrow M(S - X') - 8Y^4$	$1M + 1S$
$Z' \leftarrow 2YZ$	$1M$
$(Z')^2 \leftarrow (Z')^2$	$1S$
$(Z')^3 \leftarrow (Z')^2 \cdot Z'$	$1S$
return (X', Y', Z', Z'^2, Z'^3)	total: $\overline{5M + 6S}$
end if	or $5M + 4S$

Algorithm 4 Chudnovsky Coordinates: Point Addition

Input: $P_1 = (X_1, Y_1, Z_1, Z_1^2, Z_1^3)$, $P_2 = (X_2, Y_2, Z_2, Z_2^2, Z_2^3)$
Output: $P_1 \oplus P_2 = P_3 = (X_3, Y_3, Z_3, Z_3^2, Z_3^3)$

if $Z_1 = 0$ or $Z_2 = 0$ **then**	
return $\mathcal{O}$	
end if	
$U_1 \leftarrow X_1 \cdot Z_2^2$	$1M$
$U_2 \leftarrow X_2 \cdot Z_1^2$	$1M$
$S_1 \leftarrow Y_1 \cdot Z_2^3$	$1M$
$S_2 \leftarrow Y_2 \cdot Z_1^3$	$1M$
if $U_1 = U_2$ **then**	
if $S_1 \neq S_2$ **then**	
return $\mathcal{O}$	
else	
return POINT_DOUBLE$(X_1, Y_1, Z_1, Z_1^2, Z_1^3)$	
end if	
end if	
$H \leftarrow U_2 - U_1$	
$R \leftarrow S_2 - S_1$	
$X_3 \leftarrow R^2 - H^3 - 2 \cdot U_1 \cdot H^2$	$2M + 2S$
$Y_3 \leftarrow R \cdot (U_1 \cdot H^2 - X_3) - S_1 \cdot H^3$	$2M$
$Z_3 \leftarrow H \cdot Z_1 \cdot Z_2$	$2M$
$Z_3^2 \leftarrow Z_3^2$	$1S$
$Z_3^3 \leftarrow Z_3^2 \cdot Z_3$	$1M$
return $(X_3, Y_3, Z_3, Z_3^2, Z_3^3)$	total: $\overline{11M + 3S}$

3. Edwards Curves

3.1 Definition and transformation from Weierstrass form

Throughout the 19$^{\text{th}}$ and 20$^{\text{th}}$ century, several new forms of elliptic curves such as Hessian, Montgomery or Jacobi forms emerged. These more or less pushed the speed performance forward, however the most significant breakthrough happened at the very beginning of 2007 [12] when a new curve form emerged and was named after its author – prof. Harold M. Edwards. Within a few months after its original publications, Edwards curves gained the leading position among elliptic curves over both binary and large-p fields.

Definition 3.1.1 (Edwards Form)**.** An Edwards form of an elliptic curve is the equation

$$x^2 + y^2 = a^2(1 + x^2y^2),$$

where $a^5 \neq a$.

At this place, it is convenient to explain why such an equation defines an elliptic curve, along with the condition $a^5 \neq a$. To prove this, we shall need to state a helping proposition.

Proposition 3.1.2. *Let $y^2 = f(x)$ define a plain algebraic curve. Then its genus is 1 if and only if $f(x)$ is of degree 3 or 4 with distinct roots.*

Proof. For a complete proof, see e.g. [13, Section 3.4]. □

Using this proposition, we can prove that under certain circumstances the Edwards Curve is indeed an elliptic curve.

Proposition 3.1.3. *An Edwards curve $x^2+y^2 = a^2(1+x^2y^2)$ is an elliptic curve if and only if $a^5 \neq a$.*

Proof. We can assume $a \neq 0$, for otherwise the equation would not be a curve. Let $z = y(1 - a^2x^2)$, hence $x^2 + y^2 = a^2 + a^2x^2y^2$ transforms into $z^2 = (a^2 - x^2)(1 - a^2x^2)$. Since $a \neq 0$, the right side is a polynomial of degree 4 (namely: $a^2x^4 - (a^4 + 1)x^2 + a^2$). With respect to the previous proposition, we need that polynomial to have distinct roots, which happens if and only if its discriminant equal to $(a^4+1)^2 - 4a^4 = (a^4 - 1)^2$ is nonzero. In other words, $a^5 \neq a$ if and only if $x^2 + y^2 = a^2(1 + x^2y^2)$ is an elliptic curve. □

Our goal is to prove that every Weierstrass curve can be transformed into Edwards curve. For algebraic number fields, this is wholly true as we will see. To be able to prove it, we must introduce a few new terms. At first, since the underlying field's characteristics is 0, we can safely assume the equation to be $y^2 = f(x)$, the constrains on $f(x)$ being the same as in 3.1.2. In this case the field of rational functions is simplified to elements of the form $u(x) + v(x)y$, $u(x), v(x) \in K(x)$. Indeed, it is the appropriate elliptic function field.

Definition 3.1.4. Let E_1/K_1 and E_2/K_2 be two elliptic function fields defined by $y_1^2 = f_1(x_1)$ and $y_2^2 = f_2(x_2)$. Then the curves E_1 and E_2 are *birationally equivalent*, if their fields of rational functions are isomorphic, i.e. when $E_1/K_1 \cong E_2/K_2$. Moreover, we say that elliptic function fields E_1/K_1 and E_2/K_2 are *equivalent* if there exists a field K', such that K' is an algebraic extension of both K_1 and K_2 and $E_1/K' \cong E_2/K'$.

The notion of elliptic function field equivalence is neccesary because of the following fact. In the definition of elliptic function field, we demand that $\tilde{K} = K$, but adjoining new constants (extending K) affects the whole function field and does not allow for birational equivalence, although the field of functions still corresponds to the same curve. This notion will be the main tool to prove that all elliptic curves over number fields can be transformed into Edwards curves.

Theorem 3.1.5. *An elliptic function field is equivalent to the field of rational functions on $x^2 + y^2 = a^2(1 + x^2y^2)$ for some a.*

Proof. Let K be an algebraic number field and let $f(x) \in K[x]$ be a polynomial of degree 4 with distinct roots. If necessary, adjoin constants to K so that $f(x)$ splits into linear factors, i.e. $f(x) = c(x-\alpha_1)(x-\alpha_2)(x-\alpha_3)(x-\alpha_4)$, and perhaps even $\sqrt{c}$ so that $y^2 = f(x)$ transforms into $v^2 = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)(x-\alpha_4)$ for $v = y/\sqrt{c}$. It follows that we may without loss of generality assume that $f(x) = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)(x-\alpha_4)$ for some distinct $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in K$.

Let us have two thusly defined elliptic function fields, say by $z^2 = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)(x-\alpha_4)$ and $v^2 = (u-\beta_1)(u-\beta_2)(u-\beta_3)(u-\beta_4)$. Then they are equivalent if there is a linear transformation $x \to \frac{Ax+B}{Cx+D}$ ($A, B, C, D \in K$ and $AD \neq BC$), which carries α_i to β_i for all $i \in \{1, 2, 3, 4\}$. Denote $u = \frac{Ax+B}{Cx+D}$, then $u - \beta_i = \frac{(AD-BC)(x-\alpha_i)}{(Cx+D)(C\alpha_i+D)}$ for each i. Multiplying all four these subtractions yields $(u-\beta_1)(u-\beta_2)(u-\beta_3)(u-\beta_4) = c\frac{(x-\alpha_1)(x-\alpha_2)(x-\alpha_3)(x-\alpha_4)}{(Cx+D)^4} = c\left(\frac{z}{(Cx+D)^2}\right)^2$ for some $c \in K$ constant. It follows that (perhaps after adjoining $\sqrt{c}$) there is a birational change of variables from (z, x) to (v, u), under which $z^2 = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)(x-\alpha_4)$ corresponds to $v^2 = (u-\beta_1)(u-\beta_2)(u-\beta_3)(u-\beta_4)$.

Consider the transformation

$$x \to \frac{(\alpha_4 - \alpha_2)(x - \alpha_3)}{(\alpha_2 + \alpha_4)(x + \alpha_3) - 2\alpha_3 x - 2\alpha_2\alpha_4}$$

which maps $\alpha_2 \to -1$, $\alpha_3 \to 0$ and $\alpha_4 \to 1$. Therefore $y^2 = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)(x-\alpha_4)$ and $v^2 = (u-\phi)\cdot(u+1)\cdot u\cdot(u-1)$ define the same function field if

$$\begin{aligned}\phi &= \frac{(\alpha_4 - \alpha_2)(\alpha_1 - \alpha_3)}{(\alpha_2 + \alpha_4)(\alpha_1 + \alpha_3) - 2\alpha_3\alpha_1 - 2\alpha_2\alpha_4} \\ &= \frac{\alpha_1\alpha_4 + \alpha_2\alpha_3 - \alpha_1\alpha_2 - \alpha_2\alpha_3}{\alpha_1\alpha_2 + \alpha_1\alpha_4 + \alpha_2\alpha_3 + \alpha_3\alpha_4 + -2\alpha_1\alpha_3 - 2\alpha_2\alpha_4}.\end{aligned}$$

Rewrite $x^2 + y^2 = a^2(1 + x^2y^2)$ as $\left(\frac{y}{a}\right)^2 = (x-a)(x-\frac{1}{a})(x+a)(x+\frac{1}{a})$, then its function field is equivalent to that of $v^2 = (u-\phi)\cdot(u+1)\cdot u\cdot(u-1)$ when $\phi = \frac{-1-1-1-1}{1-1+1-1+2a^2+2a^{-2}} = -\frac{2}{a^2+a^{-2}}$, i.e. a is a solution of $a^4 + \frac{2}{\phi}a^2 + 1 = 0$. In other words, starting from $z^2 = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)(x-\alpha_4)$, we can transform

this equation into $v^2 = (u - \phi) \cdot (u + 1) \cdot u \cdot (u - 1)$, which is in turn equivalent to that of $x^2 + y^2 = a^2(1 + x^2y^2)$ with a as before. The proof is for $\deg(f) = 4$ complete.

If $f(x)$ has distinct roots and $\deg(f) = 3$, we can assume its constant term is nonzero. Otherwise, we replace $f(x)$ by $f(x + c)$. Dividing $z^2 = f(x)$ by x^4 yields $\left(\frac{z}{x^2}\right)^2 = f_1\left(\frac{1}{x}\right)$, where f_1 is a polynomial of degree 4, which is an already known situation. □

Note that this proof was built over number fields. Our attention is, however, focused on finite fields. In non-binary finite fields, it clearly holds as well, but one must be aware of one important aspect, namely that the major part of elliptic curves isomorphism classes might need a field extension to perform the transformation[1]. That is not very useful in limited environments such as smartcards, therefore other ways to extend the number of isomorphism classes available were pursued.

Not more than six months after the original appearance of Edwards curves their generalization was publicised.

Definition 3.1.6 (Generalised Edwards curves.)**.** A generalised Edwards curve is of the form

$$x^2 + y^2 = c^2(1 + dx^2y^2),$$

where $cd(1 - c^4d) \neq 0$.

Remark 3.1.7. Because this definition allows for a wider class of elliptic curves not requiring a field extension, we henceforwards identify Edwards curves with generalized Edward curves (the original ones being with $d = 1$ and $a = c$).

The proof that this equation is an elliptic curve is similar, this time the substitution being $z = y(1 - c^2dx^2)$, the resultant polynomial $f(x) = c^2dx^4 - (1 + c^4d)x^4 + c^2$, whose discriminant equals $(1 + c^4d)^2 - 4c^4d = (1 - c^4d)^2$. The conditions $c \neq 0 \neq d$ are clear.

Definition 3.1.8. Let E be an elliptic curve over K. Let $0 \neq d \in K$, then the *quadratic twist of* E, denoted by E^d, is defined by equation $dy^2 = x^3 + ax + b$.

Remark 3.1.9. Note that the curves E and E^d are isomorphic over $K(\sqrt{d})$. Hence they are birationally equivalent over K if and only if d is a square.

Theorem 3.1.10. *Let K be a finite field,* $\mathrm{char}(K) \neq 2$. *Let E be an elliptic curve over K such that the group $E(K)$ has an element of order* 4 *and a unique element of order* 2. *Then there exists a non-square $d \in K$ such that $x^2 + y^2 = 1 + dx^2y^2$ is birationally equivalent to E over K.*

Proof. Let us start from the long Weierstrass form $s^2 + a_1rs + a_3s = r^3 + a_2r^2 + a_4r + a_6$. Since $\mathrm{char}(K) \neq 2$, we may assume $a_1 = a_3 = 0$ (otherwise, set $\bar{s} = s + (a_1r + a_3)/2$). Let P denote a point of order 4 and $2P = (r_2, s_2)$. If $(r_2, s_2) \neq (0, 0)$, set $\bar{r} = r - r_2$ to remedy this situation. It follows that $a_6 = 0$. We have transformed E into $s^2 = r^3 + a_2r^2 + a_4r$.

[1] According to Bernstein and Lange in [6], more than one quarter of all isomorphism classes of elliptic curves can be transformed to Edwards curves over the same field.

Write P as (r_1, s_1) and note that $s_1 \neq 0$, for otherwise P has order 2. As a consequence, $r_1 \neq 0$ as well. From the derivation of Group Law it is clear that $2P = (0,0)$ is equivalent to the fact that the tangent line to E at P goes through $(0,0)$. In other words, $s_1 - 0 = (r_1 - 0)\lambda$, where $\lambda = (3r_1^2 + 2a_2r_1 + a_4)/2s_1$. Therefore, $3r_1^3 + 2a_2r_1^2 + a_4r_1 = 2s_1^2 = 2r_1^3 + 2a_2r_1^2 + 2a_4r_1$, the second equivalence being derived from the fact that P is on E. Thus $r_1^3 = a_4r_1$, i.e. $r_1^2 = a_4$. Substituting this back to the equation $s_1^2 = r_1^3 + a_2r_1^2 + a_4r_1$ yields $a_2 = s_1^2/r_1^2 - 2r_1$. Putting $d = 1 - 4r_1^3/s_1^2$ we obtain $a_2 = 2r_1((1+d)/(1-d))$.

Note that since $r_1 \neq 0$, $d \neq 1$. Also $d \neq 0$, for otherwise the right side of E's equation would be $r^3 + a_2r^2 + a_4r = r^3 + 2r_1r^2 + r_1^2r = r(r + r_1)^2$ and E would not be an elliptic curve. If d is a square, then apart from $(0,0)$, there is another point of order 2, namely $(r_1(\sqrt{d}+1)/(\sqrt{d}-1), 0)$.

Let E' and E'' be two quadratic twists of E defined by respectively $(r_1/(1-d))s^2 = r^3 + a_2r^2 + a_4$ and $(dr_1/(1-d))s^2 = r^3 + a_2r^2 + a_4r$. Since K is finite and d is a non-square, either $r_1/(1-d)$ or $dr_1/(1-d)$ is a square in K, so E is isomorphic to either E' or E''. Substitute $u = r/r_1$ and $v = s/r_1$ to see that E' is isomorphic to $(1/(1-d))v^2 = u^3 + s((1+d)/(1-d))u^2 + u$ and E'' is isomorphic to $(d(1+d)/(1-d))v^2 = u^3 + 2((1+d)/(1-d))u^2 + u$.

Our effort is now to show that $x^2 + y^2 = 1 + dx^2y^2$ is birationally equivalent to $(1/(1-d))v^2 = u^3 + s((1+d)/(1-d))u^2 + u$ and therefore to E'. The rational map $(u,v) \to (x,y)$ is defined by $x = 2u/v$ and $y = (u-1)/(u+1)$. Only finitely many exceptional points satisfy $v(u+1) = 0$. The inverse rational map $(x,y) \to (u,v)$ is defined by $u = (1+y)(1-y)$ and $v = 2(1+y)/(1-y)x$, again only finitely many exceptions with $(1-y)x = 0$ may occur. By a tedious, yet straightforward calculation one can show that the inverse rational map produces (u,v) satisfying $(1/(1-d))v^2 = u^3 + s((1+d)/(1-d))u^2 + u$.

Substitute $1/d$ for d and $-u$ for u to see that $x^2 + y^2 = 1 + (1/d)x^2y^2$ is birationally equivalent to $(1/(1-1/d))v^2 = (-u)^3 + 2((1+1/d)/(1-1/d))(-u)^2 + (-u)$, i.e. to $(d(1+d)/(1-d))v^2 = u^3 + 2((1+d)/(1-d))u^2 + u$ and consequently to E''. The proof is complete. □

Remark 3.1.11. Theorem 3.1.5 established a possibility of transforming every elliptic curve to an Edwards one, but gave no advice on which are birationally equivalent. Clear criteria are only available due to theorem 3.1.10, which also provides a lower bound on the amount of such elliptic curves. To transform $x^2 + y^2 = 1 + dx^2y^2$ to $\bar{x}^2 + \bar{y}^2 = \bar{c}^2(1 + \bar{d}\bar{x}^2\bar{y}^2)$, simply set $\bar{x} = \bar{c}x$ and $\bar{y} = \bar{c}y$ for some $d = \bar{d}\bar{c}^4$. Since K is finite, at least $1/4$ of its nonzero elements are 4^{th} powers, so $d/\bar{d}$ is a 4^{th} power for at least $1/4$ of $\bar{d} \in K \setminus \{0,1\}$.

3.2 Addition law on Edwards curves

We have already defined Edwards curves and proved at least one quarter of elliptic curves over a non-binary finite field K are birationally equivalent to an appropriate Edwards curve over K. What is left is to establish an addition law and prove that it preserves the group structure introduced in section 1.4.

Definition 3.2.1. Let K be a finite field, $\text{char}(K) \neq 2$. Let $c, d \in K$ satisfy $cd(1 - cd^4) \neq 0$ (i.e. $E : x^2 + y^2 = c^2(1 + dx^2y^2)$ is an Edwards curve over K).

By an *Edwards addition law* we mean the map

$$(x_1, y_1), (x_2, y_2) \longmapsto \left(\frac{x_1y_2 + y_1x_2}{c(1 + dx_1x_2y_1y_2)}, \frac{y_1y_2 - x_1x_2}{c(1 - dx_1x_2y_1y_2)} \right). \tag{3.1}$$

Remark 3.2.2. By a straightforward calculation, it is easy to verify that for any $P = (x_1, y_1)$ on the curve, $P = P + (0, c)$, therefore $(0, c)$ is a neutral element of the addition law. Moreover, it is unique. Additionally, $-P = (-x_1, y_1)$:

$$\begin{aligned}(x_1, y_1), (-x_1, y_1) \longmapsto & \left(\frac{x_1y_1 + y_1(-x_1)}{c(1 - dx_1^2y_1^2)}, \frac{y_1^2 + x_1^2}{c(1 + dx_1^2y_1^2)} \right) \\ &= \left(\frac{0}{c(1 - dx_1^2y_1^2)}, \frac{c^2(1 + dx_1^2y_1^2)}{c(1 + dx_1^2y_1^2)} \right) = (0, c).\end{aligned}$$

To finish classification of points on Edwards curves, note that $(0, -c)$ has order 2 and $(c, 0)$ and $(-c, 0)$ have order 4. This is essentialy the reason why we demand that a point of order 4 exists on an elliptic curve before transforming it in theorem 3.1.10. If such a point does not exists on E, we have to construct a field extension K' such that $E(K')$ has an element of order 4 and then apply theorem 3.1.10 to E over K'.

At first it is appropriate to prove that the map is well-defined, i.e. that the denominators never vanish. This is the core of the following proposition.

Proposition 3.2.3. *Let K be a field,* $\mathrm{char}(K) \neq 2$. *Let $c, d, e \in K$ be nonzero elements with $e = 1 - cd^4$. Assume d is not a square in K. Let $x_1, y_1, x_2, y_1 \in K$ satisfy $x_1^2 + y_1^2 = c^2(1 + dx_1^2y_1^2)$ and $x_2^2 + y_2^2 = c^2(1 + dx_2^2y_2^2)$. Then $dx_1x_2y_1y_2 \notin \{-1, 1\}$.*

Proof. Write $\epsilon = dx_1x_2y_1y_2$ and suppose that $\epsilon \in \{-1, 1\}$. Then $x_1, x_2, y_1, y_2 \neq 0$. Additionaly, $dx_1^2y_1^2(x_2^2 + y_2^2) = dx_1^2y_1^2[c^2(1 + dx_2^2y_2^2)] = c^2(dx_1^2y_1^2 + d^2x_1^2y_1^2x_2^2y_2^2) = c^2(dx_1^2y_1^2 + \epsilon^2) = c^2(1 + dx_1^2y_1^2) = x_1^2 + y_1^2$. So

$$\begin{aligned}(x_1 + \epsilon y_1)^2 = x_1^2 + y_1^2 + 2\epsilon x_1y_1 &= dx_1^2y_1^2(x_2^2 + y_2^2) + 2x_1y_1dx_1x_2y_1y_2 \\ &= dx_1^2y_1^2(x_2^2 + 2x_2y_2 + y_2^2) = dx_1^2y_1^2(x_2 + y_2)^2.\end{aligned}$$

If $x_2 + y_2 \neq 0$ then $d = ((x_1 + \epsilon y_1)/x_1y_1(x_2 + y_2))^2$ and d is a square in K, contradiction. Also when $x_2 - y - 2 \neq 0$, then $d = ((x_1 - \epsilon y_1)/x_1y_1(x_2 - y_2))^2$, so d is a square, again a contradiction. Therefore both $x_2 + y_2$ and $x_2 - y_2$ are 0, but this is possible only when $x_2 = y_2 = 0$, contradiction. □

We have proved that the map is defined in all cases, but it is not natural that the image is on E as well.

Theorem 3.2.4. *Let K be a field,* $\mathrm{char}(K) \neq 2$, *let $c, d \in$ be nonzero with $cd^4 \neq 1$. Let $x_1, x_2, y_1, y_2 \in K$ satisfy $x_1^2 + y_1^2 = c^2(1 + dx_1^2y_1^2)$ and $x_2^2 + y_2^2 = c^2(1 + dx_2^2y_2^2)$. Define x_3, y_3 as in* (3.1). *Then $x_3^2 + y_3^2 = c^2(1 + dx_3^2y_3^2)$.*

Proof. Define $T = (x_1y_2 + y_1x_2)^2(1 - dx_1x_2y_1y_2)^2 + (y_1y_2 - x_1x_2)^2(1 + dx_1x_2y_1y_2)^2$, by a long series of equations it is possible to verify that $T = (x_1^2 + y_1^2 - (x_2^2 + y_2^2)dx_1^2y_1^2)(x_2^2 + y_2^2 - (x_1^2 + y_1^2)dx_2^2y_2^2)$.

As the next step, subtract $(x_2^2+y_2^2)dx_1^2y_1^2 = c^2(1+dx_2^2y_2^2)dx_1^2y_1^2$ from $x_1^2+y_1^2 = c^2(1+dx_1^2y_1^2)$ to see that $x_1^2+y_1^2-(x_2^2+y_2^2)dx_1^2y_1^2 = c^2(1-d^2x_1^2x_2^2y_1^2y_2^2)$. Switching the role of both equations yields $x_2^2+y_2^2-(x_1^2+y_1^2)dx_2^2y_2^2 = c^2(1-d^2x_1^2x_2^2y_1^2y_2^2)$. It follows that $T = c^4(1-d^2x_1^2x_2^2y_1^2y_2^2)^2$.

It remains to substitute into the Edwards addition law. We have

$$\begin{aligned} x_3^2+y_3^2-c^2dx_3^2y_3^2 &= \frac{(x_1y_2+y_1x_2)^2}{c^2(1+dx_1x_2y_1y_2)^2}+\frac{(y_1y_2-x_1x_2)^2}{c^2(1-dx_1x_2y_1y_2)^2} \\ -\frac{c^2d(x_1y_2+y_1x_2)^2(y_1y_2-x_1x_2)^2}{c^4(1+dx_1x_2y_1y_2)^2(1-dx_1x_2y_1y_2)^2} &= \frac{T}{c^2(1+dx_1x_2y_1y_2)^2(1-dx_1x_2y_1y_2)^2} \\ = \frac{T}{c^2(1-d^2x_1^2x_2^2y_1^2y_2^2)} &= c^2, \end{aligned}$$

which proves that $x_3^2+y_3^2 = c^2(1+dx_3^2y_3^2)$ as desired. □

Therefore, we have proved that the total of two points is again a point on the curve. To complete the construction, we need that this addition law corresponds to one on the birationally equivalent curve.

Theorem 3.2.5. *Let conditions of theorem 3.2.4 be satisfied. Let $e = 1-cd^4$ and let E be the elliptic curve $(1/e)v^2 = u^3+(4/e-2)u^2+u$. For each $i \in \{1,2,3\}$ define P_i as follows: $P_i = \infty$ if $(x_i,y_i) = (0,c)$; $P_i = (0,0)$ if $(x_i,y_i) = (0,-c)$ and $P_i = (u_i,v_i)$ if $x_i \neq 0$, where $u_i = (c+y_i)/(c-y_i)$ and $v_i = 2c(c+y_i)/(c-y_i)x_i$. Then $P_i \in E(K)$ and $P_1+P_2 = P_3$.*

Proof. The first task is to show that $P_i \in E(K)$. If $(x_i,y_i) = (0,c)$, then $P_i = \infty \in E(K)$. If $(x_i,y_i) = (0,-c)$, then $P_i = (0,0) \in E(K)$. Otherwise, $P_i = (u_i,v_i) \in E(K)$ by theorem 3.1.10.

To prove that also $P_1+P_2 = P_3$, we must split the proof into several parts. If $(x_1,y_1) = (0,c)$ then readily $(x_2,y_2) = (x_3,y_3)$. P_1 is the point at infinity and $P_2 = P_3$, so $P_1+P_2 = \infty+P_2 = P_2 = P_3$; similarly when $(x_2,y_2) = (0,c)$. Onwards we will assume that $(x_1,y_1) \neq (0,c) \neq (x_2,y_2)$. If $(x_3,y_3) = (0,c)$, then $(x_2,y_2) = (-x_1,y_1)$. If $(x_1,y_1) = (0,-c)$ then also $(x_2,y_2) = (0,-c)$ and $P_1 = P_2 = (0,0)$; otherwise x_1, x_2 are nonzero. Thus $u_1 = (c+y_1)/(c-y_1) = u_2$ and $v_1 = 2cu_1/x_1 = -2cu_2/x_2 = -v_2$, so $P_1 = -P_2$. We have dealt with all situations when one of the three points is $(0,c)$.

If $(x_1,y_1) = (0,-c)$ then $(x_3,y_3) = (-x_2,-y_2)$. Now $(x_2,y_2) \neq (0,c)$ and $(x_2,y_2) \neq (0,-c)$ (in that case $(x_3,y_3) = (0,c)$, contradiction), so $x_2 \neq 0$. Thus $P_1 = (0,0)$ and $P_2 = (u_2,v_2)$ where $u_2 = (c+y_2)/(c-y_2)$ and $v_2 = 2cu_2/x_2$. By the standard addition $(0,0)+(u_2,v_2) = (r_3,s_3)$ for $r_3 = (1/e)(v_2/u_2)^2 - (4/e-2) - u_3 = 1/u_2$ and $s_3 = (u_2/v_2)(-r_3) = -v_2/u_2^2$. At the same time, $P_3 = (u_3,v_3)$ where $u_3 = (c+y_3)/(c-y_3) = (c-y_2)/(c+y_2) = 1/u_2 = r_3$ and $v_3 = 2cu_3/x_3 = -2c/u_2x_2 = -v_2/u_2^2 = s_3$. Consequently, $P_1+P_2 = P_3$ – similar process is available when $(x_2,y_2) = (0,-c)$.

Assume for the rest of the proof that $x_1 \neq 0 \neq x_2$. If $(x_3,y_3) = (0,-c)$ then $(x_1,y_1) = (x_2,-y_2)$ so $u_1 = (c+y_1)/(c-y_1) = (c-y_2)/(c+y_2) = 1/u_2$ and $v_1 = 2cu_1/x_1 = v_2/u_2^2$. Moreover, $P_3 = (0,0)$, so the addition law says that $-P_3+P_2 = (0,0)+P_2 = (1/u_2,-v_2/u_2^2) = (u_1,-v_1) = -P_1$, i.e. that $P_1+P_2 = P_3$. We can now assume that also $x_3 \neq 0$, hence $P_3 = (u_3,v_3)$ with $u_3 = (c+y_3)/(c-y_3)$ and $v_3 = 2cu_3/x_3$.

If $P_1 = -P_2$, then $u_1 = u_2$ and $v_2 = -v_1$, so $x_2 = -x_1$ and $y_2 = c(u_2 - 1)/(u_2 + 1) = c(u_1 - 1)/(u_1 + 1) = y_1$, so $(x_3, y_3) = (0, c)$ which has already been discussed. Assume that $P_1 \neq -P_2$. The two last remaining cases, namely (1) $u_1 = u_2$ and $v_1 \neq -v_2$ and (2) $u_1 \neq u_2$, are possible to verify through a straightforward calculation. □

We have proved that $P_1 + P_2 = P_3$ in any case, whatever the input points are. This property is called *completeness of the addition formulas*. It is particularly important, since the algorithms for Weierstrass curves required testing for point doubling or addition or special cases (one of the points being the point at infinity or the equation $P_1 = -P_2$). All this can be omitted and the implementor can only use just this one formula. We don't have to emphasize the effect Edwards addition law has on the sensitive side-channel attacks, since all computations (additions and doublings) might be performed by the same algorithm at the same cost of time and memory.

However, a dedicated algorithm for point doubling is both available and faster. When $(x_1, y_1) = (x_2, y_2)$, the first denominator becomes $c(1 + dx_1^2y_1^2)$, which can be rewritten as $(x_1^2 + y_1^2)/c$. The second one can be expressed as $c(1 - dx_1^2y_1^2) = c(1 + dx_1^2y_1^2) - 2cdx_1^2y_1^2 = (x_1^2 + y_1^2)/c - 2c - 2(x_1^2 + y_1^2)/c = (2c^2 - (x_1^2 + y_1^2))/c$. Therefore,

$$2(x_1, y_1) = \left(\frac{2x_1y_1}{c(1 + dx_1^2y_1^2)}, \frac{y_1^2 - x_1^2}{c(1 - dx_1^2y_1^2)}\right) = \left(\frac{2x_1y_1c}{x_1^2 + y_1^2}, \frac{(y_1^2 - x_1^2)c}{2c^2 - (x_1^2 - y_1^2)}\right). \quad (3.2)$$

Even more speed can be gained by rewriting $2x_1y_1$ as $(x_1 + y_1)^2 - x_1^2 - y_1^2$.

Affine coordinates. Points are represented by a double (x, y), the point at infinity is $(0, c)$, $-(x, y) = (-x, y)$.

- *addition*: $2I + 6M + 2c + 1d$

  $A = x_1 \cdot x_2; \quad B = y_1 \cdot y_2; \quad C = A - B;$
  $D = (x_1 + y_1) \cdot (x_2 + y_2) - A - B; \quad E = d \cdot A \cdot B; \quad F = c \cdot (1 + E);$
  $G = c \cdot (1 - E); \quad H = 1/F; \quad I = 1/G;$
  $x_3 = D \cdot H; \quad y_3 = C \cdot I.$

- *doubling*: $2I + 2M + 4S + 1c$

  $A = x_1^2; \quad B = y_1^2; \quad C = (x_1 + y_1)^2; \quad D = A + B; \quad E = 2c^2 - D;$
  $F = 1/D; \quad G = 1/E; \quad H = C - D; \quad I = c(B - A);$
  $x_3 = H \cdot F; \quad y_3 = I \cdot G.$

Projective coordinates. Points are represented by a triple $(X : Y : Z)$, if $Z \neq 0$, they correspond to the affine point $(X/Z, Y/Z)$. The neutral element is $(0 : c : 1)$ and inverse to $(X : Y : Z)$ is $(-X : Y : Z)$.

- *addition*: $10M + 1S + 1c + 1d$

  $A = Z_1 \cdot Z_2; \quad B = A^2; \quad C = X_1 \cdot X_2; \quad D = Y_1 \cdot Y_2; \quad E = d \cdot C \cdot D;$
  $F = B - E; \quad G = B + E; \quad X_3 = A \cdot F \cdot ((X_1 + Y_1) \cdot (X_2 + Y_2) - C - D);$
  $Y_3 = A \cdot G \cdot (D - C); \quad Z_3 = c \cdot F \cdot G.$

- *addition (alternative)*: $7M + 5S + 1c + 1d$. Looking at the algorithm used for computing point addition, it is possible to trade some multiplications for squarings. More precisely, $A(B-E)$, $A(B+E)$ and $(B-E)(B+E)$ can be obtained from A^2 $(= B)$, B^2, E^2, $(A+B)^2$ and $(A+E)^2$. Therefore $3M$ is replaced by $4S$, which is advantageous if $S/M < 0.75$[2].

- *mixed addition*: this is the case when $Z_2 = 1$, i.e. the calculation $A = Z_1 \cdot Z_2$ can be omitted, saving $1M$.

- *doubling*: $3M + 4S + 3c$

$$B = (X_1 + Y_1)^2; \quad C = X_1^2; \quad D = Y_1^2; \quad E = C + D; \quad H = (c \cdot Z_1)^2;$$
$$J = E - 2H; \quad X_3 = c \cdot (B - E) \cdot J; \quad Y_3 = c \cdot E \cdot (C - D); \quad Z_3 = E \cdot J.$$

- *readdition*: there are no calculations depending on only one point's coordinates, hence it is not possible to save time by caching them.

3.3 Binary Edwards Curves

Since the very beginning of this chapter, we have assumed that $\mathrm{char}(K) \neq 2$ (otherwise a general elliptic curve would not have the shape $y^2 = f(x)$ for some $f \in K[x]$ of degree 3 or 4). But the advantages of Edwards curves are so prominent that their analogue over a binary field would be desirable. This effort was not surprisingly taken by D. Bernstein and T. Lange in cooperation with R. Farashahi within a year of appearance of the original Edwards Curves.

Definition 3.3.1. Let K be a field with $\mathrm{char}(K) = 2$. Let $d_1, d_2 \in K$ such that $d_1 \neq 0$ and $d_2 \neq d_1^2 + d_1$. The *binary Edwards curve with coefficients* d_1 *and* d_2 is the affine curve

$$E_{B,d_1,d_2}: \quad d_1(x+y) + d_2(x^2+y^2) = xy + xy(x+y) + x^2y^2.$$

Remark 3.3.2. One can immediately notice that this curve is symmetric in x and y, i.e. if (x_1, y_1) is a point on the curve, then (y_1, x_1) lies on it as well.

Theorem 3.3.3. *Every binary Edwards curve is nonsingular.*

Proof. As follows from the definition, $d_1 \neq 0$ and $d_2 \neq d_1^2 + d_1$. The partial derivations of the curve equation are $\partial E/\partial x = d_1 + y + y^2$ and $\partial E/\partial y = d_1 + x + x^2$. By definition, a singular point (x_1, y_1) must satisfy $d_1 + x_1 + x_1^2 = 0$ and $d_1 + y_1 + y_1^2 = 0$. Hence, $x_1 + y_1 = x_1^2 + y_1^2 = (x_1 + y_1)^2$, so $x_1 = y_1$ or $x_1 = y_1 + 1$.

In the former case, substituting $x_1 = y_1$ into the curve equation yields $0 = x_1^2 + x_1^4$. But since $d_1 + x_1 + x_1^2 = 0$, $d_1^2 = x_1^2 + x_1^4 = 0$, contradiction. The latter case implies $d_1 + d_2 = y_1^2 + y_1^4$ and again using the partial derivation, $d_1^2 = y_1^2 + y_1^4 = d_1 + d_2$, contradicting $d_2 \neq d_1 + d_1^2$. □

[2]For example, according to [8], the S/M ratio in NIST prime fields is about 0.8. This assumption is in general justifiable for random primes. However, for some special primes the modular reduction can be rendered negligible, allowing for the S/M ratio of about 0.6 (this is the case of Mersenne primes, for instance). For example, the *Curve25519* has $S/M \approx 0.67$, see [4].

Recall that the trasformation of elliptic curves to Edwards curves over non-binary fields was only possible if their fields of rational functions were birationally equivalent (generally, they were all so over an extension of the underlying field). We would like to know which binary elliptic curves might be transformed to binary Edwards curves as well. Since we do not consider supersingular curves here, an ordinary binary elliptic curve can be written as $v^2 + uv = u^3 + a_2u^2 + a_6$ where $a_6 \neq 0$. Consider the map $(x, y) \mapsto (u, v)$ defined by

$$u = \frac{d_1(d_1^2 + d_1 + d_2)(x + y)}{(xy + d_1(x + y))},$$
$$v = d_1(d_1^2 + d_1 + d_2)\left(\frac{x}{xy + d_1(x + y)} + d_1 + 1\right),$$

it constitutes a birational equivalence from E_{B,d_1,d_2} to the elliptic curve

$$v^2 + uv = u^3 + (d_1^2 + d_2)u^2 + d_1^4(d_1^4 + d_1^2 + d_2^2).$$

An inverse map is as follows:

$$x = \frac{d_1(u + d_1^2 + d_1 + d_2)}{u + v + (d_1^2 + d_1)(d_1^2 + d_1 + d_2)},$$
$$y = \frac{d_1(u + d_1^2 + d_1 + d_2)}{v + (d_1^2 + d_1)(d_1^2 + d_1 + d_2)}.$$

The rational map $(x, y) \mapsto (u, v)$ has only one exceptional case: $(0, 0)$. Let φ be an extension of this map and define $\varphi(0, 0) = P_\infty$. Then φ is a function on all affine points of E_{B,d_1,d_2}. (If $xy + d_1(x + y) = 0$, i.e. $xy = d_1(x + y)$, then $d_2(x^2 + y^2) = xy(x+y)+x^2y^2 = d_1(x+y)^2+d_1^2(x+y)^2$, therefore $(d_2+d_1^2+d_1)(x^2+y^2) = 0$, so $x^2 + y^2 = 0$ and $x = y$. But substituting back to $xy = d_1(x + y)$ leads to $xy = 0$, so $x^2 = 0$ and $(x, y) = (0, 0)$.)

Thus binary Edwards curves are birationally equivalent to ordinary binary Weierstrass curves, which is what we set on to prove.

Definition 3.3.4. Let E_{B,d_1,d_2} be a binary Edwards curve. Let (x_1, x_2) and (x_2, y_2) be points on that curve. Define $(x_3, y_3) = (x_1, y_1) + (x_2 + y_2)$ where

$$x_3 = \frac{d_1(x_1 + x_2) + d_2(x_1 + y_1)(x_2 + y_2) + (x_1 + x_1^2)(x_2(y_1 + y_2 + 1) + y_1y_2)}{d_1 + (x_1 + x_1^2)(x_2 + y_2)},$$
$$y_3 = \frac{d_1(y_1 + y_2) + d_2(x_1 + y_1)(x_2 + y_2) + (y_1 + y_1^2)(y_2(x_1 + x_2 + 1) + x_1x_2)}{d_1 + (y_1 + y_1^2)(x_2 + y_2)}.$$

The proof that this addition law is indeed a group law and the resultant group is isomorphic to the group on Weierstrass curve (by the map φ defined above) is rather technical and does not involve deep mathematics. We shall turn our attention to special aspects of binary Edwards curves, namely their class allowing again for complete addition law.

Theorem 3.3.5. *Let K be a field,* $\mathrm{char}(K) = 2$. *Let $d_1, d_2 \in K$, $d_1 \neq 0$. Suppose that $t^2 + t + d_2 \neq 0$ for all $t \in K$. Then the addition law on the binary Edwards curve E_{B,d_1,d_2} is complete.*

Proof. We must show that the denominators $d_1+(x_1+x_1^2)(x_2+y_2)$ and $d_1+(y_1+y_1^2)(x_2+y_2)$ are nonzero for all $(x_1,y_1),(x_2,y_2)\in E_{B,d_1,d_2}(K)$. Should $x_2=y_2$, both denominators become d_1, which is nonzero by assumption. From now on, suppose that $x_2\neq y_2$ and $d_1+(x_1+x_1^2)(x_2+y_2)=0$, i.e. $d_1/(x_2+y_2)=x_1+x_1^2$.

From the curve equation it follows that

$$\begin{aligned}\frac{d_1}{x_2+y_2}&=\frac{d_1(x_2+y_2)}{x_2^2+y_2^2}=\frac{d_2(x_2^2+y_2^2)+x_2y_2+x_2y_2(x_2+y_2)+x_2^2+y_2^2}{x_2+y_2}\\&=d_2+\frac{x_2y_2+x_2y_2(x_2+y_2)+y_2^2}{x_2^2+y_2^2}+\frac{y_2^2+x_2^2y_2^2}{x_2^2+y_2^2}\\&=d_2+\frac{y_2+x_2y_2}{x_2+y_2}+\frac{y_2^2+x_2^2y_2^2}{x_2^2+y_2^2}\quad(=x_1+x_1^2).\end{aligned}$$

Consequently, for $t=x_1+(y_2+x_2y_2)/(x_2+y_2)\in K$, $t^2+t+d_2=0$, contradiction. Therefore $d_1+(x_1+x_1^2)(x_2+y_2)\neq 0$. The second case is similar. □

This theorem gives a clear clue which binary Edwards curves are favourable and motivates the following definition.

Definition 3.3.6 (Complete binary Edwards curve)**.** Let K be a field, $\mathrm{char}(K)=2$. Let $d_1,d_2\in K$ such that $d_1\neq 0$. Assume that $t^2+t+d_2\neq 0$ for all $t\in K$. The *complete binary Edwards curve with coefficients d_1 and d_2* is the affine curve

$$E_{B,d_1,d_2}: d_1(x+y)+d_2(x^2+y^2)=xy+xy(x+y)+x^2y^2.$$

Remark 3.3.7. Note that this is nothing else but a constraint on the previous definition. Instead of requiring $d_1^2+d_1+d_2\neq 0$, we insist that this condition holds for all $t\in K$, not just d_1. Over $\mathbb{F}_{2^n}$ it is equivalent to $\mathrm{Tr}_{\mathbb{F}_{2^n}/\mathbb{F}_2}(d_2)=1$.

To prove that complete binary Edwards curves are still general enough, we need the following theorem.

Theorem 3.3.8. *Let n be an integer, $n\geq 3$. Each ordinary elliptic curve over $\mathbb{F}_{2^n}$ is birationally equivalent over $\mathbb{F}_{2^n}$ to a complete binary Edwards curve.*

Proof. At first we use the knowledge from chapter 1 to say that every ordinary elliptic curve over $\mathbb{F}_{2^n}$ is isomorphic to $v^2+uv=u^3+a_2u^2+a_6$ for some $a_2\in\mathbb{F}_{2^n}$ and $a_6\in\mathbb{F}_{2^n}^*$. Note that if $\mathrm{Tr}(a_2)=\mathrm{Tr}(a_2')$, then there exists $b\in\mathbb{F}_{2^n}$ such that $a_2'=a_2+b+b^2$ and $v\mapsto v+bu$ is an isomorphism from $v^2+uv=u^3+a_2u^2+a_6$ to $v^2+uv=u^3+(a_2+b+b^2)u^2+a_6$. Fix a_2 and a_6 for the rest of the proof.

Define

$$D_{\delta,\epsilon}=\{d_1\in\mathbb{F}_{2^n}^*:\ \mathrm{Tr}(d_1)=\delta,\ \mathrm{Tr}(\sqrt{a_6}/d_1^2)=\epsilon\}$$

for every $\delta,\epsilon\in\mathbb{F}_2$. If $d_1\in D_{\mathrm{Tr}(a_2)+1,1}$, then the pair (d_1,d_2) (where $d_2=d_1^2+d_1+\sqrt{a_6}/d_1^2$) fulfils $\mathrm{Tr}(d_2)=\mathrm{Tr}(\sqrt{a_6}/d_1^2)=1$ and hence defines a complete binary Edwards curve E_{B,d_1,d_2}. Moreover, since $d_1^4(d_1^4+d_1^2+d_2^2)=a_6$ from the definition of d_2, this curve is birationally equivalent to $v^2+uv=u^3+(d_1^2+d_2)u^2+a_6$ and therefore to $v^2+uv=u^3+a_2u^2+a_6$ as well, for $\mathrm{Tr}(d_1^2+d_2)=\mathrm{Tr}(d_1)+\mathrm{Tr}(d_2)=\mathrm{Tr}(a_2)+1+1=\mathrm{Tr}(a_2)$.

We must show that $D_{\mathrm{Tr}(a_2)+1,1}$ is nonempty, which can be achieved by counting elements of D_{01} and D_{11}. At first, $\#D_{00}+\#D_{01}$ is the number of $d_1\in\mathbb{F}_{2^n}^*$ such that $\mathrm{Tr}(d_1)=0$, which implies $\#D_{00}+\#D_{01}=2^{n-1}-1$. Next, $\#D_{01}+\#D_{11}=$

2^{n-1}, since as d_1 runs through $\mathbb{F}_{2^n}^*$, so does $\sqrt{a_6}/d_1^2$, so its trace equals 1 exactly 2^{n-1} times.

Now consider $d_1 \in \mathbb{F}_{2^n}^*$ such that $\mathrm{Tr}(d_1 + \sqrt{a_6}/d_1^2) = 0$. $\#D_{00} + \#D_{11}$ clearly denotes number of such d_1s and for every one there exist exactly two $s \in \mathbb{F}_{2^n}$ such that $s^2 + s = d_1 + \sqrt{a_6}/d_1^2$, constituting two points $(U_1, V_1) = (d_1, d_1 s)$ on the elliptic curve $V^2 + UV = U^3 + \sqrt{a_6}$. By Hasse's theorem, this curve has $2^n + 1 + t$ points for some $t \in [-2\sqrt{2^n}, 2\sqrt{2^n}]$. With the exception of the point at infinity and $(0, 0)$, every other point can be uniquely derived as described, so $\#D_{00} + \#D_{11} = 2^{n-1} + (t-1)/2$.

Substituting back to the two already established equations, $2\#D_{01} = (\#D_{00} + \#D_{01}) + (\#D_{01} + \#D_{11}) - (\#D_{00} + \#D_{11}) = 2^{n-1} - 1 + 2^{n-1} - 2^{n-1} - (t-1)/2 = 2^{n-1} - (t-1)/2$ and $2\#D_{11} = 2^n - 2\#D_{01} = 2^{n-1} + (t+1)/2$. We can suppose $n \geq 3$, hence $(\sqrt{2^n} - 1)^2 \geq (\sqrt{8} - 1)^2 > 2$. Therefore, $2^n > 2\sqrt{2^n} + 1 \geq |t| + 1$ and both D_{01} and D_{11} are nonempty. □

Remark 3.3.9. The construction thus goes as follows: given an ordinary binary Weierstrass curve with coefficients a_2 and a_6, we choose an arbitrary d_1 satisfying $\mathrm{Tr}(d_1) = \mathrm{Tr}(a_2) + 1$ and $\mathrm{Tr}(\sqrt{a_6}/d_1^2) = 1$. If so, we compute $d_2 = d_1^2 + d_1 + \sqrt{a_6}/d_1^2$. We finish having a complete binary Edwards curve E_{B,d_1,d_2} birationally equivalent to the original curve. The theorem says that we can do this procedure for at least one d_1, but much stronger proposition is at hand: in fact, about fifty per cent of d_1 with $\mathrm{Tr}(d_1) = \mathrm{Tr}(a_2) + 1$ are eligible, giving the implementors a wide choice of appropriate d_1 allowing for very fast multiplications.

Affine coordinates.

- *addition*: $2I + 8M + 2S + 3d$

  $$\begin{aligned}
  &w_1 = x_1 + y_1; \quad w_2 = x_2 + y_2; \quad A = x_1^2 + x_1; \quad B = y_1^2 + y_1;\\
  &C = d_2 w_1 \cdot w_2; \quad D = x_2 \cdot y_2;\\
  &x_3 = y_1 + (C + d_1(w_1 + x_2) + A \cdot (D + x_2))/(d_1 + A \cdot w_2);\\
  &y_3 = x_1 + (C + d_1(w_1 + y_2) + B \cdot (D + y_2))/(d_1 + B \cdot w_2).
  \end{aligned}$$

  Note that according to theorem 3.3.5 the denominators can not be zero when the curve is complete, hence the addition law is complete. Additionally, when $I/M > 3$ it might be favourable to take advantage of the Montgomery inversion trick (see e.g. [21]) and trade $2I$ for $1I + 3M$.

- *doubling*: $1I + 2M + 4S + 2d$

  $$\begin{aligned}
  x_3 &= 1 + \frac{d_1 + d_2(x_1^2 + y_1^2) + y_1^2 + y_1^4}{d_1 + x_1^2 + y_1^2 + (d_2/d_1)(x_1^4 + y_1^4)},\\
  y_3 &= 1 + \frac{d_1 + d_2(x_1^2 + y_1^2) + x_1^2 + x_1^4}{d_1 + x_1^2 + y_1^2 + (d_2/d_1)(x_1^4 + y_1^4)}.
  \end{aligned}$$

  When $d_1 = d_2$, we can save one multiplication:

  $$\begin{aligned}
  &A = x_1^2; \quad B = A^2; \quad C = y_1^2; \quad E = A + C;\\
  &F = 1/(d_1 + E + B + D); \quad x_3 = (d_1 E + A + B) \cdot F; \quad y_3 = x_3 + 1 + d_1 F.
  \end{aligned}$$

  This requires only $1I + 1M + 4S + 2d$.

Projective coordinates.

- *addition*: $18M + 2S + 7d$

  Since in binary fields the cost of field squaring is significantly lower than multiplications[3] and we can choose the parameters d_1 and d_2 to allow for very fast multiplications, the most efficient algorithm is as follows:

  $$A = X_1 \cdot X_2; \quad B = Y_1 \cdot Y_2; \quad C = Z_1 \cdot Z_2; \quad D = d_1 C; \quad E = C^2;$$
  $$F = d_1^2 E; \quad G = (X_1 + Z_1) \cdot (X_2 + Z_2); \quad H = (Y_1 + Z_1) \cdot (Y_2 + Z_2);$$
  $$I = A + G; \quad J = B + H; \quad K = (X_1 + Y_1) \cdot (X_2 + Y_2);$$
  $$U = C \cdot (F + d_1 K \cdot (K + I + J + C));$$
  $$V = U + D \cdot F + K \cdot (d_2(d_1 E + G \cdot H + A \cdot B) + (d_2 + d_1) I \cdot J);$$
  $$X_3 = V + D \cdot (A + D) \cdot (G + D); \quad Y_3 = V + D \cdot (B + D) \cdot (H + D);$$
  $$Z_3 = U.$$

- *doubling*: $2M + 6S + 3d$

  $$A = X_1^2; \quad B = A^2; \quad C = Y_1^2; \quad D = C^2; \quad E = Z_1^2; \quad F = d_1 E^2;$$
  $$G = (d_2/d_1)(B + D); \quad H = A \cdot E; \quad I = C \cdot E; \quad J = H + I; \quad K = G + d_2 J;$$
  $$X_3 = K + H + D; \quad Y_3 = K + I + B; \quad Z_3 = F + J + G.$$

- *readdition*: the aforementioned formulas do not feature computations that could be cached for further use, so adding one point with several other ones can not be sped up.

[3] The M/S ratio can be made up to 35, see [1] or [31].

Conclusion

In previous chapters we have introduced elliptic curves, their arithmetics and presented two different forms. Of course, as has been mentioned, Weierstrass and Edwards forms are not the only ones, but it would be purposeless to name, define and derive all of them. It suffices to refer to [32] for Hessian curves, [18] for Jacobi Quartic curves and [23] for Koblitz curves.

Our attention to the former two has good mathematical and practical background. Firstly, Weierstrass curves present for historical reasons a very good shape helping to understand the topic, especially in case of the group law's graphical expression. Moreover, every elliptic curve can be expressed in Weierstrass form, so proofs done for that form apply to the whole environment. Weierstrass curves were also the first ones on which some speedups were being looked for and where the initial tricks emerged.

However, although other forms offered some improvements in terms of time consumption, they did not overwhelmingly exceeded Weierstrass curves. Moreover, they possessed the same setbacks – namely the necessity to distinguish between various cases and consequently the need to implement a whole *algorithm*, not just formula. That is naturally ideal for side-channel attacks revealing the secret key, a phenomenon of the last fifteen years. The effort of researchers then focused on finding such formulas that would incorporate all cases and hence both simplify the algorithm and prevent side-channel attacks. Unfortunately, it has never been completed – nearest to their goal were strongly unified formulas (without the need to distinguish between general addition and doubling), but they still had some expceptional cases which had to be dealt with.

Together with their speed, this is the reason why Edwards curves have been awarded such place here. Only they possess the ideal property: *completeness*, i.e. one formula works for every pair of input, no matter what its elements are. Of course, there are dedicated algorithms for point doubling faster than general addition, but there is no neccessity in using them. If the implementor does not seek to maximize speed and rather seeks to minimize side-channel information leaks, he can feel free to use one formula in the whole application. Attacks like power, timing or electromagnetic analysis are rendered infeasible – for free. As a bonus, Edwards curves are the most efficient in terms of speed, as the following tables illustrate. To sort the table, column ECDSA-384 (i.e. an average signing operation in ECDSA with key length of 384 bits) was used.

However, for an implementor a serious concern arises when the applications needs to reflect standards. Since ECSDA dates more than ten years back, it could not have contained Edwards curves. Instead, it presents Weierstrass curves (with $a_4 = -3$ in large-p characteristics as this choice is more effective) over both binary and large-prime fields and Koblitz curves as a special case of binary curves optimized for higher speed. Edwards curves are not yet included in any standardization documents, so whereas they might be used to factor large numbers or in private proprietary solutions without much concern, their troublefree use in public sphere is still doubtful. One can only hope that the standardization process catches up with the latest technological progress soon to further favour ECDSA.

algorithm	addition	doubling	ECDSA-384
Edwards projective	$11.2M$	$7.3M$	$4953.6M$
Edwards projective 2	$12.2M$	$7.3M$	$5145.6M$
Weierstrass Jacobian	$14M$	$9.1M$	$6182.4M$
Weierstrass projective, $a_4 = -3$	$14M$	$11M$	$6912M$
Weierstrass projective, $a_4 \neq -3$	$14M$	$12.1M$	$7334.4M$
Weierstrass affine	$23M$	$24M$	$13632M$
Edwards affine + Montgomery trick	$21.9M$	$29.1M$	$15379.2M$
Edwards affine	$46.4M$	$46.1M$	$26611.2M$

Table 3.1: $\mathrm{char}(K) > 3$: operations count assuming $I = 20M$, $S = 1M$ and $a = c = d = 0.1M$.

algorithm	addition	doubling	ECDSA-384
Edwards projective	$11M$	$6.5M$	$4608M$
Edwards projective 2	$11.2M$	$6.5M$	$4646.4M$
Weierstrass Jacobian	$12.8M$	$7.5M$	$5337.6M$
Weierstrass projective, $a_4 = -3$	$13.6M$	$10.4M$	$6604.8M$
Weierstrass projective, $a_4 \neq -3$	$13.6M$	$11.1M$	$6873.6M$
Weierstrass affine	$22.8M$	$23.6M$	$13440M$
Edwards affine + Montgomery trick	$21.9M$	$28.3M$	$15072M$
Edwards affine	$46.4M$	$45.3M$	$26304M$

Table 3.2: $\mathrm{char}(K) > 3$: operations count assuming $I = 20M$, $S = 0.8M$ and $a = c = d = 0.1M$.

algorithm	addition	doubling	ECDSA-384
Edwards projective 2	$10.55M$	$5.98M$	$4321.92M$
Edwards projective	$10.87M$	$5.98M$	$4383.36M$
Weierstrass Jacobian	$12.02M$	$6.46M$	$4788.48M$
Weierstrass projective, $a_4 = -3$	$13.34M$	$10.01M$	$6405.12M$
Weierstrass projective, $a_4 \neq -3$	$13.34M$	$10.45M$	$6574.08M$
Weierstrass affine	$12.67M$	$13.34M$	$7555.2M$
Edwards affine + Montgomery trick	$11.9M$	$17.78M$	$9112.32M$
Edwards affine	$26.4M$	$24.78M$	$14584.32M$

Table 3.3: $\mathrm{char}(K) > 3$: operations count assuming $I = 20M$, $S = 0.67M$ and $a = c = d = 0.1M$.

algorithm	addition	doubling	ECDSA-384
Edwards projective	$18M$	$2M$	$4224M$
Weierstrass Jacobian	$14M$	$4M$	$4224M$
Weierstrass projective	$14M$	$7M$	$5376M$
Weierstrass affine	$12M$	$12M$	$6912M$
Edwards affine + Montgomery + $d_1 = d_2$	$21M$	$11M$	$8256M$
Edwards affine + Montgomery trick	$21M$	$12M$	$8640M$
Edwards affine	$28M$	$12M$	$9984M$

Table 3.4: $\mathrm{char}(K) = 2$: operations count assuming $I = 10M$, $S = 0M$ and $a = c = d = 0M$.

algorithm	addition	doubling	ECDSA-384
Edwards projective	$18.4M$	$3.2M$	$4761.6M$
Weierstrass Jacobian	$15M$	$5M$	$4800M$
Weierstrass projective	$14.2M$	$7.6M$	$5644.8M$
Weierstrass affine	$12.2M$	$12.2M$	$7027.2M$
Edwards affine + Montgomery + $d_1 = d_2$	$21.4M$	$11.8M$	$8640M$
Edwards affine + Montgomery trick	$21.4M$	$12.8M$	$9024M$
Edwards affine	$28.4M$	$12.8M$	$10368M$

Table 3.5: $\mathrm{char}(K) = 2$: operations count assuming $I = 10M$, $S = 0.2M$ and $a = c = d = 0M$.

It is indeed interesting that with such advantages over DSA the spread of ECDSA is not wider. A good example at the national level are biometric passports. The general outlines allow for RSA, DSA and ECDSA, but elliptic curves are the case only in Switzerland and Germany. The vast majority of other countries (USA or EU members) still prefer older (and better known) alternative.

Through the whole thesis, we have been separating the case $\mathbb{F}_p$ and $\mathbb{F}_{2^n}$. A very natural question arises: which underlying field is better? The answer is what one might expect: it depends on the situation. Binary fields allow for carry-less and thus very fast addition (it is indeed just XOR). Also, the number of representations of such a finite field is richer: it can be a *polynomial base*, i.e. vector of polynomial coefficients $(1, \alpha, \alpha^2, \ldots, \alpha^{n-1})$ (then the field is identified with $\mathbb{F}_{2^n}[x]/(f(x))$, where $(f(x))$ is the principal ideal generated by some irreducible polynomial $f(x)$ of degree n and α is its root). Additionaly, one can favour *normal base*, which is a vector $(\alpha, \alpha^2, \alpha^{2^2} \ldots, \alpha^{2^{n-1}})$ for some $\alpha \in \mathbb{F}_{2^n}$. This is preferable in hardware applications, since squaring is nothing but a cyclic shift.

On the other hand, large numbers in $\mathbb{F}_p$ must be dealt with otherwise. In computers, the "human-readable" decimal basis is higly inefficient, so one ends up with binary representation of large numbers anyway. In practise, the base is usually half-size of the processor word (so 2^{16} for 32-bit processors and 2^{32} for 64-bit processors, for instance). Large fields favour sotfware implementations, because the reductions needed are rather multi-line algorithms.

Elliptic curves arithmetics are a similar case – there is no formula being the

most effective in all cases. It therefore depends on the circumstances varying from the underlying field through the device to the purpose for which the whole computation is performed. In the end there must be someone to collect all these pieces of information, weigh them against each other and decide which path to follow. Typically though, when one has a freedom of choice, one picks Edwards curves over both binary and large-prime fields. We must hope that this option reaches public standards soon and does not stay a mathematicians' toy for a long time.

Bibliography

[1] Aranha, Diego S. and López, Julio and Hankerson, Darrel and Rodríguez-Henríquez, Francisco. *Efficient Binary Field Arithmetics Using Vector Instructions Sets.* URL: <http://caramel.loria.fr/sem-slides/201109291030.pdf> [cit. 2012-06-29]

[2] Atkin, Arthur O. L. and Morain, François. *Elliptic Curves and Primality Proving.* Mathematics of Computation 61, Pages 29–68, 1993.

[3] Bernstein, Daniel J. and Farashahi, Reza R. and Lange, Tanja. *Binary Edward Curves.* CHES 2008, LNCS 5154, Pages 244–265, 2008.

[4] Bernstein, Daniel J. and Lange, Tanja. *Curve25519: new Diffie-Hellman speed records.* Public key cryptography—PKC 2006, 9th international conference on theory and practice in public-key cryptography, New York, NY, USA, April 24–26, 2006. Lecture Notes in Computer Science 3958, Springer, 2006. Pages 207–228. ISBN 3-540-33851-9.

[5] Bernstein, Daniel J. and Lange, Tanja. *Explicit-Formulas Database.* URL: <http://hyperelliptic.org/EFD/index.html> [cit. 2012-06-21]

[6] Bernstein, Daniel J. and Lange, Tanja. *Faster addition and doubling on elliptic curves.* Advances in cryptology—ASIACRYPT 2007, 13th international conference on the theory and application of cryptology and information security, Kuching, Malaysia, December 2–6, 2007. Lecture Notes in Computer Science 4833, Springer, 2007. Pages 29–50. ISBN 978-3-540-76899-9.

[7] Blake, Ian and Seroussi, Gadiel and Smart, Nigel. *Elliptic Curves in Cryptography.* Cambridge: Cambridge University Press, 1999. ISBN 0-521-65374-6.

[8] Brown, M. and Hankerson, D. and López, J. and Menezes, A. *Software Implementation of the NIST Elliptic Curves Over Prime Fields.* CT-RSA 2001 Proceedings of the 2001 Conference on Topics in Cryptology: The Cryptographer's Track at RSA. Springer-Verlag London, 2001. Pages 250–265. ISBN 3-540-41898-9.

[9] Chudnovsky, David V. and Chudnovsky, Gregory V. *Sequences of numbers generated by addition in formal groups and new primality and factorization tests.* Advances in Applied Mathematics, Volume 7, Issue 4, December 1986, Pages 385–434. ISSN 0196-8858.

[10] Diffie, Whitfield and Hellman, Martin E. *New Directions in Cryptography.* IEEE Transactions on Information Theory 22, Pages 644–654, 1976.

[11] Drápal, Aleš. *Lecture Notes on Elliptic Curves.*

[12] Edwards, Harold M. *A Normal Form For Elliptic Curves.* Bulletin of the American Mathematical Society, Volume 44, Number 3, July 2007, Pages 393–422.

[13] Edwards, Harold M. *Essays In Constructive Mathematics.* New York: Springer, 2004. ISBN 0-387-219781.

[14] ElGamal, Taher. *A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms.* IEEE Transactions on Information Theory 31, Pages 469–472, 1985.

[15] Grossschädl, Johann and Kamendje, Guy-Armand. *Instruction Set Extension for Fast Elliptic Curve Cryptography over Binary Finite Fields $GF(2m)$* 14th IEEE Proceedings, ASAP 2003, Pages 455–468, 2003.

[16] Hartshorne, Robin. *Algebraic Geometry.* Corrected eighth printing. New York: Springer Verlag, 1997. ISBN 0-387-90244-9.

[17] Hayes, Brian and Ribet, Kenneth E. *Fermat's Last Theorem And Modern Arithmetic.* Americal Scientist 82, Pages 144–156, 1994.

[18] Hisil, Huseyin and Wong, Kenneth Koon-ho and Carter, Gary and Dawson, Ed. *Jacobi Quartic Curves Revisited.* ACISP 2009, LNCS 5594, Pages 452–468, 2009.

[19] Husemöller, Dale. *Elliptic Curves.* Second Edition. New York: Springer Verlag, 2004. ISBN 0-387-95490-2.

[20] Johnson, Don and Menezes, Alfred and Vanstone, Scott. *The Elliptic Curve Digital Signature Algorithm (ECDSA).* Certicom, Inc., 2001.

[21] Kaliski, Burton S. *The Montgomery inverse and its applications.* IEEE Transactions on Computers, 44(8), Pages 1064–1065, 1995.

[22] Kleiner, Israel. *From Fermat to Wiles: Fermat's Last Theorem Becomes a Theorem.* Elemente der Mathematik 55, Pages 19–37, 2000.

[23] Koblitz, Neil. *CM-Curves with Good Cryptographic Properties.* Advances in Cryptology - CRYPTO '91, LNCS 576, Pages 279–287, 1992.

[24] Koblitz, Neil. *Elliptic Curve Cryptosystems.* Mathematics of Computation 48, Pages 203–209, 1987.

[25] Lenstra, Henrik W. *Factoring Integers With Elliptic Curves.* Annals of Mathematics 126, Pages 649–673, 1987.

[26] Miller, Victor S. *Use of Elliptic Curves in Cryptography.* In: Lecture Notes in Computer Sciences 218. Advances in cryptology—CRYPTO 85, Pages 417–426. New York: Springer-Verlag, 1986. ISBN 0-387-16463-4.

[27] National Institute for Standards and Technology. *Digital Signature Standard.* FIPS Publication 186, 1993.

[28] Schmitt, Susanne and Zimmer, Horst G. *Elliptic Curves: A Computational Approach.* Berlin: Walter de Gruyter GmbH, 2003. ISBN 3-11-016808-1.

[29] Silverman, Joseph S. *Advanced Topics in the Arithmetic of Elliptic Curves.* New York: Springer Verlag, 1994. ISBN 0-387-94328-5.

[30] Silverman, Joseph S. *The Arithmetics of Elliptic Curves.* New York: Springer Verlag, 1986. ISBN 0-387-96203-4.

[31] Smart, Nigel P. *A Comparison of Different Finite Fields for Use in Elliptic Curve Cryptosystems.* University of Bristol, June 2000.

[32] Smart, Nigel P. *The Hessian Form of an Elliptic Curve.* CHES 2001, LNCS 2162, Pages 118–125, 2001.

[33] Vanstone, Scott. *Responses to NIST's Proposal.* Communications of the ACM 35, Pages 50–52, 1992.

[34] Wiles, Andrew. *Modular Elliptic Curves And Fermat's Last Theorem.* Annals of Mathematics 142, Pages 443–551, 1994.

[35] Wiles, Andrew. *Ring-theoretic Properties of Certain Hecke Algebras.* Annals of Mathematics 142, Pages 553–572, 1994.