Chapter 13

Arithmetic of Elliptic Curves

Christophe Doche and Tanja Lange

Contents in Brief

Elliptic curves constitute one of the main topics of this book. They have been proposed for applications in cryptography due to their fast group law and because so far no subexponential attack on their discrete logarithm problem (cf. Section 1.5) is known. We deal with security issues in later chapters and concentrate on the *group* arithmetic here. In an actual implementation this needs to be built on an efficient implementation of finite field arithmetic (cf. Chapter 11).

In the sequel we first review the background on elliptic curves to the extent needed here. For a more general presentation of elliptic curves, see Chapter 4. Then we address the question of efficient implementation in large odd and in even characteristics. We refer mainly to [HaMe$^+$ 2003] for these sections.

Note that there are several softwares packages or libraries able to work on elliptic curves, for example PARI/GP [Pari] and **apecs** [Apecs]. The former is a linkable library that also comes with an interactive shell, whereas the latter is a Maple package. Both come with full sources. The computer algebra systems Magma [Magma] and SIMATH [Simath] can deal with elliptic curves, too.

Elliptic curves have received a lot of attention throughout the past almost 20 years and many papers report experiments and timings for various field sizes and coordinates. We do not want to repeat the results but refer to [Ava 2004a, CoMi$^+$ 1998] and Section 14.7 for odd characteristic and [HaLó$^+$ 2000, LóDa 1998, LóDa 1999] for even characteristic. Another excellent and comprehensive reference comparing point multiplication costs and implementation results is [HaMe$^+$ 2003, Tables 3.12, 3.13 and 3.14 and Chap. 5].

13.1 Summary of background on elliptic curves

13.1.1 First properties and group law

We start with a practical definition of the concept of an elliptic curve.

Definition 13.1 An *elliptic curve E over a field K* denoted by E/K is given by the *Weierstraß equation*

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \tag{13.1}$$

where the coefficients $a_1, a_2, a_3, a_4, a_6 \in K$ are such that for each point (x_1, y_1) with coordinates in $\overline{K}$ satisfying (13.1), the partial derivatives $2y_1 + a_1x_1 + a_3$ and $3x_1^2 + 2a_2x_1 + a_4 - a_1y_1$ do not vanish simultaneously.

The last condition says that an elliptic curve is *nonsingular* or *smooth*. A point on a curve is called *singular* if both partial derivatives vanish (cf. the Jacobi criterion 4.94). For shorter reference we group the coefficients in (13.1) to the equation

$$E : y^2 + h(x)y = f(x), \quad h(x), f(x) \in K[x], \quad \deg(h) \leqslant 1, \deg(f) = 3 \text{ with } f \text{ monic.}$$

The smoothness condition can also be expressed more intrinsically. Indeed, let

$$\begin{aligned} &b_2 = a_1^2 + 4a_2, \quad b_4 = a_1a_3 + 2a_4, \\ &b_6 = a_3^2 + 4a_6, \quad b_8 = a_1^2a_6 - a_1a_3a_4 + 4a_2a_6 + a_2a_3^2 - a_4^2. \end{aligned}$$

In odd characteristic, the transformation $y \mapsto y - (a_1x + a_3)/2$ leads to an isomorphic curve given by

$$y^2 = x^3 + \frac{b_2}{4}x^2 + \frac{b_4}{2}x + \frac{b_6}{4} \cdot \tag{13.2}$$

The cubic polynomial above has only simple roots over the algebraic closure $\overline{K}$ if and only if its discriminant is nonzero. The equation of the discriminant is therefore useful to determine if (13.2) is an elliptic curve or not. In addition, it is relevant for characteristic 2 fields as well.

Definition 13.2 Let E be a curve defined over K by (13.1) and let b_2, b_4, b_6 and b_8 as above. The *discriminant of the curve E* denoted by Δ satisfies

$$\Delta = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6.$$

The curve E is nonsingular, and thus is an elliptic curve, if and only if Δ is nonzero. In this case, we introduce the *j-invariant* of E, that is $j(E) = (b_2^2 - 24b_4)^3/\Delta$.

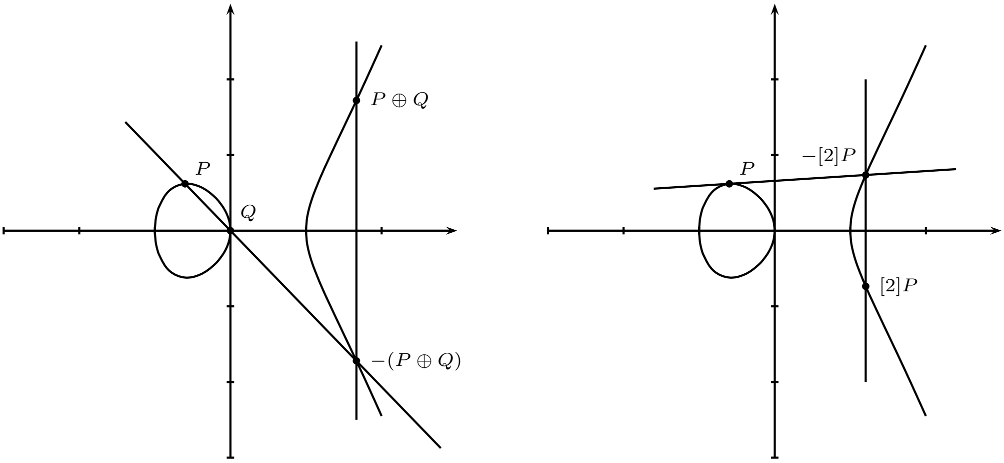
Example 13.3 In $\mathbb{F}_p$ with $p = 2003$, an elliptic curve is given by

$$E_1 : y^2 + 2xy + 8y = x^3 + 5x^2 + 1136x + 531. \tag{13.3}$$

Indeed, we have $b_2 = 24, b_4 = 285, b_6 = 185, \Delta = 1707 \neq 0$ and $j = 171$.

We now show how to turn the set of points of E into a group with group operation denoted by $\oplus$. For this we visualize it over the reals as in Figure 13.1 and assume $h(x) = 0$.

Figure 13.1 Group law on elliptic curve $y^2 = f(x)$ over $\mathbb{R}$.



To add two points $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ in general position one draws a line connecting them. There is a third point of intersection. Mirroring this point at the x-axis gives the sum $P \oplus Q$. The same construction can be applied to double a point where the connecting line is replaced by the tangent at P.

Furthermore, we need to define the sum of two points with the same x-coordinate since for them the group operation cannot be performed as stated. As $y^2 = f(x)$ there are at most 2 such points (x_1, y_1) and $(x_1, -y_1)$. Furthermore, we have to find the neutral element of the group.

The way out is to include a further point P_∞ called the *point at infinity*. It can be visualized as lying far out on the y-axis such that any line $x = c$, for some constant c, parallel to the y-axis passes through it. This point is the neutral element of the group. Hence, the line connecting (x_1, y_1) and $(x_1, -y_1)$ passes through P_∞. As it serves as the neutral element, the inflection process leaves it unchanged such that $(x_1, y_1) \oplus (x_1, -y_1) = P_\infty$, i.e., $(x_1, -y_1) = -P$.

This explanation might sound a little like hand-waving and only applicable to $\mathbb{R}$. We now derive the addition formulas for an arbitrary field K, which hold universally. For a proof we refer to Chapter 4.

Take $P \neq Q$ with $x_1 \neq x_2$ as above and let us compute the coordinates of $R = P \oplus Q = (x_3, y_3)$. The intersecting line has slope

$$\lambda = \frac{y_1 - y_2}{x_1 - x_2}$$

and passes through P. Its equation is thus given by

$$y = \lambda x + \frac{x_1 y_2 - x_2 y_1}{x_1 - x_2}.$$

We denote the constant term by μ and remark $\mu = y_1 - \lambda x_1$. The intersection points with the curve are obtained by equating the line and E

$$(\lambda x + \mu)^2 + (a_1 x + a_3)(\lambda x + \mu) = x^3 + a_2 x^2 + a_4 x + a_6.$$

This leads to the equation $r(x) = 0$ where

$$r(x) = x^3 + (a_2 - \lambda^2 - a_1\lambda)x^2 + (a_4 - 2\lambda\mu - a_3\lambda - a_1\mu)x + a_6 - \mu^2 - a_3\mu.$$

We already know two roots of $r(x)$, namely the x-coordinates of the other two points. Since

$$r(x) = (x - x_1)(x - x_2)(x - x_3)$$

one has $\lambda^2 + a_1\lambda - a_2 = x_1 + x_2 + x_3$. As x_1, x_2 are defined over K so is x_3 and $\tilde{y}_3 = \lambda x_3 + \mu$. The inflection at the x-axis has to be translated to the condition that the second point has the same x-coordinate and also satisfies the curve equation. We observe that if $P = (x_1, y_1)$ is on the curve then so is $(x_1, -y_1 - a_1 x_1 - a_3)$, which corresponds to $-P$ since the point at infinity is the neutral element for this law. Accordingly, we find $y_3 = -\lambda x_3 - \mu - a_1 x_3 - a_3$.

Doubling $P = (x_1, y_1)$ works just the same with the slope obtained by implicit derivating. Thus we have $P \oplus Q = (x_3, y_3)$ and

$$\begin{aligned} -P &= (x_1, -y_1 - a_1 x_1 - a_3), \\ P \oplus Q &= (\lambda^2 + a_1\lambda - a_2 - x_1 - x_2, \lambda(x_1 - x_3) - y_1 - a_1 x_3 - a_3), \text{ where} \\ \lambda &= \begin{cases} \dfrac{y_1 - y_2}{x_1 - x_2} & \text{if } P \neq \pm Q, \\[2ex] \dfrac{3x_1^2 + 2a_2 x_1 + a_4 - a_1 y_1}{2y_1 + a_1 x_1 + a_3} & \text{if } P = Q. \end{cases} \end{aligned}$$

It is immediate from the pictorial description that this law is commutative, has the point at infinity as neutral element, and that the inverse of (x_1, y_1) is given by $(x_1, -y_1 - a_1 x_1 - a_3)$. The associativity can be shown to hold by simply applying the group law and comparing elements. We leave the lengthy computation to the reader. Note that Chapter 4 gives extensive background showing in an abstract way the group of points on E to form a group. For a more geometrical proof, relying on Bezout's theorem, see e.g., [CAS 1991].

Example 13.4 One can easily check that the points $P_1 = (1118, 269)$ and $Q_1 = (892, 529)$ lie on the curve $E_1/\mathbb{F}_p$ as defined by (13.3). Then

$$\begin{aligned} -P_1 &= (1118, 1493), \\ P_1 \oplus Q_1 &= (1681, 1706), \\ [2]P_1 &= (1465, 677) \end{aligned}$$

are also on E_1.

The point at infinity can be motivated by giving an alternative description of elliptic curves. Equation (13.1) expresses the curve in *affine coordinates*. The same elliptic curve E in *projective coordinates* is then given by the equation

$$E : Y^2 Z + a_1 XYZ + a_3 YZ^2 = X^3 + a_2 X^2 Z + a_4 XZ^2 + a_6 Z^3.$$

Let us denote by $(X_1 : Y_1 : Z_1)$ an element of the *projective* 2*-space* $\mathbb{P}^2/K$, i.e., a class of $\overline{K}^3 \setminus \{(0,0,0)\}$ modulo the relation

$$(X_1 : Y_1 : Z_1) \sim (X_2 : Y_2 : Z_2) \iff \text{there is } \lambda \in \overline{K}^* \mid X_2 = \lambda X_1,\ Y_2 = \lambda Y_1 \text{ and } Z_2 = \lambda Z_1.$$

By abuse of notation, we identify a class with any of its representatives and call $(X_1 : Y_1 : Z_1)$ a *projective point*. We remark that only a single point of E satisfies $Z_1 = 0$, namely the *point at infinity*, which is in this case $P_\infty = (0 : 1 : 0)$. When $Z_1 \neq 0$, there is a simple correspondence between the projective point $(X_1 : Y_1 : Z_1)$ and the affine point (x_1, y_1) using the formula

$$(x_1, y_1) = (X_1/Z_1, Y_1/Z_1) \tag{13.4}$$

As the representation $(X_1 : Y_1 : Z_1)$ is not normalized, one can perform arithmetic in projective coordinates without any inversion. Note also that *generalized projective coordinates* involving suitable powers of Z_1 in (13.4) are commonly used, cf. Sections 13.2.1 and 13.3.1.

Example 13.5 The point $P_1' = (917 : 527 : 687)$ lies on the curve E_1 of equation (13.3) expressed in projective coordinates, i.e.,

$$E_1 : Y^2Z + 2XYZ + 8YZ^2 = X^3 + 5X^2Z + 1136XZ^2 + 531Z^3.$$

In fact, P_1' is in the same class as $(1118 : 269 : 1)$ and thus corresponds to the affine point $P_1 = (1118, 269)$.

13.1.2 Scalar multiplication

Take $n \in \mathbb{N} \setminus \{0\}$ and let us denote the *scalar multiplication by n on E* by $[n]$, or $[n]_E$ to avoid confusion. Namely,

$$\begin{aligned}[n] : E &\to E \\ P &\mapsto [n]P = \underbrace{P \oplus P \oplus \cdots \oplus P}_{n \text{ times}}.\end{aligned}$$

This definition extends trivially to all $n \in \mathbb{Z}$, setting $[0]P = P_\infty$ and $[n]P = [-n](-P)$ for $n < 0$. Chapter 9 deals with exponentiation, i.e., the computation of x to some power n. In the context of elliptic curves, this corresponds to $[n]P$. Thus multiplications, squarings, and divisions are replaced by additions, doublings, and subtractions on E.

As an example, we give the analogue of Algorithm 9.10 with additive notation.

Algorithm 13.6 Sliding window scalar multiplication on elliptic curves

INPUT: A point P on an elliptic curve E, a nonnegative integer $n = (n_{l-1} \dots n_0)_2$, a parameter $k \geqslant 1$ and the precomputed points $[3]P$, $[5]P, \dots, [(2^k - 1)]P$.
OUTPUT: The point $[n]P$.

```
Q ← P_∞ and i ← l − 1
while i ⩾ 0 do
    if n_i = 0 then Q ← [2]Q and i ← i − 1
    else
        s ← max(i − k + 1, 0)
        while n_s = 0 do s ← s + 1
        for h = 1 to i − s + 1 do Q ← [2]Q
        u ← (n_i ... n_s)_2                 [n_i = n_s = 1 and i − s + 1 ⩽ k]
        Q ← Q ⊕ [u]P                       [u is odd so that [u]P is precomputed]
```

```
            $i \leftarrow s - 1$
    return $Q$
```

Remark 13.7 Since subtractions can be obtained in a straightforward way, signed-digit representations of n are well suited to compute $[n]P$, cf. Section 9.1.4.

Example 13.8 With the settings of Example 13.4, let us compute $[763]P_1$ with Algorithm 13.6 and a window of size 3. We precompute $[3]P_1 = (1081, 1674)$, $[5]P_1 = (851, 77)$, $[7]P_1 = (663, 1787)$ and since $763 = (\underset{5}{\underline{101}}\,\underset{7}{\underline{111}}\,\underset{5}{\underline{101}}\,\underset{1}{\underline{1}})_2$, the intermediate values of Q are

$$\begin{array}{rrr}
[5]P_1 = (851, 77), & [10]P_1 = (4, 640), & [20]P_1 = (836, 807), \\
[40]P_1 = (1378, 1696), & [47]P_1 = (1534, 747), & [94]P_1 = (1998, 1094), \\
[188]P_1 = (1602, 1812), & [376]P_1 = (478, 1356), & [381]P_1 = (1454, 981), \\
[762]P_1 = (1970, 823), & [763]P_1 = (1453, 1428). &
\end{array}$$

Using the NAF expansion of $763 = (\underset{3}{\underline{10\bar{1}}}00000\underset{-5}{\underline{\bar{1}0\bar{1}}})_{\mathrm{s}}$ instead, one obtains

$$\begin{array}{rrr}
[3]P_1 = (1081, 1674), & [6]P_1 = (255, 1499), & [12]P_1 = (459, 1270), \\
[24]P_1 = (41, 1867), & [48]P_1 = (1461, 904), & [96]P_1 = (1966, 1808), \\
[192]P_1 = (892, 529), & [384]P_1 = (1928, 1803), & [768]P_1 = (799, 1182), \\
[763]P_1 = (1453, 1428). & &
\end{array}$$

The last step, namely $[763]P_1 = [768]P_1 \oplus [-5]P_1$, needs $[-5]P_1 = (851, 216)$ which is obtained directly from $[5]P_1$.

13.1.3 Rational points

When we consider a point P on an elliptic curve E/K, it is implicit that P has its coordinates in $\overline{K}$. To stress that P has its coordinates in K, we introduce a new concept.

Definition 13.9 Let E be an elliptic curve defined over K. The points lying on E with coordinates in K form the set of *K-rational points of E* denoted by $E(K)$. We have

$$E(K) = \{(x_1, y_1) \in K^2 \mid y_1^2 + a_1x_1y_1 + a_3y_1 = x_1^3 + a_2x_1^2 + a_4x_1 + a_6\} \cup \{P_\infty\}.$$

The structure of the group of $\mathbb{F}_q$-rational points is easy to describe. Indeed, by Corollary 5.77, $E(\mathbb{F}_q)$ is either cyclic or isomorphic to a product of two cyclic groups, namely $E(\mathbb{F}_q) \simeq \mathbb{Z}/d_1\mathbb{Z} \times \mathbb{Z}/d_2\mathbb{Z}$ where $d_1 \mid d_2$ and $d_1 \mid q - 1$.

For cryptographic applications one usually works in a subgroup of prime order ℓ. Hence, one is interested in curves and finite fields such that $|E(\mathbb{F}_q)| = c\ell$ for some small *cofactor* c. See [GaMc 2000] for conjectural probabilities that the number of points is a prime or has a small cofactor.

Finding a random $\mathbb{F}_q$-rational point P on an elliptic curve $E/\mathbb{F}_q$ is quite easy. See Sections 13.2 and 13.3 for examples. If the curve has a cofactor $c > 1$ then this random point needs not lie inside the group of order ℓ. However, the point $Q = [c]P$ either equals P_∞, in which case one has to try with a different random point P, or is a point in the prime order subgroup.

Example 13.10 Let us consider the curve E_1 as defined by (13.3). One can check that $|E_1(\mathbb{F}_p)| = 1956 = 12 \times 163$. So, there are 1955 affine points with coordinates in $\mathbb{F}_p$ and the point at infinity P_∞ lying on E_1. The point $P_1 = (1118, 269)$ is of order 1956 which implies that the group $E_1(\mathbb{F}_p)$ is cyclic generated by P_1. The point $Q_1 = (892, 529)$ is of prime order 163.

13.1.4 Torsion points

Definition 13.11 Let E/K be an elliptic curve and $n \in \mathbb{Z}$. The *kernel of* $[n]$, denoted by $E[n]$, satisfies

$$E[n] = \{P \in E(\overline{K}) \mid [n]P = P_\infty\}.$$

An element $P \in E[n]$ is called a *n-torsion point*.

Example 13.12 As $E_1(\mathbb{F}_p)$ is cyclic of order $1956 = 2^2 \times 3 \times 163$, there are n-torsion points in $E_1(\mathbb{F}_p)$ for every n dividing 1956. For instance, $R_1 = (1700, 299)$ on E_1 satisfies $R_1 = -R_1$. Thus R_1 is a 2-torsion rational point. If n is not a divisor of 1956, the corresponding n-torsion points have coordinates in some extension of $\mathbb{F}_p$. For example, there is a 9-torsion point with coordinates in the field $\mathbb{F}_{p^3} \simeq \mathbb{F}_p[\theta]$ with θ such that $\theta^3 + \theta^2 + 2 = 0$. Indeed, we can check that

$$\begin{aligned} S_1 &= (1239\theta^2 + 1872\theta + 112, 1263\theta^2 + 334\theta + 1752) \in E_1(\mathbb{F}_{p^3}), \\ [3]S_1 &= (520, 1568) \in E_1(\mathbb{F}_p), \\ [8]S_1 &= (1239\theta^2 + 1872\theta + 112, 265\theta^2 + 1931\theta + 19) = -S_1 \end{aligned}$$

so that S_1 is a 9-torsion point.

See also the related notion of division polynomial in Section 4.4.2.a.

Theorem 13.13 Let E be an elliptic curve defined over K. If the characteristic of K is either zero or prime to n then

$$E[n] \simeq \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}.$$

Otherwise, when $\mathrm{char}(K) = p$ and $n = p^r$, then either

$$E[p^r] = \{P_\infty\}, \text{ for all } r \geqslant 1 \quad \text{or} \quad E[p^r] \simeq \mathbb{Z}/p^r\mathbb{Z}, \text{ for all } r \geqslant 1.$$

Definition 13.14 Let $\mathrm{char}(K) = p$ and let E be defined over K. If $E[p^r] = \{P_\infty\}$ for one and in fact for all positive integers r, then the curve is called *supersingular*. Otherwise the curve is called *ordinary*.

A curve defined over a prime field $\mathbb{F}_p, p > 3$ is supersingular if and only if $|E(\mathbb{F}_p)| = p + 1$, cf. Proposition 13.31. Note also that if $\mathrm{char}(\mathbb{F}_q) = 2$ or 3, E is supersingular if and only if its j-invariant is zero.

Example 13.15 The curve $E_1/\mathbb{F}_p$ is ordinary. This implies that $E_1[p]$ is a subgroup of $\big(E_1(\overline{\mathbb{F}}_p), \oplus\big)$ isomorphic to $(\mathbb{F}_p, +)$.

13.1.5 Isomorphisms

Some changes of variables do not fundamentally alter an elliptic curve. Let us first describe the transformations that keep the curve in Weierstraß form.

13.1.5.a Admissible change of variables and twists

Let E/K be an elliptic curve

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6.$$

The maps

$$x \mapsto u^2x' + r \quad \text{and} \quad y \mapsto u^3y' + u^2sx' + t$$

with $(u, r, s, t) \in K^* \times K^3$ are invertible and transform the curve E into

$$E' : y'^2 + a'_1 x' y' + a'_3 y' = x'^3 + a'_2 x'^2 + a'_4 x' + a'_6,$$

where the a'_i's belong to K and can be expressed in terms of the a_i's and u, r, s, t. Via the inverse map, we associate to each point of E a point of E' showing that both curves are *isomorphic over* K. These changes of variables are the only ones leaving invariant the shape of the defining equation and, hence, they are the only *admissible change of variables*.

In case (u, r, s, t) belongs to $\overline{K}^* \times \overline{K}^3$ whereas the curves E and E', as above, are still defined over K, then E and E' are *isomorphic over* $\overline{K}$ or *twists* of each other.

Corollary 13.16 Assume that the characteristic of K is prime to 6 and let E be given by a short Weierstraß equation

$$E : y^2 = x^3 + a_4 x + a_6.$$

- If $a_4 = 0$ then for every $a'_6 \in K^*$ the curve E is isomorphic to $E' : y^2 = x^3 + a'_6$ over $K\big((a_6/a'_6)^{1/6}\big)$.
- If $a_6 = 0$ then for every $a'_4 \in K^*$ the curve E is isomorphic to $E' : y^2 = x^3 + a'_4 x$ over $K\big((a_4/a'_4)^{1/4}\big)$.
- If $a_4 a_6 \neq 0$ then for every $v \in K^*$ the curve E is isomorphic to $\widetilde{E}_v : y^2 = x^3 + a'_4 x + a'_6$ with $a'_4 = v^2 a_4$ and $a'_6 = v^3 a_6$ over the field $K(\sqrt{v})$.

The curves $\widetilde{E}_v$ are called *quadratic twists* of E. Note that E is isomorphic to $\widetilde{E}_v$ over K if and only if v is a square in K^*. Therefore up to isomorphisms there is only one quadratic twist of a curve with $a_4 a_6 \neq 0$.

Remark 13.17 Likewise one can define the quadratic twist of E by a quadratic nonresidue v as $\widetilde{E}_v : vy^2 = x^3 + a_4 x + a_6$, which is isomorphic to the above definition, as can be seen by dividing by v^3 and transforming $y \mapsto y/v, x \mapsto x/v$.

From this form one sees that E and $\widetilde{E}_v$ together contain exactly two points (x, y_i) for each field element $x \in \mathbb{F}_q$.

Proposition 13.18 Let E/K and E'/K be two elliptic curves. If E and E' are isomorphic over K then they have the same j-invariant. Conversely, if $j(E) = j(E')$ then E and E' are isomorphic over $\overline{K}$.

Using an adequate isomorphism over K, it is always possible to find a *short Weierstraß equation* that actually depends on the characteristic of the field and on the value of the j-invariant. All the cases and equations are summarized in Table 13.2.

Table 13.2 Short Weierstraß equations.

char K	Equation	Δ	j
$\neq 2, 3$	$y^2 = x^3 + a_4 x + a_6$	$-16(4a_4^3 + 27a_6^2)$	$1728 a_4^3/4\Delta$
3	$y^2 = x^3 + a_4 x + a_6$	$-a_4^3$	0
3	$y^2 = x^3 + a_2 x^2 + a_6$	$-a_2^3 a_6$	$-a_2^3/a_6$
2	$y^2 + a_3 y = x^3 + a_4 x + a_6$	a_3^4	0
2	$y^2 + xy = x^3 + a_2 x^2 + a_6$	a_6	$1/a_6$

Example 13.19 The change of variables $(x, y) \mapsto (x-2, y-x-2)$ transforms the curve E_1 given by (13.3) into

$$E_2 : y^2 = x^3 + 1132x + 278.$$

The point $P_1 = (1118, 269)$ is mapped to $P_2 = (1120, 1391) \in E_2(\mathbb{F}_p)$.

Let v be a quadratic nonresidue modulo $p = 2003$ and let $u \in \mathbb{F}_{p^2}$ be a square root of v. Then the change of variables $(x, y) \mapsto (x/u^2, y/u^3)$ is an $\mathbb{F}_{p^2}$-isomorphism between

$$E_2 : y^2 = x^3 + 1132x + 278.$$

and its *quadratic twist by* v, namely

$$\widetilde{E}_{2,v} : y^2 = x^3 + 1132v^2x + 278v^3.$$

We have $\Delta(\widetilde{E}_{2,v}) = v^6\Delta(E_2)$ and $j(\widetilde{E}_{2,v}) = j(E_2) = 171$.

The curves E_2 and $\widetilde{E}_{2,v}$ are defined over $\mathbb{F}_p$ whereas the isomorphism has coefficients in $\mathbb{F}_{p^2}$.

Remark 13.20 There are many other ways to represent an elliptic curve. For instance, we can cite the *Legendre form*

$$y^2 = x(x-1)(x-\lambda)$$

or the *Jacobi model*

$$y^2 = x^4 + ax^2 + b.$$

Over a field of characteristic greater than 3, it is also possible to represent an elliptic curve as the intersection of two quadrics with a rational point [Cas 1991]. The resulting *Jacobi form* is used in [LiSm 2001] to prevent SPA/DPA attacks, cf. Section 29.1.2.c. Quite recently, some attention has been given to another representation, namely the Hessian form, which presents some advantages from an algorithmic and cryptographic point of view [Sma 2001, Fri 2001, JoQu 2001].

13.1.5.b Hessian form

Let $\mathbb{F}_q$ be a finite field where q is a prime power such that $q \equiv 2 \pmod 3$ and consider an elliptic curve E over $\mathbb{F}_q$ with a $\mathbb{F}_q$-rational point of order 3. These assumptions are not fundamentally necessary but they make the construction of the Hessian form easier and let the equation be defined over $\mathbb{F}_q$. In particular, one can assume that E is given by the equation

$$E : y^2 + a_1xy + a_3y = x^3,$$

moving a point of order 3 to the origin, if necessary.

Let $\delta = (a_1^3 - 27a_3)$ so that $\Delta = a_3^3\delta$. Now if $q \equiv 2 \pmod 3$ every element $\alpha \in \mathbb{F}_q$ is a cube. Thus every α has a unique cube root, denoted by $\alpha^{1/3}$, which is equal to plus or minus the square root of $\alpha^{(q+1)/3}$. This implies that

$$\mu = \frac{1}{3}\big((-27a_3\delta^2 - \delta^3)^{1/3} + \delta\big) \in \mathbb{F}_q.$$

With these settings, to every point (x_1, y_1) on E corresponds $(X_1 : Y_1 : Z_1)$ with

$$X_1 = \frac{a_1(2\mu - \delta)}{3\mu - \delta}x_1 + y_1 + a_3, \qquad Y_1 = \frac{-a_1\mu}{3\mu - \delta}x_1 - y_1, \qquad Z_1 = \frac{-a_1\mu}{3\mu - \delta}x_1 - a_3$$

on the cubic

$$H : X^3 + Y^3 + Z^3 = cXYZ \quad \text{where} \quad c = 3\frac{\mu - \delta}{\mu}\cdot$$

Definition 13.21 The equation H is called the *Hessian form* of E.

One of the main features of elliptic curves expressed in Hessian form is the simplicity of the group law, which is independent of the parameter c.

Namely, take $P = (X_1 : Y_1 : Z_1)$ and $Q = (X_2 : Y_2 : Z_2)$ on H such that $P \neq Q$, then the point with coordinates $(X_3 : Y_3 : Z_3)$ such that

$$X_3 = Y_1^2 X_2 Z_2 - Y_2^2 X_1 Z_1, \qquad Y_3 = X_1^2 Y_2 Z_2 - X_2^2 Y_1 Z_1, \qquad Z_3 = Z_1^2 X_2 Y_2 - Z_2^2 X_1 Y_1$$

is on H and corresponds to $P \oplus Q$.

One can check that the neutral element for that law is $(1 : -1 : 0)$ and that the opposite of P_1 is $-P_1 = (Y_1 : X_1 : Z_1)$.

The coordinates of $[2]P$ are

$$X_3 = Y_1(Z_1^3 - X_1^3), \qquad Y_3 = X_1(Y_1^3 - Z_1^3), \qquad Z_3 = Z_1(X_1^3 - Y_1^3).$$

An addition requires 12 field multiplication and 6 squarings, whereas a doubling needs 6 multiplications and 3 squarings and both operations can be implemented in a highly parallel way [SMA 2001]. It is also interesting to note that $[2]P$ is equal to $(Z_1 : X_1 : Y_1) \oplus (Y_1 : Z_1 : X_1)$. As a consequence the same formulas can be used to double, add, and subtract points, which makes Hessian curves interesting against side-channel attacks [JoQu 2001] (cf. Section 29.1.2.b).

To find the Hessian form of an elliptic curve $E/\mathbb{F}_q$ in the general case [FRI 2001], we remark that the j-invariant of H is equal to

$$j = \frac{c^3(c^3+216)^3}{c^9 - 81c^6 + 2187c^3 - 19683}.$$

So the Hessian form of E is defined over $\mathbb{F}_q$ if and only if there exists $c \in \mathbb{F}_q$ such that

$$c^3(c^3+216)^3 - j(c^9 - 81c^6 + 2187c^3 - 19683) = 0$$

where j is the j-invariant of E.

Example 13.22 Take

$$E_2 : y^2 = x^3 + 1132x + 278$$

defined over $\mathbb{F}_p$ with $p = 2003$. Moving the point $(522, 1914) \in E_2(\mathbb{F}_p)$ of order 3 to the origin by the transformation

$$(x, y) \mapsto (x + 522, y + 555x + 1914)$$

gives the curve

$$E_3 : y^2 + 1110xy + 1825y = x^3.$$

So, from above, $\delta = 1427$ and $\mu = 1322$ so that E_3, consequently E_2 and E_1, are all isomorphic to

$$H : X^3 + Y^3 + Z^3 = 274XYZ.$$

The point $(1118, 269)$ on E_1 is sent to $(1120, 1391)$ on E_2, from where it is in turn mapped to $(598, 85)$ on E_3, which is finally sent to $(1451 : 672 : 935)$ on H.

Note that all these transformations respect the group laws of the different curves. Indeed, a K-isomorphism between the curves E and E' always gives rise to a group homorphism between $E(K)$ and $E'(K)$. However, these notions are different. That is why we introduce a new concept in the following.

13.1.6 Isogenies

Definition 13.23 Two curves E/K and E'/K are *isogenous over* K if there exists a morphism $\psi : E \to E'$ with coefficients in K mapping the neutral element of E to the neutral element of E'. From this simple property, it is possible to show that ψ is a group homomorphism from $E(K)$ to $E'(K)$.

One important property is that for every isogeny ψ, there exists a unique isogeny $\hat{\psi} : E' \to E$ called the *dual isogeny* such that

$$\hat{\psi} \circ \psi = [m]_E \quad \text{and} \quad \psi \circ \hat{\psi} = [m]_{E'}.$$

The *degree of the isogeny* ψ is equal to this m. For more background on isogenies, we refer to Section 4.3.4

Proposition 13.24 Two elliptic curves E and E' defined over $\mathbb{F}_q$ are isogenous over $\mathbb{F}_q$ if and only $|E(\mathbb{F}_q)| = |E'(\mathbb{F}_q)|$.

Example 13.25 Take

$$E_2 : y^2 = x^3 + 1132x + 278$$

and

$$E_4 : y^2 = x^3 + 500x + 1005.$$

These two curves have the same cardinality, $|E_2(\mathbb{F}_p)| = |E_4(\mathbb{F}_p)| = 1956$. Then E_2 and E_4 must be isogenous over $\mathbb{F}_{2003}$. The isogeny of degree 2 is given by the formula [Ler 1997]

$$\psi : (x, y) \longmapsto \left(\frac{x^2 + 301x + 527}{x + 301}, \frac{yx^2 + 602yx + 1942y}{x^2 + 602x + 466} \right).$$

For instance, the points, $P_2 = (1120, 1391)$ and $Q_2 = (894, 1425)$ in $E_2(\mathbb{F}_p)$ are respectively mapped by ψ on $P_4 = (565, 302)$ and $Q_4 = (1818, 1002)$ which lie on E_4. Now

$$\begin{aligned} P_2 \oplus Q_2 &= (1683, 1388), \\ P_4 \oplus Q_4 &= (1339, 821), \\ \psi(P_2 \oplus Q_2) &= (1339, 821), \\ &= \psi(P_2) \oplus \psi(Q_2). \end{aligned}$$

Note that E_2 and E_4 are isogenous but not isomorphic since $j(E_2) = 171$ whereas $j(E_4) = 515$. Furthermore, the group structure is different as $E_2(\mathbb{F}_p)$ is cyclic while $E_4(\mathbb{F}_p)$ is the direct product of a group of order 2 generated by $(1829, 0)$ and a group of order 978 generated by $(915, 1071)$.

13.1.7 Endomorphisms

The multiplication by n is an endomorphism of the curve E for every $n \in \mathbb{Z}$. The set of all endomorphisms of E defined over K will be denoted by $\mathrm{End}_K(E)$ or more simply by $\mathrm{End}(E)$, and thus contains at least $\mathbb{Z}$.

Definition 13.26 If $\mathrm{End}(E)$ is strictly bigger than $\mathbb{Z}$ we say that E has *complex multiplication*.

Let E be a nonsupersingular elliptic curve over $\mathbb{F}_q$. Such an E always has complex multiplication. Indeed, the Frobenius automorphism of $\mathbb{F}_q$ extends to the points of the curve by sending P_∞ to itself and $P = (x_1, y_1)$ to $\phi_q(P) = (x_1^q, y_1^q)$. One can easily check that the point $\phi_q(P)$ is again a point on the curve irrespective of the field of definition of P. Hence, ϕ_q is an endomorphism of E, called the *Frobenius endomorphism* of $E/\mathbb{F}_q$. It is different from $[n]$ for all $n \in \mathbb{Z}$.

Example 13.27 Take $P_1 = (1120, 391)$ on $E_1/\mathbb{F}_p$. Since P_1 has coordinates in $\mathbb{F}_p$, $\phi_p(P_1)$ is simply equal to P_1. At present, let us consider a point on E_1 with coordinates in an extension of $\mathbb{F}_p$. For instance, in Example 13.10, we give the point S_1 of order 9 in $E_1(\mathbb{F}_{p^3})$. We have

$$\begin{aligned} S_1 &= (1239\theta^2 + 1872\theta + 112, 1263\theta^2 + 334\theta + 1752), \\ \phi_p(S_1) &= (217\theta^2 + 399\theta + 1297, 681\theta^2 + 811\theta + 102), \\ \phi_p^2(S_1) &= (547\theta^2 + 1735\theta + 297, 59\theta^2 + 858\theta + 325), \\ \phi_p^3(S_1) &= (1239\theta^2 + 1872\theta + 112, 1263\theta^2 + 334\theta + 1752) = S_1. \end{aligned}$$

All of them are also 9-torsion points.

13.1.8 Cardinality

The cardinality of an elliptic curve E over $\mathbb{F}_q$, i.e., the number of $\mathbb{F}_q$-rational points, is an important aspect for the security of cryptosystems built on $E(\mathbb{F}_q)$, cf. Section 19.3.

The theorem of Hasse–Weil relates the number of points to the field size.

Theorem 13.28 (Hasse–Weil) Let E be an elliptic curve defined over $\mathbb{F}_q$. Then

$$|E(\mathbb{F}_q)| = q + 1 - t \text{ and } |t| \leqslant 2\sqrt{q}.$$

Remarks 13.29

(i) The integer t is called the *trace of the Frobenius endomorphism.*

(ii) For any integer $t \in [-2\sqrt{p}, 2\sqrt{p}]$ there is at least one elliptic curve E defined over $\mathbb{F}_p$ whose cardinality is $p + 1 - t$.

Concerning admissible cardinalities, the more general result is proved in [Wat 1969].

Theorem 13.30 Let $q = p^d$. There exists an elliptic curve E defined over $\mathbb{F}_q$ with $|E(\mathbb{F}_q)| = q + 1 - t$ if and only if one of the following conditions holds:

1. $t \not\equiv 0 \pmod p$ and $t^2 \leqslant 4q$.
2. d is odd and either (i) $t = 0$ or (ii) $p = 2$ and $t^2 = 2q$ or (iii) $p = 3$ and $t^2 = 3q$.
3. d is even and either (i) $t^2 = 4q$ or (ii) $p \not\equiv 1 \pmod 3$ and $t^2 = q$ or (iii) $p \not\equiv 1 \pmod 4$ and $t = 0$.

One associates to ϕ_q the polynomial

$$\chi_E(T) = T^2 - tT + q.$$

It is called the *characteristic polynomial of the Frobenius endomorphism*, since

$$\chi_E(\phi_q) = \phi_q^2 - [t]\phi_q + [q] = [0].$$

So, for each $P \in E(\overline{\mathbb{F}}_q)$, we have

$$\phi_q^2(P) \oplus [-t]\phi_q(P) \oplus [q]P = P_\infty.$$

As points in $E(\mathbb{F}_q)$ are fixed under ϕ_q they form the kernel of $(\mathrm{Id} - \phi_q)$ and $|E(\mathbb{F}_q)| = \chi_E(1)$.

From the complex roots τ and $\overline{\tau}$ of $\chi_E(\phi_q)$ one can compute the group order of $E(\mathbb{F}_{q^k})$, that is

$$|E(\mathbb{F}_{q^k})| = q^k + 1 - \tau^k - \overline{\tau}^k, \quad \text{for all } k \geqslant 1. \tag{13.5}$$

More explicitly, one has

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

where the sequence $(t_k)_{k\in\mathbb{N}}$ satisfies $t_0 = 2$, $t_1 = t$ and $t_{k+1} = tt_k - qt_{k-1}$, for $k \geqslant 1$.

We also have the following properties.

Proposition 13.31 Let E be a curve defined over a field $\mathbb{F}_q$ of characteristic p. The curve E is supersingular if and only if the trace t of the Frobenius satisfies

$$t \equiv 0 \pmod{p}.$$

Proposition 13.32 Let E be a curve defined over $\mathbb{F}_q$ and let $\widetilde{E}$ be the quadratic twist of E. Then

$$|E(\mathbb{F}_q)| + |\widetilde{E}(\mathbb{F}_q)| = 2q + 2.$$

This can be easily seen to hold from Remark 13.17. One immediately gets $\chi_{\widetilde{E}}(T) = T^2 + tT + q$. When one tries to find a curve with a suitable cryptographic order, that is, an order with a large prime factor, Proposition 13.32 is especially useful since it gives two candidates for each computation, cf. Chapter 17.

Example 13.33 The cardinality of $E_2(\mathbb{F}_p)$ is 1956. Therefore, ϕ_p satisfies

$$\chi_{E_2}(T) = T^2 - 48T + 2003.$$

Let $R_2 = (443\theta^2 + 1727\theta + 1809, 929\theta^2 + 280\theta + 946)$. Then

$$\begin{aligned} \phi_p(R_2) &= (857\theta^2 + 1015\theta + 766, 126\theta^2 + 1902\theta + 419), \\ \phi_p^2(R_2) &= (703\theta^2 + 1264\theta + 1568, 948\theta^2 + 1824\theta + 119) \end{aligned}$$

and one can check that

$$\phi_p^2(R_2) - [48]\phi_{2003}(R_2) + [2003]R_2 = P_\infty.$$

Also, we deduce that $|E_2(\mathbb{F}_{p^2})| = 4013712$ and $|E_2(\mathbb{F}_{p^3})| = 8036231868$.

Finally the cardinality of the curve

$$\widetilde{E}_2 : y^2 = x^3 + 774x + 1867$$

which is the twist of E_2 by the quadratic nonresidue 78, satisfies $|\widetilde{E}_2| = 2052$, and the characteristic equation of the Frobenius of $\widetilde{E}_2/\mathbb{F}_p$ is

$$\chi_{\widetilde{E}_2}(T) = T^2 + 48T + 2003.$$

13.2 Arithmetic of elliptic curves defined over $\mathbb{F}_p$

In this section we consider curves defined over finite prime fields. As they should be used in cryptographic applications, we can assume p to be large, hence, at least $p > 3$. We remark that all considerations in this section hold true for an elliptic curve defined over an arbitrary finite field $\mathbb{F}_q$ if $\operatorname{char}(\mathbb{F}_q) > 3$ and for supersingular curves over field of characteristic 3.

We already know that an elliptic curve E can be represented with respect to several coordinate systems, e.g., affine or projective coordinates. In the following we deal with efficient addition and doubling in the group of points E. To this aim we introduce five different coordinate systems in which the speeds of addition and doubling differ. We measure the time by the number of field operations needed to perform the respective operation.

In characteristic $p > 3$, one can always take for E, cf. Table 13.2, an equation of the form

$$E : y^2 = x^3 + a_4x + a_6,$$

where a_4 and a_6 are in $\mathbb{F}_p$. The points lying on the curve can have coordinates in $\mathbb{F}_p$ or in some extension $\mathbb{F}_q/\mathbb{F}_p$, for instance in an optimal extension field, cf. Section 11.3. This has two advantages. First, it is straightforward to obtain the cardinality of $E(\mathbb{F}_q)$ using (13.5) and one can use the Frobenius ϕ_p to speed up computations, cf. Section 15.1.

In the remainder of this section we deal with addition and doubling in different coordinate systems, give strategies for choosing optimal coordinates for scalar multiplication and introduce Montgomery coordinates and their arithmetic. Finally, we show how to compress the representation of a point.

An elementary multiplication in $\mathbb{F}_q$ (resp. a squaring and an inversion) will be abbreviated by M (resp. S and I).

13.2.1 Choice of the coordinates

This section is based on [CoMi+ 1998].

In Section 13.1.1 we explained the group law in general. Here we shall give formulas for the coordinates of the result of the

- addition of two points P and $Q \in E(\mathbb{F}_p)$ provided $P \neq \pm Q$,
- doubling of P.

13.2.1.a Affine coordinates ($\mathcal{A}$)

We can assume that E is given by

$$y^2 = x^3 + a_4x + a_6.$$

By the arguments above, we know that the opposite of the point (x_1, y_1) lying on E is $(x_1, -y_1)$. Also we have:

Addition

Let $P = (x_1, y_1)$, $Q = (x_2, y_2)$ such that $P \neq \pm Q$ and $P \oplus Q = (x_3, y_3)$. In this case, addition is given by

$$x_3 = \lambda^2 - x_1 - x_2, \qquad y_3 = \lambda(x_1 - x_3) - y_1, \qquad \lambda = \frac{y_1 - y_2}{x_1 - x_2}.$$

Doubling

Let $[2]P = (x_3, y_3)$. Then

$$x_3 = \lambda^2 - 2x_1, \qquad y_3 = \lambda(x_1 - x_3) - y_1, \qquad \lambda = \frac{3x_1^2 + a_4}{2y_1}.$$

For these formulas one can easily read off that an addition and a doubling require I + 2M + S and I + 2M + 2S, respectively.

Doubling followed by an addition

Building on the ideas in [EILA+ 2003], the authors of [CIJO+ 2003] show how to speed up the computation of a doubling followed by an addition using $[2]P \oplus Q$ as $(P \oplus Q) \oplus P$. The basic idea, i.e., omitting the computation of the intermediate values y_3 and x_3, saves one multiplication and the new formulas are more efficient whenever a field inversion is more expensive than 6 multiplications. The formulas are as follows where we assume that $P \neq \pm Q$ and $[2]P \neq -Q$

$$\begin{array}{lll} A = (x_2 - x_1)^2, & B = (y_2 - y_1)^2 & C = A(2x_1 + x_2) - B, \\ D = C(x_2 - x_1), & E = D^{-1}, & \lambda = CE(y_2 - y_1), \\ \lambda_2 = 2y_1 A(x_2 - x_1)E - \lambda, & x_4 = (\lambda_2 - \lambda)(\lambda + \lambda_2) + x_2, & y_4 = (x_1 - x_4)\lambda_2 - y_1, \end{array}$$

needing I + 9M + 2S.

13.2.1.b Projective coordinates ($\mathcal{P}$)

In *projective coordinates*, the equation of E is

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

The point $(X_1 : Y_1 : Z_1)$ on E corresponds to the affine point $(X_1/Z_1, Y_1/Z_1)$ when $Z_1 \neq 0$ and to the point at infinity $P_\infty = (0 : 1 : 0)$ otherwise. The opposite of $(X_1 : Y_1 : Z_1)$ is $(X_1 : -Y_1 : Z_1)$.

Addition

Let $P = (X_1 : Y_1 : Z_1)$, $Q = (X_2 : Y_2 : Z_2)$ such that $P \neq \pm Q$ and $P \oplus Q = (X_3 : Y_3 : Z_3)$. Then set

$$A = Y_2 Z_1 - Y_1 Z_2, \qquad B = X_2 Z_1 - X_1 Z_2, \qquad C = A^2 Z_1 Z_2 - B^3 - 2B^2 X_1 Z_2$$

so that

$$X_3 = BC, \qquad Y_3 = A(B^2 X_1 Z_2 - C) - B^3 Y_1 Z_2, \qquad Z_3 = B^3 Z_1 Z_2.$$

Doubling

Let $[2]P = (X_3 : Y_3 : Z_3)$ then put

$$A = a_4 Z_1^2 + 3X_1^2, \qquad B = Y_1 Z_1, \qquad C = X_1 Y_1 B, \qquad D = A^2 - 8C$$

and

$$X_3 = 2BD, \qquad Y_3 = A(4C - D) - 8Y_1^2 B^2, \qquad Z_3 = 8B^3.$$

No inversion is needed, and the computation times are 12M + 2S for a general addition and 7M + 5S for a doubling. If one of the input points to the addition is given by $(X_2 : Y_2 : 1)$, i.e., directly transformed from affine coordinates, then the requirements for an addition decrease to 9M + 2S.

13.2.1.c Jacobian and Chudnovsky Jacobian coordinates ($\mathcal{J}$ and $\mathcal{J}^c$)

With *Jacobian coordinates* the curve E is given by

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

The point $(X_1 : Y_1 : Z_1)$ on E corresponds to the affine point $(X_1/Z_1^2, Y_1/Z_1^3)$ when $Z_1 \neq 0$ and to the point at infinity $P_\infty = (1 : 1 : 0)$ otherwise. The opposite of $(X_1 : Y_1 : Z_1)$ is $(X_1 : -Y_1 : Z_1)$.

Addition

Let $P = (X_1 : Y_1 : Z_1)$, $Q = (X_2 : Y_2 : Z_2)$ such that $P \neq \pm Q$ and $P \oplus Q = (X_3 : Y_3 : Z_3)$. Then set

$$A = X_1 Z_2^2, \quad B = X_2 Z_1^2, \quad C = Y_1 Z_2^3, \quad D = Y_2 Z_1^3, \quad E = B - A, \quad F = D - C$$

and

$$X_3 = -E^3 - 2AE^2 + F^2, \qquad Y_3 = -CE^3 + F(AE^2 - X_3), \qquad Z_3 = Z_1 Z_2 E.$$

Doubling

Let $[2]P = (X_3 : Y_3 : Z_3)$. Then set

$$A = 4X_1 Y_1^2, \qquad B = 3X_1^2 + a_4 Z_1^4$$

and

$$X_3 = -2A + B^2, \qquad Y_3 = -8Y_1^4 + B(A - X_3), \qquad Z_3 = 2Y_1 Z_1.$$

The complexities are $12\mathrm{M} + 4\mathrm{S}$ for an addition and $4\mathrm{M} + 6\mathrm{S}$ for a doubling. If one of the points is given in the form $(X_1 : Y_1 : 1)$ the costs for addition reduce to $8\mathrm{M} + 3\mathrm{S}$.

The doubling involves one multiplication by the constant a_4. If it is small this multiplication can be performed by some additions and hence be neglected in the operation count. Especially if $a_4 = -3$ one can compute $T = 3X_1^2 - 3Z_1^4 = 3(X_1 - Z_1^2)(X_1 + Z_1^2)$ leading to only $4\mathrm{M} + 4\mathrm{S}$ for a doubling. Brier and Joye [BrJo 2003] study the use of isogenies to map a given curve to an isogenous one having this preferable parameter. Their conclusion is that for most randomly chosen curves there exists an isogeny of small degree mapping it to a curve with $a_4 = -3$, which justifies that the curves in the standards have this parameter.

The parameter $a_4 = 0$ is even more advantageous as the costs drop down to $3\mathrm{M} + 4\mathrm{S}$. However, this choice is far more special and the endomorphism ring $\mathrm{End}(E)$ contains a third root of unity.

In Jacobian coordinates, doublings are faster and additions slower than for the projective coordinates. To improve additions, a point P can be represented as a quintuple $(X_1, Y_1, Z_1, Z_1^2, Z_1^3)$. These coordinates are called *Chudnovsky Jacobian coordinates*. Additions and doublings are given by the same formulas as for $\mathcal{J}$ but the complexities are $11\mathrm{M} + 3\mathrm{S}$ and $5\mathrm{M} + 6\mathrm{S}$.

13.2.1.d Modified Jacobian coordinates ($\mathcal{J}^m$)

Modified Jacobian coordinates were introduced by Cohen et al. [CoMi^{+} 1998]. They are based on $\mathcal{J}$ but the internal representation of a point P is the quadruple $(X_1, Y_1, Z_1, a_4 Z_1^4)$. The formulas are essentially the same as for $\mathcal{J}$. The main difference is the introduction of $C = 8{Y_1}^4$ so that $Y_3 = B(A - X_3) - C$ and $a_4 Z_3^4 = 2C(a_4 Z_1^4)$ with the notation of Section 13.2.1.c. An addition takes $13\mathrm{M} + 6\mathrm{S}$ and a doubling $4\mathrm{M} + 4\mathrm{S}$. If one point is in affine coordinates, an addition takes $9\mathrm{M} + 5\mathrm{S}$. As I takes on average between 9 and $40\mathrm{M}$ and S is about $0.8\mathrm{M}$, this system offers the fastest doubling procedure.

13.2.1.e Example

Take

$$E_2 : y^2 = x^3 + 1132x + 278$$

and let $P_2 = (1120, 1391)$ and $Q_2 = (894, 1425)$ be two affine points on E_2. We recall below the equation and the internal representation of P_2 and Q_2 for each coordinate system. Note that for projective like systems we put Z to some random value and multiply X and Y by the respective powers.

System	Equation	P_2	Q_2
$\mathcal{A}$	$y^2 = x^3 + 1132x + 278$	$(1120, 1391)$	$(894, 1425)$
$\mathcal{P}$	$Y^2Z = X^3 + 1132XZ^2 + 278Z^3$	$(450 : 541 : 1449)$	$(1774 : 986 : 1530)$
$\mathcal{J}$	$Y^2 = X^3 + 1132XZ^4 + 278Z^6$	$(1213 : 408 : 601)$	$(1623 : 504 : 1559)$
$\mathcal{J}^c$	—	$(1213, 408, 601, 661, 667)$	$(1623, 504, 1559, 842, 713)$
$\mathcal{J}^m$	—	$(1213, 408, 601, 1794)$	$(1623, 504, 1559, 1232)$

With these particular values of P_2 and Q_2, let us compute $P_2 \oplus Q_2$, $[2]P_2$ and $[763]P_2$ within the different systems using the double and add method.

System	$P_2 \oplus Q_2$	$[2]P_2$	$[763]P_2$
$\mathcal{A}$	$(1683, 1388)$	$(1467, 143)$	$(1455, 882)$
$\mathcal{P}$	$(185 : 825 : 1220)$	$(352 : 504 : 956)$	$(931 : 1316 : 1464)$
$\mathcal{J}$	$(763 : 440 : 1934)$	$(1800 : 1083 : 1684)$	$(752 : 1146 : 543)$
$\mathcal{J}^c$	$(763, 440, 1934, 755, 1986)$	$(1800, 1083, 1684, 1611, 862)$	$(752, 1146, 543, 408, 1214)$
$\mathcal{J}^m$	$(763, 440, 1934, 1850)$	$(1800, 1083, 1684, 1119)$	$(752, 1146, 543, 1017)$

For each computation, one can check that we obtain a result equivalent to the affine one.

13.2.2 Mixed coordinates

To compute scalar multiples of a point one can use all the methods introduced in Chapter 9, especially the signed-digit representations, which are useful, as the negative of P is obtained by simply negating the y-coordinate.

The main idea here is to mix the different systems of coordinates defined above. This idea was already mentioned in adding an affine point to one in another system. In general, one can add points expressed in two different systems and give the result in a third one. For example $\mathcal{J} + \mathcal{J}^c = \mathcal{J}^m$ means that we add points in Jacobian and Chudnovsky Jacobian coordinates and express the result in the modified Jacobian system. So, we are going to choose the most efficient combination for each action we have to perform. See Table 13.3 on page 284 for a precise count of the required operations.

Precomputations

The following analysis is given in [CoMi$^+$ 1998, Section 4]. Suppose that we want to compute $[n]P$. We shall use the NAF$_w$ representation of n; see Section 9.1.4. So, we need to precompute $[i]P$ for each odd i such that $1 < i < 2^{w-1}$. For these precomputations, it is useful to choose either $\mathcal{A}$ if some inversions can be performed in the precomputation stage, or $\mathcal{J}^c$ otherwise, as these systems give rise to the most efficient (mixed) addition formulas. If $\mathcal{A}$ is selected, the Montgomery trick of simultaneous inversions in $\mathbb{F}_p$ should be used, cf. Algorithm 11.15. This leads to

$$(w-1)\mathrm{I} + \big(5 \times 2^{w-2} + 2w - 12\big)\mathrm{M} + \big(2^{w-2} + 2w - 5\big)\mathrm{S}$$

Table 13.3 *Operations required for addition and doubling.*

Doubling		Addition	
Operation	Costs	Operation	Costs
$2\mathcal{P}$	$7\mathrm{M}+5\mathrm{S}$	$\mathcal{J}^m+\mathcal{J}^m$	$13\mathrm{M}+6\mathrm{S}$
$2\mathcal{J}^c$	$5\mathrm{M}+6\mathrm{S}$	$\mathcal{J}^m+\mathcal{J}^c=\mathcal{J}^m$	$12\mathrm{M}+5\mathrm{S}$
$2\mathcal{J}$	$4\mathrm{M}+6\mathrm{S}$	$\mathcal{J}+\mathcal{J}^c=\mathcal{J}^m$	$12\mathrm{M}+5\mathrm{S}$
$2\mathcal{J}^m=\mathcal{J}^c$	$4\mathrm{M}+5\mathrm{S}$	$\mathcal{J}+\mathcal{J}$	$12\mathrm{M}+4\mathrm{S}$
$2\mathcal{J}^m$	$4\mathrm{M}+4\mathrm{S}$	$\mathcal{P}+\mathcal{P}$	$12\mathrm{M}+2\mathrm{S}$
$2\mathcal{A}=\mathcal{J}^c$	$3\mathrm{M}+5\mathrm{S}$	$\mathcal{J}^c+\mathcal{J}^c=\mathcal{J}^m$	$11\mathrm{M}+4\mathrm{S}$
$2\mathcal{J}^m=\mathcal{J}$	$3\mathrm{M}+4\mathrm{S}$	$\mathcal{J}^c+\mathcal{J}^c$	$11\mathrm{M}+3\mathrm{S}$
$2\mathcal{A}=\mathcal{J}^m$	$3\mathrm{M}+4\mathrm{S}$	$\mathcal{J}^c+\mathcal{J}=\mathcal{J}$	$11\mathrm{M}+3\mathrm{S}$
$2\mathcal{A}=\mathcal{J}$	$2\mathrm{M}+4\mathrm{S}$	$\mathcal{J}^c+\mathcal{J}^c=\mathcal{J}$	$10\mathrm{M}+2\mathrm{S}$
—	—	$\mathcal{J}+\mathcal{A}=\mathcal{J}^m$	$9\mathrm{M}+5\mathrm{S}$
—	—	$\mathcal{J}^m+\mathcal{A}=\mathcal{J}^m$	$9\mathrm{M}+5\mathrm{S}$
—	—	$\mathcal{J}^c+\mathcal{A}=\mathcal{J}^m$	$8\mathrm{M}+4\mathrm{S}$
—	—	$\mathcal{J}^c+\mathcal{A}=\mathcal{J}^c$	$8\mathrm{M}+3\mathrm{S}$
—	—	$\mathcal{J}+\mathcal{A}=\mathcal{J}$	$8\mathrm{M}+3\mathrm{S}$
—	—	$\mathcal{J}^m+\mathcal{A}=\mathcal{J}$	$8\mathrm{M}+3\mathrm{S}$
—	—	$\mathcal{A}+\mathcal{A}=\mathcal{J}^m$	$5\mathrm{M}+4\mathrm{S}$
—	—	$\mathcal{A}+\mathcal{A}=\mathcal{J}^c$	$5\mathrm{M}+3\mathrm{S}$
$2\mathcal{A}$	$\mathrm{I}+2\mathrm{M}+2\mathrm{S}$	$\mathcal{A}+\mathcal{A}$	$\mathrm{I}+2\mathrm{M}+\mathrm{S}$

for the precomputations. Note also that it is possible to avoid some doublings as explained in Remark 9.11 (iii).

Scalar multiplication

A scalar multiplication $[n]P$ consists of a sequence of doublings and additions. If a signed windowing method is used with precomputations, there are often runs of doublings interfered with only a few additions. Thus it is worthwhile to distinguish between intermediate doublings, i.e., those followed by a further doubling, and final doublings, which are followed by an addition and choose different coordinate systems for them. Cohen et al. propose to perform the intermediate doublings within $\mathcal{J}^m$ and to express the result of the last doubling in $\mathcal{J}$ since the next step is an addition. More explicitly, for each nonzero coefficient in the expansion of n the intermediate variable Q is replaced in each step by some

$$[2^s]Q \pm [u]P,$$

where $[u]P$ is in the set of precomputed multiples. So, we actually perform $(s-1)$ doublings of the type $2\mathcal{J}^m=\mathcal{J}^m$, a doubling of the form $2\mathcal{J}^m=\mathcal{J}$, and then an addition $\mathcal{J}+\mathcal{A}=\mathcal{J}^m$ or $\mathcal{J}+\mathcal{J}^c=\mathcal{J}^m$ depending on the coordinates of the precomputed values.

Let the windowing work as

$$n = 2^{n_0}(2^{n_1}(\cdots 2^{n_{v-1}}(2^{n_v}W[v]+W[v-1])\cdots)+W[0]),$$

where $W[i]$ is an odd integer in the range $-2^{w-1}+1 \leqslant W[i] \leqslant 2^{w-1}-1$ for all i, $W[v]>0, n_0 \geqslant 0$ and $n_i \geqslant w+1$ for $i \geqslant 1$.

In the main loop we perform $u = \sum_{i=0}^{v} n_i$ doublings and v additions. Put $l_1 = l - (w-1)/2$ and $K = 1/2 - 1/(w+1)$. On average $l_1 + K$ doublings and $(l_1 - K)/(w+1)$ additions are used.

Then we need approximately

$$\Big(l_1 + K + \frac{l_1 - K}{w+1}\Big)\mathrm{I} + \Big(2(l_1+K) + \frac{2}{w+1}(l_1 - K)\Big)\mathrm{M} + \Big(2(l_1+K) + \frac{2}{w+1}(l_1 - K)\Big)\mathrm{S}$$

to compute $[n]P$ excluding the costs for the precomputations if only affine coordinates are used,

$$\Big(4(l_1+K) + \frac{8}{w+1}(l_1 - K)\Big)\mathrm{M} + \Big(4(l_1+K) + \frac{5}{w+1}(l_1 - K)\Big)\mathrm{S}$$

if the precomputed points are in $\mathcal{A}$ and the computations are done without inversions using $\mathcal{J}$ and $\mathcal{J}^m$ for the intermediate points, and

$$\Big(4(l_1+K) + \frac{11}{w+1}(l_1 - K)\Big)\mathrm{M} + \Big(4(l_1+K) + \frac{5}{w+1}(l_1 - K)\Big)\mathrm{S}$$

if the precomputed points are in $\mathcal{J}^c$. Now depending on the ratio I/M, $\mathcal{A}$ or $\mathcal{J}^c$ should be chosen. For instance, for a 192-bits key length we choose $\mathcal{A}$ if I $<$ 33.9M and $\mathcal{J}^c$ otherwise, cf. [CoMi+ 1998].

13.2.3 Montgomery scalar multiplication

This technique was first described by Montgomery [Mon 1987] for a special type of curve in large characteristic and has been generalized to other curves and to even characteristic; see Section 13.3.4.

13.2.3.a Montgomery form

Let E_M be an elliptic curve expressed in *Montgomery form*, that is

$$E_M : By^2 = x^3 + Ax^2 + x. \tag{13.6}$$

The arithmetic on E_M relies on an efficient x-coordinate only computation and can be easily implemented to resist side-channel attacks, cf. Chapter 29. Indeed, let $P = (x_1, y_1)$ be a point on E_M. In projective coordinates, we write $P = (X_1 : Y_1 : Z_1)$ and let $[n]P = (X_n : Y_n : Z_n)$. The sum $[n+m]P = [n]P \oplus [m]P$ is given by the following formulas where Y_n never appears.

Addition: $n \neq m$

$$\begin{aligned} X_{m+n} &= Z_{m-n}\big((X_m - Z_m)(X_n + Z_n) + (X_m + Z_m)(X_n - Z_n)\big)^2, \\ Z_{m+n} &= X_{m-n}\big((X_m - Z_m)(X_n + Z_n) - (X_m + Z_m)(X_n - Z_n)\big)^2. \end{aligned}$$

Doubling: $n = m$

$$\begin{aligned} 4X_nZ_n &= (X_n + Z_n)^2 - (X_n - Z_n)^2, \\ X_{2n} &= (X_n + Z_n)^2(X_n - Z_n)^2, \\ Z_{2n} &= 4X_nZ_n\big((X_n - Z_n)^2 + \big((A+2)/4\big)(4X_nZ_n)\big). \end{aligned}$$

Thus an addition takes 4M and 2S whereas a doubling needs only 3M and 2S.

For some systems, the x-coordinate x_n of $[n]P$ is sufficient but others, like some signature schemes, need the y-coordinate as well, cf. Chapter 1. To recover $y_n = Y_n/Z_n$, we use the following formula [OKSA 2001]

$$y_n = \frac{(x_1 x_n + 1)(x_1 + x_n + 2A) - 2A - (x_1 - x_n)^2 x_{n+1}}{2By_1}, \tag{13.7}$$

where $P = (x_1, y_1)$ and x_n and x_{n+1} are the affine x-coordinates of $[n]P$ and $[n+1]P$.

13.2.3.b General case

Brier et Joye [BRJO 2002] generalized Montgomery's idea to any curve in short Weierstraß equation

$$E : y^2 = x^3 + a_4 x + a_6.$$

Their formulas require more elementary operations.

Addition: $n \neq m$

$$\begin{aligned} X_{m+n} &= Z_{m-n}\bigl(-4a_6 Z_m Z_n (X_m Z_n + X_n Z_m) + (X_m X_n - a_4 Z_m Z_n)^2\bigr), \\ Z_{m+n} &= X_{m-n}(X_m Z_n - X_n Z_m)^2. \end{aligned}$$

Doubling: $n = m$

$$\begin{aligned} X_{2n} &= (X_n^2 - a_4 Z_n^2)^2 - 8a_6 X_n Z_n^3, \\ Z_{2n} &= 4Z_n\bigl(X_n(X_n^2 + a_4 Z_n^2) + a_6 Z_n^3\bigr). \end{aligned}$$

When P is an affine point, an addition requires 9M and 2S whereas a doubling needs 6M and 3S.

To recover y_n in this case, we apply the formula

$$y_n = \frac{2a_6 + (x_1 x_n + a_4)(x_1 + x_n) - (x_1 - x_n)^2 x_{n+1}}{2y_1}.$$

13.2.3.c Transformation to Montgomery form

It is always possible to convert a curve in Montgomery form (13.6) into short Weierstraß equation, putting $a_4 = 1/B^2 - A^2/3B^2$ and $a_6 = -A^3/27B^3 - a_4 A/3B$. But the converse is false. Not all elliptic curves can be written in Montgomery form. However, this holds true as soon as $p \equiv 1 \pmod 4$ and $x^3 + a_4 x + a_6$ has three roots in $\mathbb{F}_p$. More generally, a curve in short Weierstraß form can be converted to Montgomery form if and only if

- the polynomial $x^3 + a_4 x + a_6$ has at least one root α in $\mathbb{F}_p$,
- the number $3\alpha^2 + a_4$ is a quadratic residue in $\mathbb{F}_p$.

Put $A = 3\alpha s, B = s$ where s is a square root of $(3\alpha^2 + a_4)^{-1}$ and the change of variables $(x, y) \mapsto (x/s + \alpha, y/s)$ is an isomorphism that transforms E into E_M. For such curves $(0, 0)$ is a point of order 2 and $|E(\mathbb{F}_p)|$ is divisible by 4.

Note that recent standards [SEC, NIST] recommend that the cardinality of E should be a prime number times a cofactor less than or equal to 4. One can state divisibility conditions in terms of the Legendre symbol $\left(\frac{\cdot}{p}\right)\cdot$ For a curve in Montgomery form $|E(\mathbb{F}_p)|$ is not divisible by 8 in the following cases:

$p \equiv 1 \pmod 4$			$p \equiv 3 \pmod 4$	
$\left(\frac{A+2}{p}\right)$	$\left(\frac{A-2}{p}\right)$	$\left(\frac{B}{p}\right)$	$\left(\frac{A+2}{p}\right)$	$\left(\frac{A-2}{p}\right)$
-1	$+1$	-1	-1	$+1$
$+1$	-1	-1		
$+1$	$+1$	-1		
-1	-1	$+1$		

Let v be a quadratic nonresidue and let $\widetilde{E}_v$ be the quadratic twist of E by v, cf. Example 13.19. Then either both E and $\widetilde{E}_v$ are transformable to Montgomery form or none is. Together with Schoof's point counting algorithm (see Section 17.2) this gives an efficient method for generating a curve transformable to Montgomery form whose cofactor is equal to 4.

Example 13.34 Let us show that $E_2/\mathbb{F}_p$

$$E_2 : y^2 = x^3 + 1132x + 278$$

can be expressed in Montgomery form.

First, $\alpha = 1702$ satisfies

$$\alpha^3 + 1132\alpha + 278 = 0$$

and $3\alpha^2 + a_4 = 527$ is a quadratic residue modulo $p = 2003$. Since $s = 899$ is an inverse square root of 527, we have $A = 1421$, $B = 899$ and the isomorphism $(x, y) \mapsto \big(899(x - 1702), 899y\big)$ maps the points of E_2 on the points of

$$E_{2,M} : 899y^2 = x^3 + 1421x^2 + x.$$

For instance, $P_2 = (1120, 1391)$ on E_2 is sent on $P_{2,M} = (1568, 637)$ on $E_{2,M}$.

13.2.3.d Montgomery ladder

Whatever the form of the curve, we use a modified version of Algorithm 9.5 adapted to scalar multiplication to compute $[n]P$.

Algorithm 13.35 Scalar multiplication using Montgomery's ladder

INPUT: A point P on E and a positive integer $n = (n_{\ell-1} \dots n_0)_2$.
OUTPUT: The point $[n]P$.

```
$P_1 \leftarrow P$ and $P_2 \leftarrow [2]P$
for $i = \ell - 2$ down to $0$ do
    if $n_i = 0$ then
        $P_1 \leftarrow [2]P_1$ and $P_2 \leftarrow P_1 \oplus P_2$
    else
        $P_1 \leftarrow P_1 \oplus P_2$ and $P_2 \leftarrow [2]P_2$
return $P_1$
```

Remarks 13.36

(i) At each step, one performs one addition and one doubling, which makes this method interesting against side-channel attacks, cf. Chapter 29.

(ii) We can check that $P_2 \ominus P_1$ is equal to P at each step so that $Z_{m-n} = Z_1$ in the formulas above. If P is expressed in affine coordinates this saves an extra multiplication in the addition. So the total complexity to compute $[n]P$ is $(6\mathrm{M} + 4\mathrm{S})(|n|_2 - 1)$ for elliptic curves in Montgomery form and $(14\mathrm{M} + 5\mathrm{S})(|n|_2 - 1)$ in short Weierstraß form.

Example 13.37 Let us compute $[763]P_{2,M}$ with Algorithm 13.35. We have $763 = (1011111011)_2$ and the different steps of the computation are given in the following table where P stands for $P_{2,M}$ and the question mark indicates that the y-coordinate is unknown.

i	n_i	(P_1, P_2)	P_1	P_2
9	1	$(P, [2]P)$	$(1568 : 637 : 1)$	$(35 : ? : 1887)$
8	0	$([2]P, [3]P)$	$(35 : ? : 1887)$	$(1887 : ? : 1248)$
7	1	$([5]P, [6]P)$	$(531 : ? : 162)$	$(120 : ? : 1069)$
6	1	$([11]P, [12]P)$	$(402 : ? : 1041)$	$(909 : ? : 1578)$
5	1	$([23]P, [24]P)$	$(1418 : ? : 1243)$	$(1389 : ? : 1977)$
4	1	$([47]P, [48]P)$	$(613 : ? : 37)$	$(1449 : ? : 231)$
3	1	$([95]P, [96]P)$	$(1685 : ? : 1191)$	$(1256 : ? : 842)$
2	0	$([190]P, [191]P)$	$(119 : ? : 1871)$	$(1501 : ? : 453)$
1	1	$([381]P, [382]P)$	$(1438 : ? : 956)$	$(287 : ? : 868)$
0	1	$([763]P, [764]P)$	$(568 : ? : 746)$	$(497 : ? : 822)$

To recover the y-coordinate of $[763]P_{2,M}$, we apply (13.7) with $x_1 = 1568, y_1 = 637, x_n$ and x_{n+1} respectively equal to $568/746$ and $497/822$. Finally, $[763]P_{2,M} = (280, 1733)$.

13.2.4 Parallel implementations

For the addition formulas in affine coordinates only a few field operations are used and, hence, parallelization is not too useful. In the other coordinate systems two processors can be applied to reduce the time for a group operation.

For Montgomery coordinates a parallel implementation using two processors is immediate, namely one can take care of the addition while the other performs the doubling. This is possible as both operations need about the same amount of operations, reducing the idle time.

Smart [Sma 2001] investigates parallel implementations of Hessian coordinates.

For Jacobian coordinates on arbitrary curves, Izu and Takagi [IzTa 2002a] propose a parallel version that additionally proposes methods for k-fold doubling. It can be implemented together with precomputations and windowing methods for scalar multiplication. Also [FiGi$^+$ 2002] deals with parallel implementation. We come back to efficient parallel implementations in the chapter on side-channel attacks, cf. Chapter 29.

13.2.5 Compression of points

For some applications it might be desirable to store or transmit as few bits as possible and still keep the same amount of information.

The following technique works for elliptic curves $E/\mathbb{F}_q$ over arbitrary finite fields $\mathbb{F}_q = \mathbb{F}_{p^d} = \mathbb{F}_p(\theta)$ of odd characteristic p (for details on the arithmetic of finite fields we refer to Chapter 11).

For an elliptic curve $E : y^2 = x^3 + a_2x^2 + a_4x + a_6$ there are at most two points with the same x-coordinate, namely $P = (x_1, y_1)$ and $-P = (x_1, -y_1)$. They are equal if and only if $y_1 = 0$, i.e., for the Weierstraß points.

Compression

To uniquely identify the point one saves x_1 and one bit $b(y_1)$. It is set to 0 if in the field representation $y_1 = \sum_{i=0}^{d-1} c_i\theta^i$ the value of c_0 taken as a nonnegative integer is even and set to 1 otherwise. This procedure works as $-y_1$ has $p - c_0$ as its least significant coefficient, which is of opposite parity as p is odd. Hence, one simply needs to check for the least significant bit of the least significant coefficient of y_1.

Decompression

To recover the y-coordinate from $\big(x_1, b(y_1)\big)$ some more work needs to be done. Namely, one evaluates $x_1^3 + a_2x_1^2 + a_4x_1 + a_6$, which has to be a square in $\mathbb{F}_q$ since x_1 is the x-coordinate of a point on E. Algorithms for square root computation, cf. Section 11.1.5, allow us to recover the two values $\pm\, y_1$ and the bit $b(y_1)$ determines the correct y-coordinate.

Example 13.38 On the curve $E_2/\mathbb{F}_p$ the point $P_2 = (1120, 1391) \in E(\mathbb{F}_p)$ is coded by $(1120, 1)$ while $R_2 = (443\theta^2 + 1727\theta + 1809, 929\theta^2 + 280\theta + 946) \in E(\mathbb{F}_{p^3})$ is represented by $(443\theta^2 + 1727\theta + 1809, 0)$.

13.3 Arithmetic of elliptic curves defined over $\mathbb{F}_{2^d}$

In this section we consider elliptic curves over $\mathbb{F}_{2^d}$. We first provide the transfer to short Weierstraß equations and state formulas for the arithmetic on supersingular and ordinary elliptic curves in affine coordinates. For the remainder of the section we concentrate on ordinary curves. The curves given in Weierstraß form

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

can be transformed depending on the value of a_1.

Supersingular curves

If $a_1 = 0$, we need to have $a_3 \neq 0$ as otherwise the curve is singular. The transformation $x \mapsto x + a_2$ leads to the equation

$$E : y^2 + a_3y = x^3 + a_4'x + a_6',$$

which is nonsingular as $a_3 \neq 0$. Such a curve E has no point $P = (x_1, y_1)$ of order two over $\overline{\mathbb{F}}_{2^d}$, as these satisfy $P = -P$, i.e., $y_1 = y_1 + a_3$ and this would only be true for $a_3 = 0$. Therefore, $E[2] = \{P_\infty\}$ and E is supersingular by Definition 13.14.

In Section 24.2.1, we extensively study supersingular curves as they come with an efficiently computable pairing. This has many consequences. For instance, the DLP is easier to solve for these curves. However, there also exist constructive aspects of pairings, e.g., see Chapter 24, and this justifies to investigate the arithmetic of these curves. Indeed, the arithmetic on the supersingular curve

$$E : y^2 + a_3y = x^3 + a_4x + a_6$$

is given by the following formulas where $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ are two points in $E(\mathbb{F}_{2^d})$

- $-P = (x_1, y_1 + a_3)$.
- if $P \neq \pm Q$, we have $P \oplus Q = (x_3, y_3)$ where

$$x_3 = \lambda^2 + x_1 + x_2, \qquad y_3 = \lambda(x_1 + x_3) + y_1 + a_3, \qquad \lambda = \frac{y_1 + y_2}{x_1 + x_2}.$$

- if $P \neq -P$, we have $[2]P = (x_3, y_3)$ where

$$x_3 = \lambda^2, \qquad y_3 = \lambda(x_1 + x_3) + y_1 + a_3, \qquad \lambda = \frac{x_1^2 + a_4}{a_3}.$$

Example 13.39 Let us consider $\mathbb{F}_{2^{11}}$, represented as $\mathbb{F}_2(\theta)$ with $\theta^{11} + \theta^2 + 1 = 0$. The elements of $\mathbb{F}_{2^{11}}$ will be represented using hexadecimal basis. For instance, `0x591` corresponds to the sequence of bits (0101 1001 0001) and therefore stands for the element $\theta^{10} + \theta^8 + \theta^7 + \theta^4 + 1$.

A supersingular elliptic curve over $\mathbb{F}_{2^{11}}$ is given by

$$E_5 : y^2 + a_3 y = x^3 + a_4 x + a_6$$

with $a_3 = \texttt{0x6EE}$, $a_4 = \texttt{0x1CC}$ and $a_6 = \texttt{0x3F6}$. The discriminant of E_5 is $\Delta = \texttt{0x722}$ while its j-invariant is zero.

The points $P_5 = (\texttt{0x3DF}, \texttt{0x171})$ and $Q_5 = (\texttt{0x732}, \texttt{0x27D})$ belong to $E_5(\mathbb{F}_{2^{11}})$ and

$$\begin{aligned} -P_5 &= (\texttt{0x3DF}, \texttt{0x79F}), \\ P_5 \oplus Q_5 &= (\texttt{0x314}, \texttt{0x4BC}), \\ [2]P_5 &= (\texttt{0xEF}, \texttt{0x6C3}). \end{aligned}$$

The cardinality of $E_5(\mathbb{F}_{2^{11}})$ is equal to $2^{11} + 2^6 + 1 = 2113$ which is prime. Thus the group $E_5(\mathbb{F}_{2^{11}})$ is cyclic and is generated by any one of its element.

Ordinary curves

If $a_1 \neq 0$, the transformations

$$y \mapsto a_1^3 y + \frac{a_3^2 + a_1^2 a_4}{a_1^3}, \quad x \mapsto a_1^2 x + \frac{a_3}{a_1}$$

followed by a division by a_1^6 lead to an isomorphic curve given by

$$y^2 + xy = x^3 + a_2' x^2 + a_6',$$

which is nonsingular whenever $a_6' \neq 0$. In this case, the curve is ordinary.

Remark 13.40 It is always possible to choose a_2' small in the sense that multiplications by a_2' can be carried out by a few additions only. Let c be an element of absolute trace 0, i.e., $\mathrm{Tr}_{\mathbb{F}_{2^d}/\mathbb{F}_2}(c) = 0$, such that multiplications by $a_2' + c$ can be carried out efficiently. In practice, d should be odd, (cf. Section 23.2.2.c) and in this case if $\mathrm{Tr}_{\mathbb{F}_{2^d}/\mathbb{F}_2}(a_2') = 1$ then $\mathrm{Tr}_{\mathbb{F}_{2^d}/\mathbb{F}_2}(a_2' + 1) = 0$. So in any case, c can be taken equal to a_2' or $a_2' + 1$ with the result that $a_2' + c$ is an element of $\mathbb{F}_2$. Let λ be such that $\lambda^2 + \lambda + c = 0$. Indeed, (13.8) allows for a further transformation

$$x \mapsto x, \quad y \mapsto y + \lambda x,$$

which leads to the curve

$$y^2 + xy = x^3 + (a_2' + c)x^2 + a_6'.$$

Example 13.41 An ordinary elliptic curve over $\mathbb{F}_{2^{11}}$ is given by

$$E_6 : y^2 + xy = x^3 + a_2x^2 + a_6$$

with $a_2 = \texttt{0x6EE}$ and $a_6 = \texttt{0x1CC}$. As the trace of a_2 is 1 we can put $c = a_2 + 1$ which is of trace 0 and find $\lambda = \texttt{0x51E}$ such that $\lambda^2 + \lambda = c$. Now the change of variables $x \mapsto x,\ y \mapsto y + \lambda x$, with $\lambda = \texttt{0x68B}$ transforms the curve E_6 into

$$E_7 : y^2 + xy = x^3 + x^2 + a_6.$$

The discriminant of E_7 is $\Delta = a_6$ and its j-invariant is $1/a_6 = \texttt{0x37F}$. The points $P_7 = (\texttt{0x420}, \texttt{0x5B3})$ and $Q_7 = (\texttt{0x4B8}, \texttt{0x167})$ are on E_7. The curve E_7 has 2026 rational points in $\mathbb{F}_{2^{11}}$ and $E_7(\mathbb{F}_{2^{11}})$ is cyclic generated by P_7.

13.3.1 Choice of the coordinates

The remainder of this chapter is entirely devoted to ordinary curves, i.e., curves given by

$$E : y^2 + xy = x^3 + a_2x^2 + a_6, \tag{13.8}$$

with $a_2, a_6 \in \mathbb{F}_{2^d}$ such that $a_6 \neq 0$. The coefficient a_2 can be chosen with a reduced number of terms and can even be taken in $\mathbb{F}_2$ when d is odd, cf. Remark 13.40 for explanations.

We first give a study on the addition formulas in different coordinate systems and study mixed coordinate systems, then give a generalization of Montgomery coordinates and introduce a further endomorphism on the curve, the point halving. Finally we discuss compression techniques.

As in Section 13.2, an elementary multiplication in $\mathbb{F}_{2^d}$ (respectively a squaring and an inversion) will be represented by M (respectively S and I).

This section is mainly based on [HALÓ$^+$ 2000]. As for curves over prime fields we study different systems of coordinates, namely affine, projective, Jacobian and López–Dahab. For these binary fields some extra tricks are applicable.

We shall give formulas for the

- addition of two points P and $Q \in E(\mathbb{F}_{2^d})$ provided $P \neq \pm Q$,
- doubling of P.

13.3.1.a Affine coordinates ($\mathcal{A}$)

Recall that we can choose an elliptic curve of the form

$$E : y^2 + xy = x^3 + a_2x^2 + a_6.$$

The opposite of $P = (x_1, y_1)$ equals $-P = (x_1, x_1 + y_1)$.

Addition

Let $P = (x_1, y_1)$, $Q = (x_2, y_2)$ such that $P \neq \pm Q$ then $P \oplus Q = (x_3, y_3)$ is given by

$$x_3 = \lambda^2 + \lambda + x_1 + x_2 + a_2, \qquad y_3 = \lambda(x_1 + x_3) + x_3 + y_1, \qquad \lambda = \frac{y_1 + y_2}{x_1 + x_2}\cdot$$

Doubling

Let $P = (x_1, y_1)$ then $[2]P = (x_3, y_3)$, where

$$x_3 = \lambda^2 + \lambda + a_2, \qquad y_3 = \lambda(x_1 + x_3) + x_3 + y_1, \qquad \lambda = x_1 + \frac{y_1}{x_1}\cdot$$

Thus an addition and a doubling require exactly the same number of operations, that is, $\mathrm{I} + 2\mathrm{M} + \mathrm{S}$.

Doubling followed by an addition

Extending an idea presented in Section 13.2.1.a (see also [EiLa+ 2003]), Ciet et al. [CiJo+ 2003], propose a method to compute $[2]P \oplus Q$ as as single operation. The formulas are given below where we assume that $P \neq \pm Q$ and $[2]P \neq -Q$

$$\begin{aligned} &A = x_2 + x_1, && B = y_2 + y_1, && C = A^2(x_2 + a_2) + B(B + A),\\ &D = (AC)^{-1}, && \lambda = BCD, && \lambda_2 = A^3 D x_1 + \lambda + 1,\\ &x_4 = (\lambda + \lambda_2)^2 + \lambda + \lambda_2 + x_2, && y_4 = (x_1 + x_4)\lambda_2 + y_1 + x_4, && \end{aligned}$$

requiring $\mathrm{I} + 9\mathrm{M} + 2\mathrm{S}$.

13.3.1.b Projective coordinates ($\mathcal{P}$)

With *projective coordinates* the curve is parameterized by the equation

$$Y^2Z + XYZ = X^3 + a_2X^2Z + a_6Z^3.$$

Like in odd characteristic, we let $(X_1 : Y_1 : Z_1)$ represent the affine point $(X_1/Z_1, Y_1/Z_1)$ if $Z_1 \neq 0$ and $P_\infty = (0 : 1 : 0)$ otherwise. The opposite of $(X_1 : Y_1 : Z_1)$ is $(X_1 : X_1 + Y_1 : Z_1)$.

Addition

Let $P = (X_1 : Y_1 : Z_1)$, $Q = (X_2 : Y_2 : Z_2)$ such that $P \neq \pm Q$ then $P \oplus Q = (X_3 : Y_3 : Z_3)$ is given by

$$\begin{aligned} &A = Y_1Z_2 + Z_1Y_2, && B = X_1Z_2 + Z_1X_2, && C = B^2,\\ &D = Z_1Z_2, && E = (A^2 + AB + a_2C)D + BC, && \\ &X_3 = BE, && Y_3 = C(AX_1 + Y_1B)Z_2 + (A + B)E, && Z_3 = B^3D. \end{aligned}$$

Doubling

If $P = (X_1 : Y_1 : Z_1)$ then $[2]P = (X_3 : Y_3 : Z_3)$ is given by

$$\begin{aligned} &A = X_1^2, && B = A + Y_1Z_1, && C = X_1Z_1,\\ &D = C^2, && E = (B^2 + BC + a_2D), && \\ &X_3 = CE, && Y_3 = (B + C)E + A^2C, && Z_3 = CD. \end{aligned}$$

In projective coordinates, no inversion is needed. An addition needs $16\mathrm{M} + 2\mathrm{S}$ and a doubling requires $8\mathrm{M} + 4\mathrm{S}$.

If the addition receives one input point in affine coordinates, i.e., as $(X_2 : Y_2 : 1)$, the costs reduce to $12\mathrm{M} + 2\mathrm{S}$. Such an *addition in mixed coordinates* is studied in larger generality in the next section.

All operations profit from small a_2 as one multiplication is saved.

13.3.1.c Jacobian coordinates ($\mathcal{J}$)

In *Jacobian coordinates*, the curve is given by the equation

$$Y^2 + XYZ = X^3 + a_2X^2Z^2 + a_6Z^6.$$

The point represented by $(X_1 : Y_1 : Z_1)$ corresponds to the affine point $(X_1/Z_1^2, Y_1/Z_1^3)$ when $Z_1 \neq 0$ and to $P_\infty = (1 : 1 : 0)$ otherwise. The opposite of $(X_1 : Y_1 : Z_1)$ is $(X_1 : X_1Z_1 + Y_1 : Z_1)$.

Addition

Let $P=(X_1:Y_1:Z_1)$, $Q=(X_2:Y_2:Z_2)$ such that $P\neq\pm Q$ then $P\oplus Q=(X_3:Y_3:Z_3)$ is given by

$$\begin{array}{lll} A=X_1Z_2^2, & B=X_2Z_1^2, & C=Y_1Z_2^3,\\ D=Y_2Z_1^3, & E=A+B, & F=C+D,\\ G=EZ_1, & H=FX_2+GY_2, & Z_3=GZ_2,\\ I=F+Z_3, & X_3=a_2Z_3^2+FI+E^3, & Y_3=IX_3+G^2H. \end{array}$$

Doubling

If $P=(X_1:Y_1:Z_1)$ then $[2]P=(X_3:Y_3:Z_3)$ is given by

$$\begin{array}{lll} A=X_1^2, & B=A^2, & C=Z_1^2,\\ X_3=B+a_6C^4, & Z_3=X_1C, & Y_3=BZ_3+(A+Y_1Z_1+Z_3)X_3. \end{array}$$

In Jacobian coordinates an addition requires $16\mathrm{M}+3\mathrm{S}$ in general and only $11\mathrm{M}+3\mathrm{S}$ if one input is in affine coordinates. Also if $a_2\in\{0,1\}$ we need one multiplication less in the addition of points. A doubling needs $5\mathrm{M}+5\mathrm{S}$ including one multiplication by a_6.

13.3.1.d López–Dahab coordinates ($\mathcal{LD}$)

López and Dahab [LóDa 1998] introduced a further set of coordinates in which the curve is given by the equation

$$Y^2+XYZ=X^3Z+a_2X^2Z^2+a_6Z^4.$$

The triple $(X_1:Y_1:Z_1)$ represents the affine point $(X_1/Z_1, Y_1/Z_1^2)$ when $Z_1\neq 0$ and $P_\infty=(1:0:0)$ otherwise. The opposite of $(X_1:Y_1:Z_1)$ is $(X_1:X_1Z_1+Y_1:Z_1)$.

Addition

Let $P=(X_1:Y_1:Z_1)$, $Q=(X_2:Y_2:Z_2)$ such that $P\neq\pm Q$ then $P\oplus Q=(X_3:Y_3:Z_3)$ is given by

$$\begin{array}{lll} A=X_1Z_2, & B=X_2Z_1, & C=A^2,\\ D=B^2, & E=A+B, & F=C+D,\\ G=Y_1Z_2^2, & H=Y_2Z_1^2, & I=G+H,\\ J=IE, & Z_3=FZ_1Z_2, & X_3=A(H+D)+B(C+G),\\ Y_3=(AJ+FG)F+(J+Z_3)X_3. & & \end{array}$$

A general addition $P\oplus Q$ in this coordinate system takes $13\mathrm{M}+4\mathrm{S}$ as shown by Higuchi and Takagi [HiTa 2000]. Note that the original formulas proposed in [LóDa 1998] need $14\mathrm{M}+6\mathrm{S}$.

Mixed Addition

If Q is in affine coordinates the costs drop to $10\mathrm{M}+3\mathrm{S}$. In fact, it is possible to do a bit better, since Al–Daoud et al. [AlMa$^+$ 2002] proved that only $9\mathrm{M}+5\mathrm{S}$ are sufficient in this case. The formulas are given below.

$$\begin{array}{lll} A=Y_1+Y_2Z_1^2, & B=X_1+X_2Z_1, & C=BZ_1,\\ Z_3=C^2, & D=X_2Z_3, & X_3=A^2+C(A+B^2+a_2C),\\ Y_3=(D+X_3)(AC+Z_3)+(Y_2+X_2)Z_3^2. & & \end{array}$$

Note that when $a_2 \in \{0,1\}$ one further multiplication is saved.

Doubling

If $P = (X_1 : Y_1 : Z_1)$ then $[2]P = (X_3 : Y_3 : Z_3)$ is given by [LÓDA 1998]

$$\begin{aligned} A &= Z_1^2, & B &= a_6 A^2, & C &= X_1^2, \\ Z_3 &= AC, & X_3 &= C^2 + B, & Y_3 &= (Y_1^2 + a_2 Z_3 + B)X_3 + Z_3 B. \end{aligned}$$

To analyze the complexity, first note that in practice a_2 can be chosen in $\mathbb{F}_2$, cf. Remark 13.40, saving one product.
For fixed a_2 and a_6 it is also possible to use less additions if $\sqrt{a_6}$ can be precomputed. E.g., for $a_2 = 1$ one can use

$$\begin{aligned} A &= X_1^2, & B &= \sqrt{a_6} Z_1^2, & C &= X_1 Z_1, \\ Z_3 &= C^2, & X_3 &= (A + B)^2, & Y_3 &= \bigl(AC + (Y_1 + B)(A + B)\bigr)^2 \end{aligned}$$

requiring 4M + 5S including one multiplication by $\sqrt{a_6}$.

For fixed $a_2 = 0$, X_3 and Z_3 are given as above whereas $Y_3 = \bigl(BC + (Y_1 + B)(A + B)\bigr)^2$, which also requires 4M + 5S including one multiplication by $\sqrt{a_6}$.

It is also possible to trade this multiplication by a constant and a squaring for a general multiplication [LAN 2004b], which might be interesting if the curve varies or if $\sqrt{a_6}$ is big. The formulas are as follows

$$\begin{aligned} A &= X_1 Z_1, & B &= X_1^2, & C &= B + Y_1, \\ D &= AC, & Z_3 &= A^2, & X_3 &= C^2 + D + a_2 Z_3, \\ Y_3 &= (Z_3 + D)X_3 + B^2 Z_3 \end{aligned} \tag{13.9}$$

requiring 5M + 4S including one multiplication by a_2.

13.3.1.e Example

Take the curve

$$E_7 : y^2 + xy = x^3 + x^2 + a_6 \tag{13.10}$$

with $a_6 = \texttt{0x1CC}$. We recall below the equation of E_7 as well as the coordinates of $P_7 = (\texttt{0x420}, \texttt{0x681})$ and $Q_7 = (\texttt{0x4B8}, \texttt{0x563})$ on E_7 for each coordinate system. Note that the third coordinate in projective, Jacobian and López–Dahab systems is chosen at random.

System	Equation	P_7	Q_7
$\mathcal{A}$	$y^2 + xy = x^3 + x^2 + a_6$	(0x420, 0x5B3)	(0x4B8, 0x167)
$\mathcal{P}$	$Y^2Z + XYZ = X^3 + X^2Z + a_6Z^3$	(0x64F : 0x5BA : 0x1C9)	(0x4DD : 0x1F0 : 0x3FA)
$\mathcal{J}$	$Y^2 + XYZ = X^3 + X^2Z^2 + a_6Z^6$	(0x4DA : 0x1F7 : 0x701)	(0x383 : 0x5BA : 0x1E1)
$\mathcal{LD}$	$Y^2 + XYZ = X^3Z + X^2Z^2 + a_6Z^4$	(0x6BE : 0x15F : 0x7B3)	(0x757 : 0x3EF : 0xA1C)

With these particular values of P_7 and Q_7, let us compute $P_7 \oplus Q_7$, $[2]P_7$ and $[763]P_7$ within the different systems using the double and add method.

System	$P_7 \oplus Q_7$	$[2]P_7$	$[763]P_7$
$\mathcal{A}$	$(\mathtt{0x724}, \mathtt{0x7B3})$	$(\mathtt{0x14D}, \mathtt{0x4CB})$	$(\mathtt{0x84}, \mathtt{0x475})$
$\mathcal{P}$	$(\mathtt{0x675} : \mathtt{0x6D5} : \mathtt{0x4D5})$	$(\mathtt{0x4D5} : \mathtt{0x21E} : \mathtt{0x705})$	$(\mathtt{0x582} : \mathtt{0x14} : \mathtt{0x543})$
$\mathcal{J}$	$(\mathtt{0x12} : \mathtt{0x46B} : \mathtt{0x5F})$	$(\mathtt{0x5B1} : \mathtt{0x417} : \mathtt{0x7D})$	$(\mathtt{0x2F7} : \mathtt{0x572} : \mathtt{0x3E2})$
$\mathcal{LD}$	$(\mathtt{0x7C5} : \mathtt{0x1D2} : \mathtt{0x3D2})$	$(\mathtt{0x444} : \mathtt{0x4A0} : \mathtt{0x193})$	$(\mathtt{0x2F} : \mathtt{0x265} : \mathtt{0x220})$

For each computation, the obtained result is equivalent to the affine one.

13.3.2 Faster doublings in affine coordinates

Let $P = (x_1, y_1)$ be a point lying on the ordinary curve

$$E : y^2 + xy = x^3 + a_2x^2 + a_6.$$

When the solution of a quadratic equation can be quickly found, e.g., if $\mathbb{F}_{2^d}$ is represented by a normal basis, the following method [SOL 1997] replaces one general multiplication by a multiplication by the fixed constant a_6.

Namely, compute $x_3 = x_1^2 + a_6/x_1^2$, which is also equal to $\lambda^2 + \lambda + a_2$. Then find μ such that $\mu^2+\mu = x_3+a_2$, see Section 11.2.6. So $\lambda = \mu+\varepsilon$ where $\varepsilon = 0$ or 1. Therefore $\mu x_1+x_1^2+y_1 = \varepsilon x_1$ and we deduce ε from this equation. Note that it is not necessary to perform μx_1 in full but rather to compute one well chosen coordinate in the product. Thus the computation of λ is almost free and it remains to perform $y_3 = x_1^2 + (\lambda + 1)x_3$.

To perform several doublings in a row of $P = (x_1, y_1)$, it is faster to store the intermediate values by the x-coordinate and the slope of the tangent, i.e., to represent $[2^i]P$ as (x_{2^i}, λ_{2^i}). This is possible because

$$\begin{aligned}
x_2 &= \lambda_1^2 + \lambda_1 + a_2 \text{ and} \\
\lambda_2 &= \lambda_1^2 + \lambda_1 + a_2 + \frac{\lambda_1(x_1 + \lambda_1^2 + \lambda_1 + a_2) + \lambda_1^2 + \lambda_1 + a_2 + y_1}{\lambda_1^2 + \lambda_1 + a_2} \\
&= \lambda_1^2 + a_2 + \frac{a_6}{x_1^4 + a_6}.
\end{aligned}$$

This idea leads to the following algorithm described in [LÓDA 2000b].

Algorithm 13.42 Repeated doublings

INPUT: A point $P = (x_1, y_1)$ on E such that $[2^k]P \neq P_\infty$ and an integer $k \geqslant 2$.
OUTPUT: The point $[2^k]P$ of coordinates (x_3, y_3).

```
$\lambda \leftarrow x_1 + y_1/x_1$ and $u \leftarrow x_1$
for $i = 1$ to $k - 1$ do
    $x' \leftarrow \lambda^2 + \lambda + a_2$
    $\lambda' \leftarrow \lambda^2 + a_2 + \frac{a_6}{u^4 + a_6}$
    $u \leftarrow x'$ and $\lambda \leftarrow \lambda'$
$x_3 \leftarrow \lambda^2 + \lambda + a_2$ and $y_3 \leftarrow u^2 + (\lambda + 1)x_3$
return $(x_3, y_3)$
```

This algorithm needs $k\mathrm{I} + (k+1)\mathrm{M} + (3k-1)\mathrm{S}$.

Example 13.43 Take P_7 on E_7 as defined in Example 13.41 and let us compute $[2^5]P_7$. The values of λ and u along the execution of Algorithm 13.42 are given below.

i	—	1	2	3	4
λ	0x1C	0x67	0x6F7	0x96	0x719
u	0x420	0x14D	0x479	0x344	0x1AB

At the end, $x_3 =$ 0x67C and $y_3 =$ 0x71C.

Another strategy is to use a closed formula to get $[2^k]P$ directly rather than computing successive doublings. The interest is to perform only one inversion at the cost of extra multiplications [GuPa 1997]. We do not state these formulas here as the same number of operations can be obtained by using López–Dahab coordinates for the intermediate doublings and transforming the result to affine coordinates afterwards.

13.3.3 Mixed coordinates

In the previous part we introduced different representations for the point on E together with the algorithms to perform addition and doubling. For the additions we also mentioned the number of operations needed if one of the input points is in affine coordinates. Like in odd characteristic we now study arbitrary mixes of coordinates to perform scalar multiplications where we use two (different) systems of coordinates as input and one as output. By $\mathcal{J} + \mathcal{A} = \mathcal{LD}$ we denote the addition taking as input one point in Jacobian coordinates and one in affine and giving the result in López–Dahab coordinates.

Additionally we use the abbreviations $\mathcal{A}'$ to denote the representation by (x, λ) introduced in the previous section for multiple doublings. For $\mathcal{A}'$ coordinates the table entry refers to the asymptotic complexity of a doubling in a sequence of k consecutive doublings, thus we neglect other marginal operations. Table 13.4 on page 297 gives the number of field operations needed depending on the coordinate systems. Compared to the case of odd characteristic, changes between the coordinate systems are not too interesting and are therefore not listed. We denote the costs for multiplication with a_2 by M_2 and concentrate on the most interesting cases. We do *not* take into account the effects of small a_6 as this cannot be achieved generically.

No precomputation

If the system offers no space to store precomputations one should use $\mathcal{A}$ if inversions are affordable, i.e., less than 8 times as expensive as a multiplication, and otherwise use $\mathcal{LD}$ for the doublings and $\mathcal{LD} + \mathcal{A} = \mathcal{LD}$ for the additions if the input is in affine coordinates and as $\mathcal{LD} + \mathcal{LD}$ otherwise.

Precomputations

Also in even characteristic, using the NAF_w representation, cf. Section 9.1.4 is advantageous to compute scalar multiples $[n]P$. This requires precomputing all odd multiples $[i]P$ for $1 < i < 2^{w-1}$. They can be obtained as a sequence of additions and one doubling.

If inversions in $\mathbb{F}_{2^d}$ are not too expensive one should choose affine coordinates as a system for the precomputations as they offer the fastest mixed coordinates. Like in the case of odd characteristic this does not mean that one needs to perform 2^{w-2} inversions but one can follow [CoMi+ 1998, section 4] and apply Montgomery's trick of simultaneous inversions. For details we refer to the study for odd characteristic, cf. Section 13.2.2. Then one needs:

$$(w-1)\mathrm{I} + \left(5 \times 2^{w-2} + 2w - 12\right)\mathrm{M} + \left(2^{w-2} + w - 3\right)\mathrm{S}.$$

Table 13.4 *Operations required for addition and doubling.*

Doubling		Addition	
Operation	Costs	Operation	Costs
$2\mathcal{P}$	$7\mathrm{M} + 4\mathrm{S} + \mathrm{M}_2$	$\mathcal{J} + \mathcal{J}$	$15\mathrm{M} + 3\mathrm{S} + \mathrm{M}_2$
$2\mathcal{J}$	$5\mathrm{M} + 5\mathrm{S}$	$\mathcal{P} + \mathcal{P}$	$15\mathrm{M} + 2\mathrm{S} + \mathrm{M}_2$
$2\mathcal{LD}$	$4\mathrm{M} + 4\mathrm{S} + \mathrm{M}_2$	$\mathcal{LD} + \mathcal{LD}$	$13\mathrm{M} + 4\mathrm{S}$
$2\mathcal{A} = \mathcal{P}$	$5\mathrm{M} + 2\mathrm{S} + \mathrm{M}_2$	$\mathcal{P} + \mathcal{A} = \mathcal{P}$	$11\mathrm{M} + 2\mathrm{S} + \mathrm{M}_2$
$2\mathcal{A} = \mathcal{LD}$	$2\mathrm{M} + 3\mathrm{S} + \mathrm{M}_2$	$\mathcal{J} + \mathcal{A} = \mathcal{J}$	$10\mathrm{M} + 3\mathrm{S} + \mathrm{M}_2$
$2\mathcal{A} = \mathcal{J}$	$\mathrm{M} + 2\mathrm{S} + \mathrm{M}_2$	$\mathcal{LD} + \mathcal{A} = \mathcal{LD}$	$8\mathrm{M} + 5\mathrm{S} + \mathrm{M}_2$
—	—	$\mathcal{A} + \mathcal{A} = \mathcal{LD}$	$5\mathrm{M} + 2\mathrm{S} + \mathrm{M}_2$
$2\mathcal{A}$	$\mathrm{I} + 2\mathrm{M} + \mathrm{S}$	$\mathcal{A} + \mathcal{A} = \mathcal{J}$	$4\mathrm{M} + \mathrm{S} + \mathrm{M}_2$
$2\mathcal{A}'$	$\mathrm{I} + \mathrm{M} + \mathrm{S}$	$\mathcal{A} + \mathcal{A} = \mathcal{A}'$	$2\mathrm{I} + 3\mathrm{M} + \mathrm{S}$
$2\mathcal{A}' = \mathcal{A}$	$\mathrm{M} + 2\mathrm{S}$	$\mathcal{A} + \mathcal{A}$	$\mathrm{I} + 2\mathrm{M} + \mathrm{S}$

If inversions are prohibitively expensive one should choose $\mathcal{LD}$ coordinates as they are the most efficient inversion-free system, provided that one multiplication is at least as expensive as three squarings, which is usually the case in binary fields. This way

$$\left(13 \times 2^{w-2} - 8\right)\mathrm{M} + 4 \times 2^{w-2}\mathrm{S}$$

are needed for the precomputations.

Using one I and $3(2^{w-2} - 2)\mathrm{M}$ the resulting precomputed points can be transformed to affine. Furthermore, the use of precomputations leads to long runs of doublings in the algorithms and they are much faster in $\mathcal{LD}$ than in $\mathcal{P}$, which otherwise would offer the lowest number of operations per addition.

Scalar multiplication

A scalar multiplication consists of a sequence of doublings and additions. If a signed windowing method is used with precomputations there are often runs of doublings interfered with only a few additions. Thus it is worthwhile to distinguish between intermediate doublings, i.e., those followed by a further doubling, and final doublings, which are followed by an addition, and to choose different coordinate systems for them.

We first assume that the precomputed points are in $\mathcal{A}$ as this leads to the most interesting mixes of coordinates. If one inversion per bit of the scalar is affordable one should use $\mathcal{A}'$ for the intermediate doublings.

More explicitly, the intermediate variable Q is replaced each step by some

$$[2^s]Q \pm [u]P,$$

where $[u]P$ is in the set of precomputed multiples. So we actually perform $(s-1)$ doublings of the type $2\mathcal{A}' = \mathcal{A}'$, a doubling of the form $2\mathcal{A}' = \mathcal{A}$ and then an addition $\mathcal{A} + \mathcal{A} = \mathcal{A}'$.

Let l be the binary length of n, let $l_1 = l - (w-1)/2$, and $K = 1/2 - 1/(w+1)$.

In the main loop, we perform on average $l_1 + K - v$ doublings of the form $2\mathcal{A}' = \mathcal{A}'$, v doublings of the form $2\mathcal{A}' = \mathcal{A}$, and v additions, where $v = (l_1 - K)/(w+1)$. Then we need approximately

$$\Big(l_1 + K + \frac{l_1 - K}{w+1}\Big)\mathrm{I} + \Big(l_1 + K + 3\frac{l_1 - K}{w+1}\Big)\mathrm{M} + \Big(l_1 + K + 2\frac{l_1 - K}{w+1}\Big)\mathrm{S}.$$

If the algorithm should not make use of inversions but the precomputed points are in $\mathcal{A}$, Table 13.4 shows that the doublings should be performed within $\mathcal{LD}$. This is followed by one addition of the type $\mathcal{LD} + \mathcal{A} = \mathcal{LD}$. This needs approximately

$$\Big(4(l_1 + K) + 8\frac{l_1 - K}{w+1}\Big)\mathrm{M} + \Big(4(l_1 + K) + 5\frac{l_1 - K}{w+1}\Big)\mathrm{S} + \Big(l_1 + K + \frac{l_1 - K}{w+1}\Big)\mathrm{M}_2.$$

If the precomputed points are in $\mathcal{LD}$ the most efficient way is to choose this coordinate system for all operations. In total this needs asymptotically

$$\Big(4(l_1 + K) + 13\frac{l_1 - K}{w+1}\Big)\mathrm{M} + \Big(4(l_1 + K) + 4\frac{l_1 - K}{w+1}\Big)\mathrm{S} + \Big(l_1 + K\Big)\mathrm{M}_2.$$

13.3.4 Montgomery scalar multiplication

López and Dahab [LóDa 1999] generalized Montgomery's idea, cf. Section 13.2.3, to binary curves. Let $P = (x_1, y_1)$ be a point on E. In projective coordinates, we write $P = (X_1 : Y_1 : Z_1)$ and let $[n]P = (X_n : Y_n : Z_n)$. The sum $[n+m]P = [n]P \oplus [m]P$ is given by the following formulas where Y_n does not occur.

Addition: $n \neq m$

$$\begin{aligned} Z_{m+n} &= (X_m Z_n)^2 + (X_n Z_m)^2, \\ X_{m+n} &= Z_{m+n} X_{m-n} + X_m Z_n X_n Z_m. \end{aligned}$$

Doubling: $n = m$

$$\begin{aligned} X_{2n} &= X_n^4 + a_6 Z_n^4 = \big(X_n^2 + \sqrt{a_6} Z_n^2\big)^2, \\ Z_{2n} &= X_n^2 Z_n^2. \end{aligned}$$

An addition takes 4M and 1S whereas a doubling needs only 2M and 3S, if $\sqrt{a_6}$ is precomputed. For the full scalar multiplication $[n]P$, we use *Montgomery's ladder*, cf. Algorithm 13.35, which requires $(6\mathrm{M} + 4\mathrm{S})(|n|_2 - 1)$ in total.

To recover the y-coordinate of $[n]P = (X_n : Y_n : Z_n)$ we first compute the affine x-coordinates of $[n]P$ and $[n+1]P$, that is $x_n = X_n/Z_n$ and $x_{n+1} = X_{n+1}/Z_{n+1}$ and then use the formula [LóDa 1999, OkSa 2001]

$$y_n = \frac{(x_n + x_1)\big((x_n + x_1)(x_{n+1} + x_1) + x_1^2 + y_1\big)}{x_1} + y_1. \tag{13.11}$$

Example 13.44 Let us compute $[763]P_7$ with Algorithm 13.35. The different steps of the computation are given in the following table where P stands for $P_7 \in E_7$, given by (13.10).

i	n_i	(P_1, P_2)	P_1	P_2
9	1	$(P, [2]P)$	$(\texttt{0x420} : ? : \texttt{0x1})$	$(\texttt{0x158} : ? : \texttt{0x605})$
8	0	$([2]P, [3]P)$	$(\texttt{0x158} : ? : \texttt{0x605})$	$(\texttt{0x7E9} : ? : \texttt{0x2FD})$
7	1	$([5]P, [6]P)$	$(\texttt{0x295} : ? : \texttt{0x56B})$	$(\texttt{0x620} : ? : \texttt{0x43B})$
6	1	$([11]P, [12]P)$	$(\texttt{0x5D0} : ? : \texttt{0x247})$	$(\texttt{0xA6} : ? : \texttt{0x6CE})$
5	1	$([23]P, [24]P)$	$(\texttt{0x755} : ? : \texttt{0x21B})$	$(\texttt{0x409} : ? : \texttt{0x93})$
4	1	$([47]P, [48]P)$	$(\texttt{0xBD} : ? : \texttt{0x25E})$	$(\texttt{0x26} : ? : \texttt{0x4BE})$
3	1	$([95]P, [96]P)$	$(\texttt{0x4EE} : ? : \texttt{0x51D})$	$(\texttt{0x4D6} : ? : \texttt{0x304})$
2	0	$([190]P, [191]P)$	$(\texttt{0x4C1} : ? : \texttt{0x58C})$	$(\texttt{0x553} : ? : \texttt{0x386})$
1	1	$([381]P, [382]P)$	$(\texttt{0x613} : ? : \texttt{0x7E4})$	$(\texttt{0x2BB} : ? : \texttt{0x60B})$
0	1	$([763]P, [764]P)$	$(\texttt{0x6C4} : ? : \texttt{0x105})$	$(\texttt{0x655} : ? : \texttt{0x485})$

To end the computation, we apply (13.11) to obtain that $[763]P_7 = (\texttt{0x84}, \texttt{0x475})$.

13.3.5 Point halving and applications

In this section we introduce a further map on the group of points of an elliptic curve.

Let $|E(\mathbb{F}_{2^d})| = 2^k \ell$, where ℓ is odd. If $k = 1$ then E is said to have *minimal* 2*-torsion* as curves of the form

$$E : y^2 + xy = x^3 + a_2 x^2 + a_6 \tag{13.12}$$

considered here always have one point of order 2, namely the point $T = (0, \sqrt{a_6})$. Hence, the doubling map [2] is not injective. Now assume E to have minimal 2-torsion and let G be a subgroup of odd order. If P belongs to G then there is a unique point $Q \in G$ such that $P = [2]Q$. Then denote $Q = \left[\frac{1}{2}\right]P$ and define the one-to-one *halving map* by

$$\begin{array}{rcl} \left[\frac{1}{2}\right] : G & \to & G \\ P & \mapsto & Q \text{ such that } [2]Q = P. \end{array}$$

In the following, we shall represent a point $P = (x_1, y_1)$ as (x_1, λ_1) where $\lambda_1 = x_1 + y_1/x_1$. In the context of a scalar multiplication based on halvings, this representation leads to a faster implementation as for repeated doublings.

In [Knu 1999] Knudsen develops an efficient technique to halve a point in affine coordinates lying on an elliptic curve with minimal 2-torsion. Independently, Schroeppel proposed the same method [Sch 2000c].

Note that half the curves of the form (13.12) defined over $\mathbb{F}_{2^d}$ have minimal 2-torsion, since this property is equivalent to $\mathrm{Tr}(a_2) = 1$.

Let $P = (x_1, \lambda_1) \in G$ and $Q = \left[\frac{1}{2}\right]P = (x_2, \lambda_2)$. Inverting doubling formulas, one has

$$\begin{array}{rcl} \lambda_2^2 + \lambda_2 & = & a_2 + x_1, \\ x_2^2 & = & x_1(\lambda_2 + 1) + y_1 = x_1(\lambda_2 + \lambda_1 + x_1 + 1), \\ y_2 & = & x_2(x_2 + \lambda_2). \end{array}$$

The algorithm is as follows. First find γ such that $\gamma^2 + \gamma = a_2 + x_1$. The other solution of the equation is then $\gamma + 1$, cf. Lemma 11.56. One corresponds to λ_2 and Q and the other one to $\lambda_2 + 1$ and $Q \oplus T$. If E has minimal 2-torsion it is possible to determine if γ is equal to λ_2 or not. Indeed only Q can be halved but not $Q \oplus T$. So (x_2, λ_2) is equal to $\left[\frac{1}{2}\right]P$ if and only if the equation $X^2 + X = a_2 + x_2$ has a solution in $\mathbb{F}_{2^d}$. This holds true if and only if $\operatorname{Tr}(a_2 + x_2) = 0$. Clearly $\operatorname{Tr}(a_2 + x_2) = \operatorname{Tr}(a_2^2 + x_2^2)$ and this remark saves a square root computation.

So one first obtains $w = x_1(\gamma + \lambda_1 + x_1 + 1)$, which is a candidate for x_2^2. If $\operatorname{Tr}(a_2^2 + w) = 0$ then $\lambda_2 = \gamma$ and $x_2 = \sqrt{w}$. Otherwise $\lambda_2 = \gamma + 1$ and $x_2 = \sqrt{w + x_1}$.

All these steps are summarized in the following algorithm.

Algorithm 13.45 Point halving

INPUT: The point $P = (x_1, y_1) \in G$ represented as (x_1, λ_1).
OUTPUT: The point $\left[\frac{1}{2}\right]P = (x_2, y_2)$ represented as (x_2, λ_2).

```
1. compute $\gamma$ such that $\gamma^2 + \gamma = a_2 + x_1$
2. $w \leftarrow x_1(\gamma + \lambda_1 + x_1 + 1)$
3. if $\operatorname{Tr}(a_2^2 + w) = 1$ then $\gamma \leftarrow \gamma + 1$ and $w \leftarrow w + x_1$
4. $\lambda_2 \leftarrow \gamma$ and $x_2 \leftarrow \sqrt{w}$
5. return $Q = (x_2, \lambda_2)$
```

Remarks 13.46

(i) To determine (x_2, λ_2) Algorithm 13.45 requires us to compute the solution of a quadratic equation, one square root, one multiplication, and one absolute trace. A further multiplication is necessary to obtain y_2.

(ii) See Section 11.2.6 for a description of algorithms to compute γ and $\sqrt{w}$.

(iii) The computation of the trace in Line 3 is straightforward, cf. Remarks 11.57.

(iv) Algorithm 13.45 can be easily generalized when $E(\mathbb{F}_{2^d})$ has a subgroup isomorphic to $\mathbb{Z}/2^k\mathbb{Z}$ with $k > 1$ [KNU 1999]. Nevertheless, it is necessary, in this case, to solve k equations, perform $k + 1$ multiplications, one test, and k or $k + 1$ square root computations to find (x_2, y_2), so that in practice the technique is usually not interesting for $k > 1$.

Example 13.47 The point $P_7 = (\mathtt{0x420}, \mathtt{0x5B3})$ on E_7 is a point of order 2026. This implies that $R_7 = [2]P_7 = (\mathtt{0x14D}, \mathtt{0x4CB})$ is a point of odd order $\ell = 1013$. Thus in the group $G = \langle R_7 \rangle$, the halving map is well defined. Let us compute $\left[\frac{1}{2}\right]R_7$ with Algorithm 13.45. First, we have $\lambda_1 = \mathtt{0x67}$. We deduce that $\gamma = \mathtt{0x1C}$ and $w = \mathtt{0x605}$. Since the trace of $a_2^2 + w$ is equal to one, the values of γ and w are changed to $\mathtt{0x1D}$ and $\mathtt{0x748}$. Finally, $\lambda_2 = \mathtt{0x1D}$ and $x_2 = \mathtt{0x3B8}$. It follows that the unique point $S_7 \in G$ such that $[2]S_7 = R_7$ is $(\mathtt{0x3B8}, \mathtt{0x441})$.

We also have $P_7 = S_7 \oplus T_7$ where $T_7 = (\mathtt{0x0}, \mathtt{0x19A})$ is the 2-torsion point of E_7.

Now, let us explain how to compute the scalar multiplication $[n]P$ of a point P of odd order $\ell_1 \mid \ell$. Let $m = \lceil \lg \ell_1 \rceil$. Then if

$$2^{m-1}n = \sum_{i=0}^{m-1} \widehat{n}_i 2^i \bmod \ell_1, \quad \text{with } n_i \in \{0, 1\}$$

one has

$$n \equiv \sum_{i=0}^{m-1} \frac{\widehat{n}_{m-1-i}}{2^i} \pmod{\ell_1}$$

and $[n]P$ can be obtained by the following algorithm. We additionally put $\left[\frac{1}{2}\right]P_\infty = P_\infty$.

Algorithm 13.48 Halve and add scalar multiplication

INPUT: A point $P \in E(\mathbb{F}_{2^d})$ of odd order ℓ_1 and a positive integer n.
OUTPUT: The point $[n]P$.

```
$m \leftarrow \lceil \lg \ell_1 \rceil$
$\widehat{n} \leftarrow (2^{m-1}n) \bmod \ell_1$                    $[\widehat{n} = (\widehat{n}_{m-1} \ldots \widehat{n}_0)_2]$
$Q \leftarrow P_\infty$
for $i = 0$ to $m - 1$ do
    $Q \leftarrow \left[\frac{1}{2}\right]Q$
    if $\widehat{n}_i = 1$ then $Q \leftarrow Q \oplus P$
return $Q$
```

Remarks 13.49

(i) All the window and recoding techniques seen in Chapter 9 apply as well. In particular, if $\sum_{i=0}^{m} \widehat{n}_i 2^i$ is the NAF$_w$ representation of $2^m n$ modulo ℓ_1, then

$$n \equiv \sum_{i=0}^{m} \frac{\widehat{n}_{m-i}}{2^i} \pmod{\ell_1}.$$

(ii) No method is currently known to halve a point in projective coordinates. In [HaMe$^+$ 2003] two halve-and-add algorithms for the NAF$_w$ representation are given. The one operating from the right to the left halves the input P rather than the accumulators, which can therefore be represented in projective coordinates. In this case mixed addition formulas can be used for a better efficiency.

(iii) Point halving can be used to achieve faster scalar multiplication on Koblitz curves; see Chapter 15 and [AvCi$^+$ 2004].

Example 13.50 Let us compute $[763]R_7$ with Algorithm 13.48. As R_7 is of odd order 1013, we have $m = 10$ and $2^9 \times 763 \equiv 651 \pmod{1013}$. Now $651 = (1010001011)_2$ which implies that

$$763 \equiv \frac{1}{2^9} + \frac{1}{2^8} + \frac{1}{2^6} + \frac{1}{2^2} + 1 \pmod{1013}.$$

Thus, the main steps of the computation, expressed in the form (x, λ), are

$$\begin{aligned}
\left[\tfrac{1}{2}\right]R_7 \oplus R_7 &= (\texttt{0x1}, \texttt{0x21D}),\\
\left[\tfrac{1}{2}\right]^3 R_7 \oplus \left[\tfrac{1}{2}\right]^2 R_7 \oplus R_7 &= (\texttt{0x644}, \texttt{0x184}),\\
\left[\tfrac{1}{2}\right]^7 R_7 \oplus \left[\tfrac{1}{2}\right]^6 R_7 \oplus \left[\tfrac{1}{2}\right]^4 R_7 \oplus R_7 &= (\texttt{0x77C}, \texttt{0x3EC}),\\
\left[\tfrac{1}{2}\right]^9 R_7 \oplus \left[\tfrac{1}{2}\right]^8 R_7 \oplus \left[\tfrac{1}{2}\right]^6 R_7 \oplus \left[\tfrac{1}{2}\right]^2 R_7 \oplus R_7 &= (\texttt{0x2EA}, \texttt{0x281}).
\end{aligned}$$

We deduce that $[763]R_7 = (\mathtt{0x2EA}, \mathtt{0x7C8})$ in affine coordinates.

It is also easy to obtain the multiple of a point that does not belong to G. For instance, let us compute $[763]P_7$. We have $P_7 = \left[\frac{1}{2}\right]R_7 \oplus T_7$ and since the maps $\left[\frac{1}{2}\right]$ and $[n]$ commute, it follows that

$$\begin{aligned} [763]P_7 &= \left[\tfrac{1}{2}\right][763]R_7 \oplus T_7, \\ &= (\mathtt{0x5CF}, \mathtt{0x485}) \oplus (\mathtt{0x0}, \mathtt{0x19A}), \\ &= (\mathtt{0x84}, \mathtt{0x475}). \end{aligned}$$

13.3.6 Parallel implementation

Also, for fields of even characteristic, parallel implementations have gained some interest, one of the first works being [KoTs 1993]. However, applications using affine coordinates usually try to achieve parallelism on the lower level of field arithmetic.

In the chapter on side-channel attacks, cf. Chapter 29, we discuss several parallel implementations as this is mainly of interest for small devices like smart cards. There, the additional restriction is that the implementation should be secured against some particular attacks. Common choices are Montgomery coordinates distributed on two processors. We refer the reader to that chapter for details and mention here only the work of Mishra [Mis 2004a], who derives a pipelined computation such that in a scalar multiplication the average number of clock cycles needed per group operation is only 6 when using two processors.

13.3.7 Compression of points

Let $P = (x_1, y_1)$ be a point on $E/\mathbb{F}_{2^d} : y^2 + xy = f(x)$. As for odd characteristic, we show how to represent P by $\big(x_1, b(y_1)\big)$, where $b(y_1)$ is a bit distinguishing P from $-P = (x_1, y_1 + x_1)$.

There exists exactly one Weierstraß point having $x_1 = 0$. For the other points we follow the steps in the next paragraphs.

Decompression

In even characteristic it is easier to explain decompression first. Thus, assume that P is given by $\big(x_1, b(y_1)\big)$, and $b(y_1) \in \{0, 1\}$. As x_1 is the x-coordinate of a point, the quadratic equation $y^2 + x_1 y + x_1^3 + a_2 x_1^2 + a_6$ has two solutions. It is clear that such a solution exists if $Y^2 + Y + (x_1^3 + a_2 x_1^2 + a_6)/x_1^2$ has a solution, i.e., if $\mathrm{Tr}\big((x_1^3 + a_2 x_1^2 + a_6)/x_1^2\big) = 0$, cf. Section 11.2.6. If y_1' is one solution then $y_1' + 1$ is the other. Hence, for the roots y_1' the least significant bit allows us to distinguish between the solutions and we need to resort to the equation in Y to compute the roots. To find the solutions of the original equation we put $y_1 = y_1' x_1$ for the y_1' determined by $b(y_1)$.

Compression

We have just seen that the least significant bit of $y_1' = y_1/x_1$ should be used as $b(y_1)$. Unfortunately, this requires one inversion, hence, some work is also needed to compress a point. This is in contrast to the case of odd characteristic.