Efficient Scalar Multiplication by Isogeny Decompositions

Christophe Doche[1], Thomas Icart[2], and David R. Kohel[3]

[1] Department of Computing
Macquarie University, Australia
doche@ics.mq.edu.au

[2] Laboratoire d'Informatique de l'École Polytechnique, France

[3] School of Mathematics and Statistics
University of Sydney, Australia
kohel@maths.usyd.edu.au

November 23, 2005

Abstract

On an elliptic curve, the degree of an isogeny corresponds essentially to the degrees of the polynomial expressions involved in its application. The multiplication–by–ℓ map $[\ell]$ has degree ℓ^2, therefore the complexity to directly evaluate $[\ell](P)$ is $O(\ell^2)$. For a small prime $\ell\,(= 2, 3)$ such that the additive binary representation provides no better performance, this represents the true cost of application of scalar multiplication. If an elliptic curves admits an isogeny φ of degree ℓ then the costs of computing $\varphi(P)$ should in contrast be $O(\ell)$ field operations. Since we then have a product expression $[\ell] = \hat{\varphi}\varphi$, the existence of an ℓ-isogeny φ on an elliptic curve yields a theoretical improvement from $O(\ell^2)$ to $O(\ell)$ field operations for the evaluation of $[\ell](P)$ by naïve application of the defining polynomials. In this work we investigate actual improvements for small ℓ of this asymptotic complexity. For this purpose, we describe the general construction of families of curves with a suitable decomposition $[\ell] = \hat{\varphi}\varphi$, and provide explicit examples of such a family of curves with simple decomposition for [3]. Finally we derive a new tripling algorithm to find complexity improvements to triplication on a curve in certain projective coordinate systems, then combine this new operation to non-adjacent forms for ℓ-adic expansions in order to obtain an improved strategy for scalar multiplication on elliptic curves.

Keywords. Elliptic curve cryptography, fast arithmetic, efficiently computable isogenies, efficient tripling, ℓ-adic NAF_w.

1 Introduction

Given an elliptic curve E/K, together with a point $P \in E(K)$ and an integer k, the efficient computation of the scalar multiple $[k]P$ is central in elliptic curve cryptography. Many ways to speed up this computation have been actively researched. For instance, one can cite

- the use of alternative representations for the scalar multiple k (non-adjacent forms [MO90, CMO97, TYW04], ternary/binary approach [CJLM05], Dual Base Number System [DJM99]).
- the improvement of existing operations by use of other systems of coordinates (projective , weighted projective [CMO98]) and the introduction of new basic operations like $[2]P \pm Q$, $[3]P$, $[3]P \pm Q$, $[4]P$, $[4P] \pm Q$, *cf.* [CJLM05, DIM05].
- the use of endomorphisms (first on a singular curve that appeared to be insecure [MV90], later with Koblitz curves [Kob92, Sol00, Lan05] and GLV curves [GLV01, CLSQ03]).

See [ACD$^+$05, chaps. 9, 13, and 15] and [HMV03] for a more comprehensive description of all the techniques involved.

The purpose of this article is to investigate new and more efficient ways to compute the multiplication–by–ℓ map. Indeed, given an integer $\ell \geqslant 2$, it is possible in some cases and for well chosen families of curves to split the map $[\ell]$ as the product of two isogenies. A direct computation of $[\ell]P$ involves the evaluation of rational polynomials of degree ℓ^2. The interest of this approach is that the isogenies φ and $\hat{\varphi}$ such that $[\ell] = \hat{\varphi}\varphi$ will be both of degree ℓ. Therefore it should be possible to obtain more efficient formulas to compute $[\ell]$ this way. We investigate this idea for small values of ℓ, especially 2 and 3 and obtain a more efficient tripling leading to a very fast scalar multiplication algorithm.

2 Splitting Multiplication by ℓ

In this section we describe the definitions and background results for existence and construction of an ℓ-isogeny φ such that $[\ell] = \hat{\varphi}\varphi$.

2.1 Subgroup (schemes) defined over K.

Let E be an elliptic curve over K, with defining equation

$$F(x, y) = y^2 + (a_1x + a_3)y - (x^3 + a_2x^2 + a_4x + a_6) = 0.$$

We give an elementary background on concepts and conditions for torsion subgroups to be defined over the base field K.

Definition 2.1 *Let N be an integer greater than 1 and let $E[N]$ be the group of N-torsion points in $\overline{K}$. A torsion subgroup G of $E[N]$ is said to be defined* *over K or to be K-rational if $G\backslash\{O\}$ is the zero set of a finite set of polynomials $\{f_1(x,y),\dots,f_n(x,y)\}$ in $K[x,y]/\big(F(x,y)\big)$.*

A torsion subgroup can be specified by two polynomials, one of which is the polynomial $\psi_G(x)$ whose roots are the x-coordinates of the points $P=(x,y)$ in G. If N is odd, then this polynomial suffices to define the torsion subgroup. If N is even, then the full ideal of polynomials which have zeros on G cannot be specified as a single polynomial in x. As an example, if $G=\{O,(x_0,y_0)\}$, where (x_0,y_0) is a 2-torsion point, then G is determined as the zero set of the polynomial $x-x_0$, but both $y-y_0$ and $2y+a_1x+a_3$ are zero on $\{(x_0,y_0)\}$, but are not in the ideal $(x-x_0)$.

From the odd case, the condition for a subgroup to be K-rational is not that the points have coefficients in K, but that the symmetric functions in these coefficients must lie in K. Since every finite subgroup G of $E\big(\overline{K}\big)$ is the kernel of an isogeny $\varphi_G : E\to E'$, the question of whether the subgroup can be defined over K, is related to the K-rationality of the isogeny φ_G. The following classical theorem states that these concepts are equivalent.

Theorem 2.1 *A finite subgroup G of E is K-rational if and only if G is the kernel of an isogeny $\psi : E\to E'$ defined over K.*

Since the subgroup $E[N]$ of $E\big(\overline{K}\big)$ is the kernel of the scalar multiplication $[N]$, which is defined over K, we obtain:

Corollary 2.1 *Every torsion subgroup $E[N]$ is K-rational.*

The defining polynomials for the N-torsion subgroups are the *division polynomials $\psi_N(x,y)$*, which are computable by explicit recursive formulas.

Corollary 2.2 *Let G and H be two finite K-rational subgroups of E. Then $G\cap H$ and $G+H$ are K-rational subgroups of E.*

Proof 2.1 *The intersection property holds immediately since if G and H are the zero sets of $S=\{g_1,\dots,g_r\}$, and $T=\{h_1,\dots,h_s\}$, respectively, then $G\cap H$ is the zero set of $S\cup T$. To prove that $G+H$ is K-rational we apply the theorem to the isogeny $\varphi_{H'}\circ\varphi_G$ where $H'=\varphi_G(H)$.*

Combining the previous two corollaries we obtain:

Corollary 2.3 *Suppose that E admits an isogeny $E\to E'$ with cyclic kernel of order N. Then $E[\ell]$ contains a rational subgroup of order ℓ for every ℓ dividing N.*

These corollaries permit us to find a product decomposition for any isogeny, or its defining kernel subgroup, into scalar multiplications $[\ell]$ (determined by $E[\ell]$) and isogenies of prime degree (given by a rational subgroup G of order ℓ), for primes ℓ dividing the degree of the isogeny. Since efficient algorithms for scalar multiplication $[\ell]$ by small primes have been well-investigated, in the next section we focus on prime order isogenies ℓ which "split" the isogeny $[\ell]$ into a product of isogenies φ and $\hat{\varphi}$.

2.2 Parameterizations of cyclic ℓ-torsion subgroups

The theory of modular curves gives a means of achieving explicit parameterizations of families of elliptic curves with the structure of an isogeny of degree ℓ. We describe the general background to this construction to motivate the examples.

It is well-known that the j-invariant of an elliptic curve E over any field K determines the isomorphism class of that curve over $\overline{K}$. Conversely, any value $j \neq 0, 12^3$ is the j-invariant of an elliptic curve

$$E_j : y^2 + xy = x^3 - \frac{36}{j - 12^3}x - \frac{1}{j - 12^3}.$$

The j-invariant can be identified with a generator of the function field $K\big(X(1)\big)$ of the modular curve $X(1)$, classifying elliptic curves up to isomorphism. We view the above equation E_j as a family of elliptic curves over the "j-line" $X(1)\backslash\{0, 1, \infty\} \cong \mathbb{A}^1\backslash\{0, 1\}$.

In order to determine similar models for elliptic curves which admits an ℓ-isogeny, or equivalently a K-rational cyclic subgroup G of $E[\ell]$, we use the modular curves $X_0(\ell)$ covering $X(1)$.

For the values $\ell = 2, 3, 5, 7$, and 13 the curve $X_0(\ell)$ has genus 0, which means that there exists a modular function u on $X_0(\ell)$ such that $K\big(X_0(\ell)\big) = K(u)$. The covering $X_0(\ell) \to X(1)$ is determined by an inclusion of function fields $K\big(X(1)\big) \to K\big(X_0(\ell)\big)$, which means that we can express j as a rational function in u.

For the above values of ℓ, we may use quotients of the Dedekind η function on the upper half plane

$$u(q) = \left(\frac{\eta(\tau)}{\eta(\ell\tau)}\right)^r = q^{-1}\prod_{n=1}^{\infty}\left(\frac{1 - q^n}{1 - q^{n\ell}}\right)^r$$

where $r = 24/\gcd(12, \ell - 1)$ and $q = \exp(2\pi i\tau)$, to find a relation with the q-expansion $j(q)$ for the j-function to solve for the expression for the j-function. Substituting into the above equations we then twist the curve or make a change of variables to simplify the resulting equation to obtain the models for which the ℓ-torsion contains a parameterized rational subgroup of order ℓ (over $K(u)$ or over K for any particular value of u in K). The models used in the isogeny decompositions which follow may be derived by this technique, with the kernel polynomial determined by factorization of the ℓ-division polynomial of this curve.

2.3 Parameterized models

Applying these ideas, we have build families of curves for which [2] or [3] splits into 2 isogenies of degree respectively 2 and 3. For instance, an elliptic curve defined over a field of characteristic different from 2 and 3 with a rational 3-torsion subgroup can be expressed in the form (up to twists):

$$E : y^2 = x^3 + 3u(x + 1)^2$$

with the 3-torsion subgroup defined by $x = 0$; we note that the curve E does not necessarily have a point of order 3. The image curve is defined by an equation:

$$E_t : y^2 = x^3 - u(3x - 4u + 9)^2.$$

Note that the same thing holds in characteristic 2. In fact, an elliptic curve with a rational 3-torsion subgroup can be expressed in the form (up to twists):

$$E : y^2 + (x + u)y = x^3.$$

It has a rational 3-torsion subgroup defined by $x = 0$. The image curve is defined by an equation:

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

Explicit formulas of the curves and isogenies to split [2] in characteristic greater than 2 and to split [3] in characteristic greater than 3 can be found in Section 3.

2.4 On special versus generic elliptic curves

Since we propose curves of a particular form, it is relevant to make a distinction between curves of a special form and generic curves.

A family of elliptic curves is a parameterized equation of different elliptic curves $E/K(u_1, \ldots, u_t)$ in indeterminates $u_1, \ldots, u_t$. We say that a family of elliptic curves is *geometrically special* if, for $(u_1, \ldots, u_t) \in \overline{K}^n$, there exists a finite set of j-invariants of curves in the family. Otherwise, we say that the family is *geometrically general.* Standard examples of families are the family of elliptic curves $y^2 = x^3 = ax + b$, over $K(a, b)$ which is geometrically general, or the family of Koblitz curves $y^2 + xy = x^3 + ax^2 + 1$ over $\mathbb{F}_2(a)$ which are geometrically special.

Any family of curves obtained by the CM construction are geometrically special because there exists only a finite set of j-invariants for each fixed discriminant D. Even if D is allowed to vary, in practice there are only a finite set of candidates D with $|D|$ bounded by the time to compute a class polynomial for D. Similarly, any family of supersingular elliptic curves is geometrically special, since there are only finitely many j-invariants of supersingular elliptic curves.

The curves that we introduce lie in geometrically general families because their invariants give infinitely many j-invariants $j = j(u)$, and conversely, every j-invariant arises as $j(u)$ for some u in $\overline{K}$.

We say that a family is *arithmetically special* if the properties of the curves in the family are in some way special with respect to a random curve over K. This is more imprecise, but to make it more precise one should speak of an arithmetic invariant, like group order or discriminant of the endomorphism ring which can distinguish curves in the family and those outside of it. Every special construction will be arithmetically special. For instance, Jao et al. [JMV05] observe that curves produced by CM construction are arithmetically special and distinguished by properties of the discriminant of their endomorphism rings. By construction we build curves that are arithmetically special, since they all have a cyclic ℓ-isogeny. In contrast, a curve over a finite field has a 50% chance of such a rational ℓ-isogeny, and a curve with such a rational isogeny over a number field is exceptional. Supersingular elliptic curves are arithmetically special with respect to existence of rational isogenies: over a finite degree extension L/K, all $\ell+1$ cyclic ℓ-isogenies for all ℓ become simultaneously L-rational.

Despite the fact that our families have arithmetically special ℓ-torsion, by virtue of the criterion by which they are constructed, for any prime $n \neq \ell$, the n-torsion and n-isogenies follow the general behavior, and we have no reason to expect any special properties of the group orders $|E(K)|$ for curves in our families, apart from the potential factors of ℓ which arise.

3 Efficiently Applicable Isogenies

Let us investigate at present how the multiplications by [2] and [3] can be efficiently split as a product of 2 isogenies in practice.

3.1 Elliptic curves with degree 2 isogenies

An elliptic curve defined over a field $\mathbb{F}_q$ of characteristic $\neq 2$ with a rational 2-torsion subgroup can be expressed in the form (up to twists):

$$E : y^2 = x^3 + ux^2 + 16ux$$

with a 2-torsion point $(0,0)$. The corresponding isogeny of degree 2 is:

$$(x_1, y_1) \mapsto (x_t, y_t) = \left(x_1 + u\left(1 + \frac{16}{x_1}\right),\ y_1\left(1 - \frac{16u}{x_1^2}\right)\right),$$

to an image curve defined by an equation:

$$E_t : y^2 = x^3 - 2ux^2 + u(u-64)x.$$

The isogeny dual to the first isogeny is given by

$$(x_t, y_t) \mapsto (x_2, y_2) = \left(\frac{1}{2^2}\left(x_t - 2u + \frac{u(u-64)}{x_t}\right),\ \frac{1}{2^3}y_t\left(1 - \frac{u(u-64)}{x_t^2}\right)\right).$$

The compositum of these maps gives the multiplication–by–2 map on E.

To take advantage of this splitting, let us introduce a new system of coordinates. Since they are similar to López-Dahab coordinates ($\mathcal{LD}$) introduced in characteristic 2, *cf.* [LD98], let us call them *modified López-Dahab coordinates* ($\mathcal{LD}^m$). A point (x_1, y_1) in affine coordinates ($\mathcal{A}$) on the elliptic curve E will be represented by (X_1, Y_1, Z_1, Z_1^2) where $x_1 = X_1/Z_1$ and $y_1 = Y_1/Z_1^2$. It is a simple exercise to check that (X_2, Y_2, Z_2, Z_2^2) corresponding to $(x_2, y_2) = [2](x_1, y_1)$

is given by

$$
\begin{array}{rlll}
A = X_1^2, & B = X_1^2 - 16uZ_1^2, & Y_t = Y_1 \times B, \\
X_2 = B^2, & Z_2 = 4Y_1^2, & C = X_1^2 \times uZ_1^2, \\
D = Z_2^2, & E = u(Z_2 - 4C), & Y_2 = Y_t\bigl(2X_2 + E + 256C\bigr).
\end{array}
$$

The number of elementary operations needed to obtain (X_2, Y_2, Z_2, Z_2^2) is thus $5\mathrm{M}+4\mathrm{S}$, where M and S respectively denotes a multiplication and a squaring in the field $\mathbb{F}_q$. However, if u is chosen so that a multiplication by u is negligible, the costs for a doubling drop to $3\mathrm{M}+4\mathrm{S}$. Note that it is sufficient to choose u to fit in a word, or to have a low Hamming weight representation in order to achieve this property. Clearly, the number of suitable values of u for a given p is extremely large and therefore this assumption has a limited impact on the rest of the system.

Note also that the fastest system of coordinates for doubling corresponds to modified Jacobian coordinates $\mathcal{J}^m$ (see for instance [CMO98]) where a point (x_1, y_1) is represented by (X_1, Y_1, Z_1, aZ_1^4) with $x_1 = X_1/Z_1^2$ and $y_1 = Y_1/Y_1^3$. Indeed, to perform a double on the curve $y^2 = x^3+ax+b$, one needs only 4M+4S. It is to be noted that choosing a special value for a does not change the overall complexity. The addition $\mathcal{J}^m + \mathcal{J}^m = \mathcal{J}^m$ needs $13\mathrm{M}+6\mathrm{S}$ whereas the mixed addition $\mathcal{J}^m + \mathcal{A} = \mathcal{J}^m$ only $9\mathrm{M}+5\mathrm{S}$. Again this complexity is independent of the value of the parameters so that no advantage can be obtained from a special choice of a curve in modified Jacobian coordinates.

Now, let us give addition formulas for $\mathcal{LD}^m$. We will only address the mixed coordinates case, since it is the most important in practice. So let $(X_1, Y_1, 1)$ in $\mathcal{A}$ and (X_2, Y_2, Z_2, Z_2^2) in $\mathcal{J}^m$ be two points on E. Again it is a simple exercise to check that (X_3, Y_3, Z_3, Z_3^2) is given that:

$$
\begin{array}{lll}
A = Y_1 \times Z_2^2 - Y_2, & B = X_1 \times Z_2 - X_2, & C = B \times Z_2, \\
Z_3 = C^2, & D = X_1 \times Z_3, & E = A^2, \\
F = X_2 \times B \times C, & X_3 = E - uZ_3 - D - F, & G = Z_3^2, \\
H = A \times C, & Y_3 = H \times (D - X_3) - Y_1 \times G. &
\end{array}
$$

These computations require $9\mathrm{M}+3\mathrm{S}$ if a multiplication by u is negligible. So, choosing a special value for u provides an improvement and makes modified López–Dahab coordinates faster than modified Jacobian coordinates. At present let us generalize the concept to the multiplication–by–[3] map.

3.2 Elliptic curves with degree 3 isogenies

As mentioned earlier, an elliptic curve defined over a field of characteristic different from 2 and 3 with a rational 3-torsion subgroup can be expressed in the form (up to twists):

$$E : y^2 = x^3 + 3u(x+1)^2$$

with the 3-torsion subgroup defined by $x = 0$; we note that the curve E does not necessarily have a point of order 3. The corresponding isogeny of degree 3 is:

$$(x_1, y_1) \mapsto (x_t, y_t) = \left(x_1 + 4u + 12u\frac{x_1 + 1}{x_1^2},\ y_1 \left(1 - 12u\frac{x_1 + 2}{x_1^3} \right) \right).$$

The image curve is defined by an equation:

$$E_t : y^2 = x^3 - u(3x - 4u + 9)^2$$

which subsequently has a 3-torsion subgroup defined by $x = 0$, defining the kernel of the dual isogeny. This isogeny takes form

$$\begin{aligned}(x_t, y_t) \mapsto (x_3, y_3) = \ & \left(\frac{1}{3^2} \left(x_t - 12u + \frac{12u(4u-9)}{x_t} - \frac{4u(4u-9)^2}{x_t^2} \right), \right. \\ & \left. \frac{1}{3^3} y_t \left(1 - \frac{12u(4u-9)}{x_t^2} + \frac{8u(4u-9)^2}{x_t^3} \right) \right).\end{aligned}$$

The compositum of these maps gives the multiplication–by–3 map on E.

Again, to take advantage of this splitting, we will use weighted projective coordinates. More precisely let us represent the affine point $P_1 = (x_1, y_1)$ by (X_1, Y_1, Z_1, Z_1^2) where $x_1 = X_1/Z_1^2$ and $y_1 = Y_1/Z_1^3$. These coordinates are called *new Jacobian* and are denoted by $\mathcal{J}^n$. We will also describe doublings and mixed additions for this system. The term Z_1^2 will contribute to make the mixed addition more efficient. First let us give the formulas to compute $[3]P_1 = (X_3, Y_3, Z_3, Z_3^2)$:

$$\begin{array}{lll} A = (X_1 + 3Z_1^2)^2, & B = uZ_1^2 \times A, & X_t = Y_1^2 + B, \\ Y_t = Y_1 \times (Y_1^2 - 3B), & Z_t = X_1 \times Z_1, & C = Z_t^2, \\ D = \big((4u-9)C - X_t\big)^2, & E = -3uC \times D, & X_3 = (Y_t^2 + E), \\ Y_3 = Y_t(X_3 - 4E), & Z_3 = 3X_t \times Z_t, & Z_3^2. \end{array}$$

It is easy to see that $6\mathrm{M} + 6\mathrm{S}$ are needed to obtain $[3]P_1$ in $\mathcal{J}^n$ when u is suitably chosen so that a multiplication by u is negligible. Otherwise, $8\mathrm{M} + 6\mathrm{S}$ are necessary.

Now let us see how a doubling can be efficiently obtained in that system. In fact, it is sufficient to slightly modify the formulas existing for Jacobian coordinates. We have:

$$\begin{array}{lll} A = Y_1 \times Z_1, & Z_2 = 2A, & B = 4Y_1^2 \times X_1, \\ C = B + 6uA^2, & Z_2^2 = 4A^2, & D = 3X_1^2, \\ E = D + 6uZ_1^2 \times (Z_1^2 + X_1), & X_2 = -2B + E^2, & Y_2 = -8Y_1^4 + E \times (B - X_2). \end{array}$$

Thus a doubling in $\mathcal{J}^n$ requires $4\mathrm{M} + 5\mathrm{S}$ as long as we neglect multiplications by u, otherwise a doubling can be obtained with $6\mathrm{M} + 4\mathrm{S}$.

Finally, let us detail the addition of an affine point $(X_1, Y_1, 1)$ and a point (X_2, Y_2, Z_2, Z_2^2) in $\mathcal{J}^n$. Again, they slightly differ from the ones for the addition in Jacobian coordinates, see [ACD+05].

$$
\begin{array}{lll}
A = X_1 \times Z_2^2, & B = Y_1 \times Z_2^2 \times Z_2, & C = X_2 - A, \\
D = Y_2 - B, & Z_3 = Z_2 \times C, & E = Z_3^2, \\
F = C^2, & G = C \times F, & H = A \times F, \\
X_3 = -G - 3uE - 2H + D^2, & Y_3 = -B \times G + D \times (H - X_3). &
\end{array}
$$

In total, one needs 8M+3S to compute an addition. If u is a random element in the field, then an extra multiplication is required. Note that the extra element Z_2^2 in $\mathcal{J}^n$ allows to save one squaring in the addition above.

3.2.1 Comparison with other systems.

Direct tripling formulas have been introduced by Ciet et al. [CJLM05]. The general idea is to avoid computing intermediate values for the doubling. This allows to get rid of one inversion at the cost of more multiplications. Recently, Dimitrov et al. succeeded in totally avoid using inversions [DIM05]. Usually, no special value for the parameters of the curve is considered, probably because this has a limited impact anyway on the complexity of the operations. In our case, important savings can be made if the parameter u of the curve is specially chosen, as suggested by the next table comparing the complexities of different operations in different coordinate systems. Note that we only require that a multiplication by u is trivial so that a very large scope of values are still available, like a small u or more generally u with a low Hamming weight expansion.

System	This work	[DIM05]	[CJLM05]
Equation	$y^2 = x^3 + 3u(x+1)^2$	$y^2 = x^3 + ax + b$	$y^2 = x^3 + ax + b$
Coordinates	New Jacobian $\mathcal{J}^n$	Jacobian $\mathcal{J}$	Affine $\mathcal{A}$
Tripling	8M + 6S	10M + 6S	I + 7M + 4S
special u or a	6M + 6S	9M + 6S	—
Doubling	6M + 4S	4M + 6S	I + 2M + 2S
special u or a	4M + 5S	4M + 5S	—
$a = -3$	NA	4M + 4S	—
Mixed Addition	9M + 3S	8M + 3S	I + 2M + S
special u or a	8M + 3S	—	—

Note also that there exist formulas to directly compute $[2]P \pm Q$ and $[3]P \pm Q$ with respectively I + 9M + 2S and 2I + 9M + 3S; see [CJLM05] for details.

Since we have a very efficient tripling algorithm, it is natural to consider the expansion of k in base 3 leading to a "triple and add algorithm" as well as other generalizations, like expansions in non-adjacent form. We discuss this at present.

4 Non-adjacent forms for ℓ-adic expansions

Given two integers k and $\ell \geqslant 2$, it is well-known that k can be expressed in a unique way in base ℓ. For computer applications, ℓ is usually chosen to be 2 or a power of 2. In the context of multiplication and of exponentiation/scalar multiplication other representations have been considered, for instance the binary non-adjacent form and width-w non-adjacent form, respectively denoted by NAF and NAF_w, see [ACD$^+$05].

Recently, Takagi et al. [TYW04] have generalized the concept of width-w non-adjacent form to any radix ℓ and introduced an ℓ-NAF_w.

Definition 4.1 *Let ℓ and w be two integers greater than* 1*. Then every positive integer k has a signed-digit expansion*

$$k = \sum_{i=0}^{m} k_i \ell^i$$

where

- *each k_i is zero or coprime with ℓ,*
- $|k_i| < \ell^w/2$,
- *among any w consecutive coefficients at most one is nonzero.*

An expansion of this particular form is called width-w non-adjacent form in basis ℓ, *ℓ-NAF_w for short, and is denoted by $(k_m \dots k_0)_{\ell\text{-}NAF_w}$.*

It is trivial to derive an algorithm to compute the ℓ-NAF_w generalizing the one existing for the NAF_w.

Algorithm 1. ℓ-NAF_w representation

INPUT: A positive integer k, a radix $\ell \geqslant 2$ and a parameter $w > 1$.
OUTPUT: The ℓ-NAF_w representation $(k_m \dots k_0)_{\ell\text{-}\mathrm{NAF}_w}$ of k.

```
1.  i ← 0
2.  while k > 0 do
3.      if k ≢ 0 (mod ℓ) then
4.          k_i ← k mod ℓ^w
5.          if k_i > ℓ^w/2 then k_i ← k_i − ℓ^w
```

6. $k \leftarrow k - k_i$

7. **else** $k_i \leftarrow 0$

8. $k \leftarrow k/\ell$ and $i \leftarrow i + 1$

9. **return** $(k_m \ldots k_0)_{\ell\text{-NAF}_w}$

Remarks.

- The classical NAF corresponds to the choice $\ell = w = 2$.
- It can be shown that this expansion is unique and that it has the smallest Hamming weight among all signed representations for k having digits k_i's such that $|k_i| < \ell^w/2$, see [TYW04].

It is well-known that the density of the classical NAF_w is $1/(w+1)$. This result can be generalized to ℓ-NAF_w, as shown in [TYW04]. See also [HT05] for further results.

Proposition 4.1 *The average density of the ℓ-NAF$_w$ is equal to* $\dfrac{\ell - 1}{(\ell - 1)w + 1}\cdot$

Proof 4.1 *For that matter, we compute the average length $E(\ell, w)$ of running 0's between two nonzero coefficients. From the definition, it is clear that there are at least $w - 1$ consecutive zeroes between two nonzero coefficients in the ℓ-NAF$_w$ expansion.*

Assuming that $k \not\equiv 0 \pmod{\ell}$ then $k_i \neq 0$ and $k \leftarrow k - k_i$ is now a multiple of ℓ^w. Let $t = k/\ell^w$. There are different possibilities for the integer t which can take any value. If t is not a multiple of ℓ, there will be exactly $w - 1$ consecutive zeroes until the next nonzero coefficient is found. Now the probability that t is not a multiple of ℓ is $(\ell - 1)/\ell$. In the same way, there will be exactly $w - 2 + i$ consecutive zeroes until the next nonzero coefficient is found if and only if t is a multiple of ℓ^{i-1} but not a multiple of ℓ^i. This event occurs with a probability equal to $(\ell - 1)/\ell^i$, namely $\ell - 1$ choices $(\ell^{i-1}, 2\ell^{i-1}, \ldots, (\ell - 1)\ell^{i-1})$ out of ℓ^i possible residues. This implies that the average length of running zeroes is

$$E(\ell, w) = w - 2 + \sum_{i \geq 1} i(\ell - 1)/\ell^i$$

and a simple computation gives $E(\ell, w) = w - 2 + \ell/(\ell - 1)$. Since the average density of the ℓ-NAF$_w$ is $1/\big(E(\ell, w) + 1\big)$, we obtain the expected result.

5 Experiments

In the following, we count the number of elementary operations needed to perform a scalar multiplication on an elliptic curve (with generic or special parameters) defined over a finite field $\mathbb{F}_p$ of size respectively 160 and 200 bits with various methods. More precisely we investigate

- the double and add, also known as the binary method and denoted by Bin.
- the ℓ-NAF$_w$ for $\ell = 2$ and $w = 2, 3, 4$, and 5
- the triple and add, also known as the ternary method and denoted by Tern.
- the 3-NAF$_2$
- the sextuple and add method, denoted by Sext.
- the 6-NAF$_2$
- the ternary/binary approach [CJLM05], denoted by Tern./bin.
- the Dual Base Number System (DBNS) as explained in [DIM05]. Note however that we did not try to tune the values of b_{max} and t_{max}, i.e. the biggest possible values for the powers of 2 and 3 in the expansion of k. This would certainly lead to big improvements.

In each case, we give the number $\#\mathcal{P}$ of precomputations needed to compute $[k]P$ when combined with a left-to-right approach. The density δ of the obtained expansion is also given. The different situations under scrutiny are:

A. Curve: $y^2 = x^3 + u(x+1)^3$ defined over a finite field $\mathbb{F}_p$ of odd characteristic.
   Operations:
   - tripling map [3] obtained as the composition of 2 isogenies expressed in new Jacobian coordinates
   - doubling and addition in new Jacobian coordinates

B. Curve: $y^2 = x^3 + ax + b$ defined over a finite field $\mathbb{F}_p$ of odd characteristic.
   Operations:
   - direct tripling formulas explained in [DIM05]
   - direct $[2]P \pm Q$ and $[3]P \pm Q$ explained in [CJLM05] whenever it is possible.

C. Same curve and same operations as in B. except that the direct tripling formulas come from [CJLM05].

We assume that the cost of a squaring is 0.8M. This allows us to express the complexity only in terms of inversions and multiplications. All the complexities are obtained in a theoretical way except for the ternary/binary and the DBNS approaches. In these cases, an average over 10^4 exponents has been computed. In each case, we provide the ratio between a multiplication and an inversion so that the complexities of this work and [DIM05] (resp. [CJLM05]) are equal. Thus, if I/M is bigger than the indicated value, our method will be more efficient. See Tables 1, 2, 3, and 4 for details.

6 Conclusion

We have described a family of elliptic curve defined over a prime field of large characteristic for which the multiplication–by–3 map, can be decomposed into the product of 2 isogenies. Explicit formulas indicate that a tripling can be done with $8\mathrm{M} + 6\mathrm{S}$, and even $6\mathrm{M} + 6\mathrm{S}$ if the parameter of the curve is suitably chosen. Since 3 plays an major role, we also tested generalizations of the width-w NAF expansion to deal with ℓ-adic expansions. We then tested our new tripling algorithm in different situations. When there is no memory constraints, the 3-NAF_2, 6-NAF_2, and 3-NAF_3 give excellent results for respectively only 2, 6 and 8 precomputed values and outclass their binary counterparts. Also, this system performs better than those described in [CJLM05] and [DIM05] for most methods under very realistic assumptions concerning the ratio I/M.

Of course, it would be desirable to extend this work and different directions are of interest. Indeed, the same study should be carried out in characteristic 2 and bigger values of ℓ should be investigated, the first candidate being 5. Also, the Dual Base Number System (DBNS) when combined with this new tripling method should give very good results with appropriate settings that need to be found. Also, designing direct formulas for $[2]P \pm Q$ and $[3]P \pm Q$ in new Jacobian coordinates would lead to further improvements.

References

[ACD+05] R. M. Avanzi, H. Cohen, C. Doche, G. Frey, T. Lange, K. Nguyen, and F. Vercauteren, *Handbook of Elliptic and Hyperelliptic Curve Cryptography*, CRC Press, Inc., 2005.

[CJLM05] M. Ciet, M. Joye, K. Lauter, and P. L. Montgomery, *Trading inversions for multiplications in elliptic curve cryptography*, Des. Codes Cryptogr. (2005), To appear. Also available from Cryptology ePrint Archive.

[CLSQ03] M. Ciet, T. Lange, F. Sica, and J.-J. Quisquater, *Improved algorithms for efficient arithmetic on elliptic curves using fast endomorphisms*, Advances in Cryptology – Eurocrypt 2003, Lecture Notes in Comput. Sci., vol. 2656, Springer-Verlag, 2003, pp. 388–400.

[CMO97] H. Cohen, A. Miyaji, and T. Ono, *Efficient elliptic curve exponentiation*, Information and Communication Security – ICICS 1997, Lecture Notes in Comput. Sci., vol. 1334, Springer-Verlag, Berlin, 1997, pp. 282–290.

[CMO98] ———, *Efficient elliptic curve exponentiation using mixed coordinates*, Advances in Cryptology – Asiacrypt 1998, Lecture Notes in Comput. Sci., vol. 1514, Springer-Verlag, Berlin, 1998, pp. 51–65.

[DIM05] V. S. Dimitrov, L. Imbert, and P. K. Mishra, *Efficient and secure elliptic curve point multiplication using double-base chains*, Advances in Cryptology – Asiacrypt 2005, Lecture Notes in Comput. Sci., vol. 3788, Springer-Verlag, Berlin, 2005, pp. 59–78.

[DJM99] V. S. Dimitrov, G. A. Jullien, and W. C. Miller, *Theory and applications of the double-base number system*, IEEE Trans. on Computers **48** (1999), no. 10, 1098–1106.

[GLV01] R. P. Gallant, R. J. Lambert, and S. A. Vanstone, *Faster point multiplication on elliptic curves with efficient endomorphisms*, Advances in Cryptology – Crypto 2001, Lecture Notes in Comput. Sci., vol. 2139, Springer-Verlag, Berlin, 2001, pp. 190–200.

[HMV03] D. Hankerson, A. J. Menezes, and S. A. Vanstone, *Guide to elliptic curve cryptography*, Springer-Verlag, Berlin, 2003.

[HT05] D.-G. Han and T. Takagi, *Some analysis of radix-r representations*, preprint, 2005. See `http://eprint.iacr.org/2005/402/`

[JMV05] D. Jao, S. D. Miller, and R. Venkatesan, *Do all elliptic curves of the same order have the same difficulty of discrete log?*, Advances in Cryptology – Asiacrypt 2005, Lecture Notes in Comput. Sci., vol. 3788, Springer-Verlag, Berlin, 2005.

[Kob92] N. Koblitz, *CM-curves with good cryptographic properties*, Advances in Cryptology – Crypto 1991, Lecture Notes in Comput. Sci., vol. 576, Springer-Verlag, Berlin, 1992, pp. 279–287.

[Lan05] T. Lange, *Koblitz curve cryptosystems*, Finite Fields Appl. **11** (2005), no. 2, 220–229.

[LD98] J. López and R. Dahab, *Improved algorithms for elliptic curve arithmetic in* $GF(2^n)$, Tech. Report IC-98-39, Relatório Técnico, October 1998.

[MO90] F. Morain and J. Olivos, *Speeding up the computations on an elliptic curve using addition-subtraction chains*, Inform. Theory Appl. **24** (1990), 531–543.

[MV90] A. J. Menezes and S. A. Vanstone, *The implementation of elliptic curve cryptosystems*, Advances in Cryptology – Auscrypt 1990, Lecture Notes in Comput. Sci., vol. 453, Springer-Verlag, Berlin, 1990, pp. 2–13.

[Sol00] J. A. Solinas, *Efficient arithmetic on Koblitz curves*, Des. Codes Cryptogr. **19** (2000), 195–249.

[TYW04] T. Takagi, S.-M. Yen, and B.-C. Wu, *Radix-r non-adjacent form*, Information Security Conference – ISC 2004, Lecture Notes in Comput. Sci., vol. 3225, Springer-Verlag, Berlin, 2004, pp. 99–110.

Method	$\#\mathcal{P}$	δ	A.	B.	I/M	C.	I/M
Bin.	—	1/2	2384M	80I + 1552M	10.4	160I + 1136M	7.8
NAF	—	1/3	2076M	53I + 1503M	10.8	160I + 947M	7.1
NAF_3	2	1/4	1928M	40I + 1480M	11.2	160I + 856M	6.7
NAF_4	4	1/5	1837M	32I + 1466M	11.6	160I + 800M	6.5
NAF_5	8	1/6	1780M	27I + 1457M	12	160I + 765M	6.3
Tern.	—	2/3	2057M	134I + 1321M	5.5	168I + 1164M	5.3
3-NAF_2	2	2/5	1749M	80I + 1391M	4.5	141I + 1110M	4.5
3-NAF_3	8	2/7	1623M	58I + 1419M	3.5	130I + 1088M	4.1
Sext.	—	5/6	1957M	52I + 1557M	7.7	124I + 1220M	5.9
6-NAF_2	6	5/11	1683M	28I + 1514M	6.1	124I + 1052M	5.1
Tern./bin.	—	—	1773M	36I + 1507M	7.4	127I + 1067M	5.6
DBNS	—	—	1883M	45I + 1519M	8.1	129I + 1113M	6

Table 1: Complexities with a 160bit size for a random curve

Method	$\#\mathcal{P}$	δ	A.	B.	I/M	C.	I/M
Bin.	—	1/2	2112M	80I + 1424M	8.6	160I + 1136M	6.1
NAF	—	1/3	1831M	53I + 1332M	9.4	160I + 947M	5.5
NAF_3	2	1/4	1696M	40I + 1288M	10.2	160I + 856M	5.2
NAF_4	4	1/5	1613M	32I + 1261M	11	160I + 800M	5.1
NAF_5	8	1/6	1561M	27I + 1244M	11.7	160I + 765M	5
Tern.	—	2/3	1788M	134I + 1287M	3.7	168I + 1164M	3.7
3-NAF_2	2	2/5	1507M	80I + 1330M	2.2	141I + 1110M	2.8
3-NAF_3	8	2/7	1392M	58I + 1347M	0.8	130I + 1088M	2.3
Sext.	—	5/6	1706M	52I + 1479M	4.4	124I + 1220M	3.9
6-NAF_2	6	5/11	1457M	28I + 1397M	2.1	124I + 1052M	3.3
Tern./bin.	—	—	1541M	36I + 1394M	4.1	127I + 1067M	3.7
DBNS	—	—	1643M	45I + 1415M	5	129I + 1113M	4.1

Table 2: Complexities with a 160bit size for a special curve

Method	$\#\mathcal{P}$	δ	A.	B.	I/M	C.	I/M
Bin.	—	1/2	2980M	100I + 1940M	10.4	200I + 1420M	7.8
NAF	—	1/3	2604M	67I + 1881M	10.8	200I + 1189M	7.1
NAF_3	2	1/4	2410M	50I + 1850M	11.2	200I + 1070M	6.7
NAF_4	4	1/5	2296M	40I + 1832M	11.6	200I + 1000M	6.5
NAF_5	8	1/6	2216M	33I + 1819M	12	200I + 951M	6.3
Tern.	—	2/3	2570M	168I + 1646M	5.5	210I + 1453M	5.3
3-NAF_2	2	2/5	2183M	100I + 1735M	4.5	176I + 1385M	4.5
3-NAF_3	8	2/7	2023M	72I + 1771M	3.5	162I + 1357M	4.1
Sext.	—	5/6	2424M	64I + 1932M	7.7	154I + 1511M	5.9
6-NAF_2	6	5/11	2093M	35I + 1880M	6.1	154I + 1308M	5.1
Tern./bin.	—	—	2221M	45I + 1887M	7.4	159I + 1337M	5.6
DBNS	—	—	2378M	58I + 1905M	8.1	162I + 1403M	6

Table 3: Complexities with a 200bit size for a random curve

Method	$\#\mathcal{P}$	δ	A.	B.	I/M	C.	I/M
Bin.	—	1/2	2640M	100I + 1780M	8.6	200I + 1420M	6.1
NAF	—	1/3	2297M	67I + 1668M	9.4	200I + 1189M	5.5
NAF_3	2	1/4	2120M	50I + 1610M	10.2	200I + 1070M	5.2
NAF_4	4	1/5	2016M	40I + 1576M	11	200I + 1000M	5.1
NAF_5	8	1/6	1943M	33I + 1552M	11.8	200I + 951M	5
Tern.	—	2/3	2234M	168I + 1604M	3.7	210I + 1453M	3.7
3-NAF_2	2	2/5	1881M	100I + 1659M	2.2	176I + 1385M	2.8
3-NAF_3	8	2/7	1735M	72I + 1681M	0.7	162I + 1357M	2.3
Sext.	—	5/6	2113M	64I + 1835M	4.4	154I + 1511M	3.9
6-NAF_2	6	5/11	1812M	35I + 1736M	2.2	154I + 1308M	3.3
Tern./bin.	—	—	1933M	45I + 1743M	4.2	159I + 1332M	3.8
DBNS	—	—	2077M	58I + 1777M	5.1	162I + 1404M	4.2

Table 4: Complexities with a 200bit size for a special curve