Toric forms of elliptic curves and their arithmetic

Wouter Castryck and Frederik Vercauteren

Abstract

This paper scans a large class of one-parameter families of elliptic curves for efficient arithmetic. The construction of the class is inspired by toric geometry, which provides a natural framework for the study of various forms of elliptic curves. The class both encompasses many prominent known forms and includes thousands of new forms. A powerful algorithm is described that automatically computes the most compact group operation formulas for any parameterized family of elliptic curves. The generality of this algorithm is further illustrated by computing uniform addition formulas and formulas for generalized Montgomery arithmetic.

1 Introduction

Since Lenstra's discovery of the elliptic curve factorization method [28] and the introduction of elliptic curve cryptography by Miller [29] and Koblitz [25], there has been a continuous interest in speeding up addition/doubling and (multi)-scalar multiplication on elliptic curves. Whereas Lenstra and Koblitz suggested to simply use the short Weierstrass equation and normal affine coordinates, Miller already proposed the use of Jacobian coordinates.

In a plethora of papers, many new coordinate systems and different forms were proposed. The most notable proposals are the following: Chudnovsky Jacobian coordinates [14] and modified Jacobian coordinates [16] (both using the short Weierstrass equation), Jacobi intersections [14, 27], the Hessian form [21, 35], the Jacobi quartic form [14, 9], the Montgomery form [31], the Doche/Icart/Kohel forms [17] and finally, the Edwards and twisted Edwards forms [6, 4]. All these forms and coordinate systems have been gathered in the Explicit Formulas Database by Bernstein and Lange [5], which also includes numerous speed-ups, mainly due to Bernstein and Lange themselves, and to Hisil et al. [23, 24]

The discovery of these different forms raises the question whether there are more unknown forms of interest and if so, if any of these forms lead to more efficient arithmetic. The goal of this paper is to provide an answer within a certain large class of forms. In this, we will always assume that we work over a field of sufficiently large characteristic. The class is inspired by results from toric geometry that give a natural classification of elliptic curves based on the Newton polytope of the defining polynomial, provided the latter satisfies a certain generic condition. This idea was presented by the first author at [10] – it subsequently proved useful in the construction of a characteristic 2 variant of Edwards arithmetic [7]. On the highest level, there are 16 non-equivalent base forms, only 6 of which seem to have appeared in the literature so far. On a somewhat lower level, i.e. by using unimodular transformations and specializing coefficients, one obtains an infinite number of forms, out of which we selected our class. It consists of over 50000 one-parameter families of elliptic curves, all of which we scanned for efficient arithmetic.

Of course, computing group operation formulas, let alone efficient formulas, in a large number of parameterized families soon becomes impossible by hand. To solve this problem, we propose a very general algorithm to compute efficient group operation formulas based on a combination of interpolation and lattice reduction. Alternatively, we could have used a rational simplication algorithm due to Monagan and Pearce [30], but our method is much more robust and avoids capricious Gröbner basis computations. To illustrate the robustness of our algorithm, we show how uniform addition formulas, i.e. formulas that can also be used for doubling, can be easily computed. Furthermore, we study generalized Montgomery arithmetic by considering the index 2 subfield of the function field of the curve that is invariant under negation. This notion generalizes x-coordinate only arithmetic on Montgomery curves. Again we provide non-trivial examples obtained by our algorithm.

At no point in this article, we claim immediate cryptographic applicability, neither are we blind for the current limitations of our scan: we restrict to prime fields of large enough characteristic, we restrict to affine formulas, some well-known forms are not covered by our class, and we tightly link efficiency with compactness. But the main aim of this paper is to provide techniques towards a more systematic approach in the search for efficient elliptic curve arithmetic. In particular, we emphasize that these limitations are not intrinsic to our method.

Nevertheless, there are some interesting conclusions to be drawn. First, we prove that within our class, the Edwards form is essentially the only form admitting quadratic doubling and addition formulas. Although our class is finite, it seems big enough to detect patterns, and in Proposition 9 we will give some theoretical evidence suggesting that the above conclusion is not a coincidence.

Another conclusion is that relatively good formulas are very common, so that designers for which other curve features are more important should not feel limited to the list of well-known forms. Our algorithm can then serve in finding efficient formulas for arithmetic.

Along the way, we obtain a number of theoretical results and side-way observations, regardless of our class. For instance, we obtain a good understanding of what can be expected from Montgomery arithmetic (in its most general setting) and we prove that one can never improve upon Montgomery's doubling formula, in the rough sense explained in Proposition 10. We show that many prominent known forms fit in our framework, and we illustrate that by studying the toric resolution of the curve, one can often guess good candidates for projective coordinate systems suited for efficient arithmetic. An off-topic contribution is the proof of various statistics concerning the number of isomorphism and isogeny classes of elliptic curves in certain families (Section 2.2).

The remainder of this paper is organized as follows. Section 2 recalls the notions of elliptic curve, addition formulas, doubling formulas, uniformity, and completeness. Five well-known and well-working examples are discussed, and some according statistics are proven. Generalized Montgomery arithmetic is introduced. Section 3 presents our framework based on toric geometry, along with the construction of our class of over 50000 forms. It is shown that many prominent known forms are contained in this class, including our five selected examples. Section 4 describes our interpolation algorithm to compute efficient addition/doubling formulas on families of elliptic curves and Section 5 illustrates the generality of this algorithm by providing addition/doubling formulas for several new families, uniform addition formulas for the short Weierstrass equation and generalized Montgomery arithmetic for the Jacobi quartic. Section 6 discusses the results of our scan. We also prove some prudent optimality results on Edwards and Montgomery arithmetic. Finally, Section 8 concludes the paper and presents avenues for future research.

2 Elliptic curves, addition, and doubling

2.1 Theoretical framework

Throughout this article, k denotes a perfect field (typically a finite field or a field containing $\mathbb{Q}$) and $\overline{k}$ denotes an algebraic closure.

An *elliptic curve* over k is a pair $(E, \mathcal{O})$. Here E is a curve of geometric genus one in $\mathbb{P}^2$, defined by the homogenization of an absolutely irreducible polynomial $C(x,y) \in k[x,y]$. We do not impose E to be non-singular: this is not standard, but it allows us to consider e.g. Edwards curves and Jacobian quartics as elliptic curves in a more natural way. In any case, there always exists a non-singular curve $\widetilde{E}/k$ along with a k-rational birational morphism $\lambda : \widetilde{E} \to E$ under which the non-singular part E_{ns} of E can be identified with a Zariski open subset of $\widetilde{E}$, i.e. $\lambda|_{\lambda^{-1}(E_{ns})}$ is an isomorphism. The points of $\widetilde{E}$ are called *places* of E, and a place $\mathcal{P}$ is said to *dominate* $\lambda(\mathcal{P})$. The singular points of E may be dominated by several places. The second parameter $\mathcal{O}$ is a k-rational place of E. By the above identification this is typically just a non-singular k-rational point.

The curve $\widetilde{E}$ is endowed with unique k-rational morphisms $\psi : \widetilde{E} \times \widetilde{E} \to \widetilde{E}$ and $\chi : \widetilde{E} \to \widetilde{E}$ that can be interpreted as addition and negation, turning $\widetilde{E} = \widetilde{E}(\overline{k})$ into an abelian group in which $\mathcal{O}$ serves as neutral element. Note that, for each intermediate field $k \subset k' \subset \overline{k}$, the k'-rational points $\widetilde{E}(k')$ form a subgroup of $\widetilde{E}$. We will write $\mathcal{P} + \mathcal{Q}$ for $\psi(\mathcal{P}, \mathcal{Q})$ and $-\mathcal{P}$ for $\chi(\mathcal{P})$. Changing the base place boils down to translating the group law: let $\mathcal{O}'$ be a new base place inducing new operations $+'$ and $-'$, then $\mathcal{P} +' \mathcal{Q} = ((\mathcal{P} - \mathcal{O}') + (\mathcal{Q} - \mathcal{O}')) + \mathcal{O}'$ and $-'\mathcal{P} = -(\mathcal{P} - \mathcal{O}') + \mathcal{O}'$. For each $n \in \mathbb{Z}$ the notation $[n]\mathcal{P}$ abbreviates

$$\operatorname{sgn}(n) \cdot \Big(\underbrace{\mathcal{P} + \cdots + \mathcal{P}}_{|n| \text{ times}} \Big).$$

The map $\varphi_n : \widetilde{E} \to \widetilde{E} : \mathcal{P} \to [n]\mathcal{P}$ is a degree n^2 morphism. We will be particularly interested in the doubling map φ_2.

The function fields of $\widetilde{E}$ and $\widetilde{E} \times \widetilde{E}$ are identified with the fraction fields of

$$\frac{k[x,y]}{(C(x,y))} \quad \text{and} \quad \frac{k[x_1,y_1,x_2,y_2]}{(C(x_1,y_1), C(x_2,y_2))}$$

respectively. We define a set of *addition formulas* on an elliptic curve $(E, \mathcal{O})$ to be a quartet of nonzero polynomials $f_1, g_1, f_2, g_2 \in k[x_1, y_1, x_2, y_2]$ such that

$$x \circ \psi = \frac{f_1}{g_1}, \qquad y \circ \psi = \frac{f_2}{g_2}$$

inside the function field $k(\widetilde{E} \times \widetilde{E})$. A set of *doubling formulas* is a quartet of nonzero polynomials $f_1, g_1, f_2, g_2 \in k[x, y]$ such that

$$x \circ \varphi_2 = \frac{f_1}{g_1}, \qquad y \circ \varphi_2 = \frac{f_2}{g_2}$$

inside the function field $k(\widetilde{E})$.

Let $U_{ns} = E_{ns} \cap \mathbb{A}^2$. Then addition formulas and doubling formulas can be used to perform arithmetic on generically chosen points of U_{ns}. For instance, let f_1, g_1, f_2, g_2 be a set of addition formulas and take points $\mathcal{P} = (p_1, p_2), \mathcal{Q} = (q_1, q_2) \in U_{ns}$. Then unless the denominators are zero, it makes sense to compute

$$\left(\frac{f_1(p_1,p_2,q_1,q_2)}{g_1(p_1,p_2,q_1,q_2)}, \frac{f_2(p_1,p_2,q_1,q_2)}{g_2(p_1,p_2,q_1,q_2)} \right).$$

If the result is in U_{ns} again, this exactly matches with $\mathcal{P} + \mathcal{Q}$. Point pairs of $(\widetilde{E} \times \widetilde{E}) \setminus (U_{ns} \times U_{ns})$, as well as point pairs of $U_{ns} \times U_{ns}$ where the above method fails, are called *exceptional point pairs* with respect to the given addition formulas. With respect to doubling formulas, it is straightforward to define the similar notion of *exceptional points*. Exceptional point (pair) sets are always of co-dimension ≥ 1.

2.2 Some well-known examples

Here are five famous shapes of elliptic curves. Evidently, the existing literature contains a lot more forms that have proven useful (see the references in the introduction), but the examples below are very classical – even the Edwards form, which in fact dates back to Gauss – and illustrative for the remainder of this paper.

(1) Assume $\operatorname{char} k \neq 2, 3$. A *Weierstrass curve* is an elliptic curve E defined by

$$C(x, y) = y^2 - x^3 - Ax - B \in k[x, y], \quad 4A^3 + 27B^2 \neq 0,$$

along with the unique point $\mathcal{O} = (0, 1, 0)$ at infinity. Such a curve is non-singular (thus $\widetilde{E} = E$ and $\lambda = \mathrm{id}$) and the group operations can be described using the well-known tangent-chord method. A naive calculation then gives the following addition formulas:

$$x \circ \psi = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)^2 - x_1 - x_2, \qquad y \circ \psi = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)(x_1 - x_3) - y_1,$$

where x_3 abbreviates $x \circ \psi$. Note that all point pairs for which $x_1 = x_2$ are exceptional. In particular, the above expressions are unsuitable for doubling, for which instead

$$x_3 = x \circ \varphi_2 = \left(\frac{3x^2 + A}{2y}\right)^2 - 2x, \qquad y \circ \varphi_2 = \left(\frac{3x^2 + A}{2y}\right)(x - x_3) - y$$

can be used.

(2) Assume $\operatorname{char} k \neq 3$. A *Hessian curve* is an elliptic curve E defined by

$$C(x, y) = x^3 + y^3 + 1 - 3dxy \in k[x, y], \quad d^3 \neq 1,$$

along with $\mathcal{O} = (-1, 1, 0)$. Hessian curves are non-singular, and again tangent-chord arithmetic applies. We refer to [21] for details on how to obtain the addition formulas

$$x \circ \psi = \frac{y_1^2 x_2 - y_2^2 x_1}{x_2 y_2 - x_1 y_1}, \qquad y \circ \psi = \frac{x_1^2 y_2 - x_2^2 y_1}{x_2 y_2 - x_1 y_1}$$

and for explicit doubling formulas. An interesting property of Hessian curves is that, although the diagonal of $U_{ns} \times U_{ns}$ belongs to the exceptional locus of the addition formulas, these can nevertheless be used to perform doubling, using the relation $[2](\alpha, \beta, \gamma) = (\gamma, \alpha, \beta) + (\beta, \gamma, \alpha)$. This feature is interesting against side-channel attacks. See also [35].

(3) Assume $\operatorname{char} k \neq 2$. An *Edwards curve* is an elliptic curve E which is defined by a polynomial

$$C(x, y) = x^2 + y^2 - 1 - dx^2y^2 \in k[x, y], \quad d \neq 0, 1,$$

along with the non-singular affine point $\mathcal{O} = (0, 1)$. Edwards curves allow the following elegant addition formulas:

$$x \circ \psi = \frac{x_1 y_2 + y_1 x_2}{1 + d x_1 x_2 y_1 y_2}, \qquad y \circ \psi = \frac{y_1 y_2 - x_1 x_2}{1 - d x_1 x_2 y_1 y_2}.$$

These are *uniform*, in the sense that under $x_1, x_2 \mapsto x$ and $y_1, y_2 \mapsto y$ they specialize to doubling formulas (see also Section 2.3). The curve E has two singular points, namely $(1, 0, 0)$ and $(0, 1, 0)$. It desingularizes to an intersection of quadrics

$$\widetilde{E} : \begin{cases} xy - zw = 0 \\ x^2 + y^2 - z^2 - dw^2 = 0 \end{cases}$$

in $\mathbb{P}^3$, which naturally projects onto $E \subset \mathbb{P}^2$ (the projection $\lambda : \widetilde{E} \to E$ corresponds to substituting $x \leftarrow xz$, $y \leftarrow yz$, $z \leftarrow z^2$, $w \leftarrow xy$). The place dominating $\mathcal{O}$ is $(0, 1, 1, 0)$. The places dominating $(1, 0, 0)$ are $(\sqrt{d}, 0, 0, 1)$ and $(-\sqrt{d}, 0, 0, 1)$, and the places dominating $(0, 1, 0)$ are $(0, \sqrt{d}, 0, 1)$ and $(0, -\sqrt{d}, 0, 1)$. Note that if d is a non-square, then $\widetilde{E}(k) \subset \mathbb{A}^2$, which is related to the *completeness* of Edwards addition in that case (see Section 2.3). The main references on Edwards arithmetic are [6, 4].

(4) Assume $\operatorname{char} k \neq 2$. A *Jacobi quartic* is an elliptic curve E defined by

$$C(x,y) = y^2 - x^4 + 2Ax^2 - 1 \in k[x,y], \quad A \neq \pm 2,$$

along with the affine point $\mathcal{O} = (0,1)$. In [9], Billet and Joye computed the following formulas:

$$x \circ \psi = \frac{x_1y_2 + x_2y_1}{1 - x_1^2x_2^2}, \quad y \circ \psi = \frac{\left(1 + x_1^2x_2^2\right)(y_1y_2 - 2Ax_1x_2) + 2x_1x_2\left(x_1^2 + x_2^2\right)}{1 - x_1^2x_2^2},$$

which are again uniform. The Jacobi quartic has a singular point $(0,1,0)$ at infinity. The desingularization map λ is the projection from the intersection in $\mathbb{P}^3$ of the quadrics $x^2 - zw$ and $y^2 - w^2 + 2Ax^2 - z^2$ to $\mathbb{P}^2$ (corresponding to substituting $x \leftarrow xz$, $y \leftarrow yz$, $z \leftarrow z^2$, $w \leftarrow x^2$). The place dominating $\mathcal{O}$ is $(0,1,1,0)$. The places dominating $(0,1,0)$ are $(0,1,0,1)$ and $(0,-1,0,1)$.

(5) Assume $\operatorname{char} k \neq 2$. A *Montgomery curve* is an elliptic curve E defined by a polynomial

$$C(x,y) = By^2 - x^3 - Ax^2 - x \in k[x,y], \quad B \neq 0,\ A \neq \pm 2$$

along with the unique point $\mathcal{O} = (0,1,0)$ at infinity. Tangent-chord arithmetic applies and it is easy to see that $k(x)$ is the index 2 subfield invariant under negation, so one can take $t = x$. Montgomery [31] proved the following formulas:

$$x \circ \varphi_2 = \frac{(x+1)^2(x-1)^2}{4x((x-1)^2 + \frac{A+2}{4}((x+1)^2 - (x-1)^2))}$$

$$x_{m+n}x_{m-n} = \frac{((x_m - 1)(x_n + 1) + (x_m + 1)(x_n - 1))^2}{((x_m - 1)(x_n + 1) - (x_m + 1)(x_n - 1))^2},$$

where $x_i = x \circ \varphi_i$.

Intermezzo: classification of the above forms.

In the following discussion, we always assume that k is a finite field having an appropriate characteristic ($\operatorname{char} k \geq 5$ will work everywhere). It is well-known that every elliptic curve is in k-rational birational equivalence with a Weierstrass curve, but the same is no longer true for the other forms **(2-5)**. In this intermezzo, we will give a brief classification, both up to k-birational equivalence and up to k-isogeny, and prove according statistics. It is essentially a summary of existing (yet fragmentary) material. Similar statistics have been observed, without proofs or references, in [4, Section 4]. Classication up to $\overline{k}$-birational equivalence can be done through a j-invariant computation, which was carried out in [34] in a limited number of cases. Note that in the concluding section of [34], classification up to k-isogeny is presented as an open problem.

(2) If $(E,\mathcal{O})$ is Hessian, then $\widetilde{E}(k)$ has a subgroup $\{\mathcal{O},(-1,0,1),(0,-1,1)\}$ of order 3. In particular, $3 \mid \#\widetilde{E}(k)$ is a necessary condition for an elliptic curve $(E,\mathcal{O})$ to be k-birationally equivalent to a Hessian curve. If $\#k \equiv 2 \bmod 3$ then this condition is also sufficient [15, 13.1.5.b]. If $\#k \equiv 1 \bmod 3$, then $(E,\mathcal{O})$ is Hessian if and only if $\widetilde{E}(k)$ contains all nine 3-torsion points of $\widetilde{E}(\overline{k})$. For the if-part, the proof goes as follows. Take a model in which $\mathcal{O}$ is a flex (e.g. a Weierstrass model). Then by the tangent-chord rule, its other flexes are exactly its other 3-torsion points, hence k-rational. But then an additional projective transformation puts our curve into Hessian form [20, Lemma 11.36]. For the only-if-part it suffices to observe that the flexes of a Hessian curve are precisely the intersection points with the three coordinate axes, all of them being rational if $\#k \equiv 1 \bmod 3$.

The probability that a randomly chosen Weierstrass curve can be shaped into Hessian form can then be estimated as

$$P(3 \mid \#\widetilde{E}(k)) \approx \frac{1}{2} \qquad \text{if } \#k \equiv 2 \bmod 3$$

$$P(\widetilde{E}[3](k) \cong \mathbb{Z}/3\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}) \approx \frac{1}{\#\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})} = \frac{1}{24} \qquad \text{if } \#k \equiv 1 \bmod 3,$$

following [13, Theorems 1 and 2]. The error term is $O(\#k^{-1/2})$.

Two elliptic curves $(E, \mathcal{O})$ and $(E', \mathcal{O}')$ are k-isogenous if and only if $\#\widetilde{E}(k) = \#\widetilde{E}'(k)$ by Tate's theorem. So if $\#k \equiv 2 \bmod 3$, a necessary and sufficient condition for an elliptic curve $(E, \mathcal{O})$ to be k-isogenous to a Hessian curve is $3 \mid \#\widetilde{E}(k)$. If $\#k \equiv 1 \bmod 3$, then the condition $9 \mid \#\widetilde{E}(k)$ is necessary and almost always sufficient. This follows from a very general theorem due to Tsfasman et al. [36, Theorem 3.3.15]. The exceptions are all supersingular.

(5) Montgomery curves can be intrinsically characterized by the existence of a point $\mathcal{P} \in \widetilde{E}(k)$ for which $\left(\widetilde{E}/\langle\mathcal{P}\rangle\right)(k)$ contains all four 2-torsion points: combine [32, Proposition 5] with Vélu's formulas to see this. Equivalently, an elliptic curve $(E, \mathcal{O})$ is Montgomery if the curve or its quadratic twist have a k-rational point of order 4; see also [4, Theorems 3.2 and 3.3]. In case $\#k \equiv 3 \bmod 4$, it suffices to check whether $(E, \mathcal{O})$ itself has a point of order 4 [4, Theorem 3.4]. In case $\#k \equiv 1 \bmod 4$, the curve $(E, \mathcal{O})$ is Montgomery if and only if $4 \mid \#\widetilde{E}(k)$: it suffices to verify that every (2×2)-matrix over $\mathbb{Z}/4\mathbb{Z}$ having trace 2 and determinant $\#k \equiv 1$ is conjugated to a matrix of the form

$$\begin{pmatrix} 1 & w \\ 0 & 1 \end{pmatrix} \quad \text{or of the form} \quad \begin{pmatrix} -1 & w \\ 0 & -1 \end{pmatrix}$$

(implying that $(E, \mathcal{O})$ resp. its quadratic twist have a k-rational point of order 4).

The probability that a randomly chosen Weierstrass curve can be shaped into Montgomery form can then be estimated as

$$P(\widetilde{E}(k) \text{ contains point of order } 4) \approx \frac{3}{8} \qquad \text{if } \#k \equiv 3 \bmod 4$$

$$P(4 \mid \#\widetilde{E}(k)) \approx \frac{5}{12} \qquad \text{if } \#k \equiv 1 \bmod 4,$$

following [13, Theorems 1 and 3]. The error term is $O(\#k^{-1/2})$.

Clearly, a necessary condition for an elliptic curve $(E, \mathcal{O})$ to be k-isogenous to a Montgomery curve is $4 \mid \#\widetilde{E}(k)$. This is also sufficient: an explicit isogeny is given in [4, Theorem 5.1].

(3) Up to k-rational birational equivalence, Edwards curves are precisely those elliptic curves having a k-rational point of order 4, see [4, Theorem 3.3]. Note that in particular, every Edwards curve is a Montgomery curve and, conversely, every Montgomery curve is a twist of an Edwards curve.

The probability that a randomly chosen Weierstrass curve can be shaped into Edwards form can then be estimated as

$$P(\widetilde{E}(k) \text{ contains point of order } 4) \approx \frac{3}{8} \qquad \text{if } \#k \equiv 3 \bmod 4$$

$$P(\widetilde{E}(k) \text{ contains point of order } 4) \approx \frac{1}{3} \qquad \text{if } \#k \equiv 1 \bmod 4,$$

following [13, Theorems 1 and 3]. The error term is $O(\#k^{-1/2})$.

In the case where k is finite, an application of [36, Theorem 3.3.15] classifies Edwards curves up to isogeny: if $4 \mid \#\widetilde{E}(k)$, then $(E, \mathcal{O})$ only fails to be k-isogenous to an Edwards curve if $\#k$ is a square and $\widetilde{E}$ is a supersingular curve having $(\sqrt{\#k} \pm 1)^2$ rational points, the sign to be chosen such that $4 \nmid \sqrt{\#k} \pm 1$. In particular, if $\#k \equiv 3 \bmod 4$, then $\#k$ is never a square and an explicit k-isogeny can be constructed following [4, Theorems 3.2, 3.4 and 5.1].

(4) A Weierstrass curve can be k-birationally transformed to a Jacobi quartic if and only if all 2-torsion points are k-rational. This can be read along the lines in [9, Section 3]. Using [13, Theorem 2], the proportion of Weierstrass curves shapable into Jacobi quartic form is then given by

$$P(\widetilde{E}[2](k) \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}) \approx \frac{1}{\#\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})} = \frac{1}{6}.$$

The error term is $O(\#k^{-1/2})$.

Again by [36, Theorem 3.3.15], apart from some explicitly known supersingular exceptions, if $4 \mid \#\widetilde{E}(k)$ then $(E, \mathcal{O})$ is k-isogenous to a Jacobi quartic.

2.3 Uniformity and completeness

We call a set of addition formulas *uniform* if they specialize under $x_i \mapsto x$, $y_i \mapsto y$ $(i = 1, 2)$ to a set of doubling formulas. This will be the case whenever g_1 and g_2 do not identically vanish (over $\overline{k}$) on the diagonal of $U_{ns} \times U_{ns}$. Uniform addition formulas always exist (whatever E and $\mathcal{O}$ are), simply because ψ is a morphism. Indeed, it suffices to take a point $\mathcal{P} \in U_{ns}(\overline{k})$ such that $x \circ \psi$ and $y \circ \psi$ are defined on $(\mathcal{P}, \mathcal{P})$. Then there exists an open neighborhood $W \ni (\mathcal{P}, \mathcal{P})$ in $U_{ns} \times U_{ns}$ and polynomials $f_1, g_1, f_2, g_2 \in k[x_1, y_1, x_2, y_2]$ for which $x \circ \psi = f_1/g_1$ and $y \circ \psi = f_2/g_2$, such that g_1, g_2 nowhere vanish on W. Uniformity is an interesting feature against side-channel attacks. We already encountered uniform addition formulas for Edwards and Jacobi quartic curves. In Section 5.3 we provide an example for the short Weierstrass form computed by the algorithm described in Section 4.

A set of addition formulas (resp. doubling formulas) is said to be *complete* if $U_{ns}(k) \times U_{ns}(k)$ (resp. $U_{ns}(k)$) contains no exceptional point pairs (resp. exceptional points). As soon as $U_{ns}(k) \neq \emptyset$, complete addition formulas are automatically uniform. However, whereas uniformity is a property that is invariant under base field extension, completeness is not.

Lemma 1: If $k = \overline{k}$, then complete addition or doubling formulas do not exist.

Proof. $U_{ns} = U_{ns}(\overline{k}) \neq \emptyset$, thus it suffices to prove that complete doubling formulas cannot exist. Let $\mathcal{P}$ be a place above a point at infinity. Since $\widetilde{E} \setminus U_{ns}$ is finite, there is a minimal r such that $\varphi_{2^r}^{-1}\{\mathcal{P}\}$ contains a point $\mathcal{Q} \in U_{ns}$. This will be an exceptional point. ■

The best-known example of complete addition formulas (and hence of complete doubling formulas) is provided by Edwards addition, as described above. Indeed, if d is taken to be non-square, then $dx_1x_2y_1y_2$ can never be ± 1. See [6, Theorem 3.3] for details.

2.4 Generalized Montgomery arithmetic

Let $(E, \mathcal{O})$ be an elliptic curve. The subfield of $k(\widetilde{E})$ consisting of functions f that satisfy $f = f \circ \chi$ is of the form $k(t)$ for a non-constant function $t \in k(\widetilde{E})$. Equivalently, $k(t)$ consists of all functions f that satisfy $f(\mathcal{P}) = f(-\mathcal{P})$ for all pairs $\pm\mathcal{P}$ at which f is defined. It is a subfield of index 2, corresponding to a degree 2 morphism $\widetilde{E} \to \mathbb{P}^1$. For example, in **(1)** the Weierstrass setting, the map $f \mapsto f \circ \chi$ is determined by $x \mapsto x, y \mapsto -y$. So the subfield is just $k(x)$, and one can take $t = x$. In **(2)** the Hessian setting, the map $f \mapsto f \circ \chi$ is determined by $x \mapsto y, y \mapsto x$ and one can take $t = x + y$. In **(3)** the Edwards setting, we have $x \mapsto -x, y \mapsto y$ so we can take $t = y$. In **(4)** the Jacobi quartic setting, the map is $x \mapsto -x, y \mapsto y$ and one can take $t = (y+1)/x^2$. Note that $k(y) \subsetneq k(t)$. Finally, **(5)** for Montgomery curves one can take $t = x$ as in the Weierstrass setting. The following lemma is easy to prove by noting that $\varphi_n \circ \chi = \chi \circ \varphi_n$.

Lemma 2: For all $n \in \mathbb{Z}$, we have $t \circ \varphi_n \in k(t)$.

A *t-only doubling formula* is a couple of nonzero polynomials $f, g \in k[t]$ such that $t \circ \varphi_2 = f/g$ inside $k(t)$. Concerning addition, it is in general impossible to derive $t(\mathcal{P} + \mathcal{Q})$ from $t(\mathcal{P})$ and $t(\mathcal{Q})$. Instead, one makes use of the next statement, which is easy to verify using the classical Weierstrass addition formulas (it holds in any characteristic).

Lemma 3: There exists a bivariate rational function F over k such that for all $m, n \in \mathbb{Z}$

$$(t \circ \varphi_{m+n})(t \circ \varphi_{m-n}) = F(t \circ \varphi_n, t \circ \varphi_m) \quad \text{in } k(t). \tag{1}$$

A *t-only addition formula* is a couple of nonzero polynomials $f, g \in k[t_n, t_m]$ such that $F = f/g$ satisfies (1). Here t_n and t_m are formal variables.

Then t-only arithmetic can be used to compute $[n]\mathcal{P}$ by subsequently obtaining $([n]\mathcal{P}, [n+1]\mathcal{P})$ from $([\lfloor n/2 \rfloor]\mathcal{P}, [\lfloor n/2 \rfloor + 1]\mathcal{P})$, using one t-only doubling and one t-only addition. This is the so-called *Montgomery ladder*; for more details we refer to [15, 13.2.3]. Montgomery proposed this method [31] in the context of speeding up Lenstra's elliptic curve factorization method [28], although soon after it found its way to cryptography (e.g. in Bernstein's `curve25519` [3]).

2.5 Projective coordinates

To avoid time-costly field inversions, addition and doubling are commonly done using projective coordinates, see for instance [15, 13.2.1.b]. The same principle is used for Montgomery arithmetic: one then works on the projective t-line $\mathbb{P}^1$. Now instead of projective coordinates, one can often gain a speed-up using alternative coordinate systems, the most famous being weighted projective coordinates. E.g., in the Weierstrass setting, it is more natural to work in $\mathbb{P}(2;3;1)$, these are called *Jacobian coordinates*; see [15, 13.2.1.c] for some details. Similarly, one preferably works in a $\mathbb{P}(1;2;1)$-related coordinate system for Jacobi quartics. It can also be useful to work with hyperboloidal coordinates, i.e. to work on $\mathbb{P}^1 \times \mathbb{P}^1$ (which is the quadric $xy = zw$ in $\mathbb{P}^3$). For that, one embeds a point (x, y) as $(x, y, 1, xy)$. This setting gave some of the best operation counts so far for Edwards arithmetic [24]. We refer to the Explicit Formulas Database [5] for an overview of the various other inversion-free coordinate systems that have been proposed.

3 Toric forms of elliptic curves

3.1 Nondegenerate polynomials, lattice polytopes, and equivalence

For the general background on toric varieties and nondegenerate polynomials, we refer to [18, 1].

Let $C(x, y) \in k[x, y]$ be an absolutely irreducible polynomial. Let $S \in \mathbb{Z}^2$ be the set of exponent vectors appearing in C, and denote with Δ its convex hull in $\mathbb{R}^2$. It is called the *Newton polytope* of C, which is an example of a *lattice polytope*, i.e. a convex polytope whose vertices lie in $\mathbb{Z}^2$. A *face* of Δ is either Δ itself, either a non-empty intersection of Δ with a line $aX + bY = c$ for which $\Delta \subset \{(X, Y) \in \mathbb{R}^2 \,|\, aX + bY \leq c\}$. The 1-dimensional faces are called *edges*, the 0-dimensional faces are referred to as *vertices*. The union over the edges of Δ is called the *boundary* and is denoted by $\partial\Delta$. For each subset $\tau \subset \mathbb{R}^2$, let $C_\tau(x, y)$ be obtained from $C(x, y)$ by erasing all terms whose exponent vectors lie outside of τ.

Definition/Theorem 4: *[11, Corollary 2.8]* Suppose that for each face $\tau \subset \Delta$, the system of equations

$$C_\tau(x, y) = \frac{\partial}{\partial x} C_\tau(x, y) = \frac{\partial}{\partial y} C_\tau(x, y) = 0$$

has no solutions in the *torus* $\mathbb{T}^2 = (\overline{k} \setminus 0)^2 \subset \mathbb{A}^2$, then $C(x, y)$ is called *nondegenerate with respect to its Newton polytope*. In that case, the geometric genus of the curve defined by $C(x, y) = 0$ equals $\#((\Delta \setminus \partial\Delta) \cap \mathbb{Z}^2)$.

Regardless of the condition of nondegeneracy, $\#((\Delta \setminus \partial\Delta) \cap \mathbb{Z}^2)$ is an upper bound for the geometric genus of the curve defined by $C(x, y) = 0$. This is called *Baker's inequality*, for a proof see [2, Theorem 4.2].

We now consider *$\mathbb{Z}$-affine maps*

$$f : \mathbb{R}^2 \to \mathbb{R}^2 : \begin{bmatrix} X \\ Y \end{bmatrix} \mapsto A \cdot \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} a \\ b \end{bmatrix}$$

for $a, b \in \mathbb{Z}$ and $A \in \mathrm{GL}_2(\mathbb{Z})$. Two lattice polytopes $\Delta, \Delta' \subset \mathbb{R}^2$ are called *equivalent* if there exists a $\mathbb{Z}$-affine map f such that $f(\Delta) = \Delta'$. Two absolutely irreducible polynomials $C(x, y), C'(x, y) \in k[x, y]$ are called *equivalent* if C' can be obtained from C by applying a $\mathbb{Z}$-affine map to its exponent vectors. This procedure actually induces an isomorphism between their respective loci in $\mathbb{T}^2$. Equivalent polynomials have equivalent Newton polytopes, and share their being nondegenerate or not.

In the spirit of Theorem 4, we define the *genus* of a lattice polytope Δ to be $\#((\Delta \setminus \partial\Delta) \cap \mathbb{Z}^2)$. For a fixed $g \geq 1$, there is a finite number of equivalence classes of lattice polytopes of genus g. If $g = 1$, there are 16 equivalence classes. Lattice polytopes representing these are shown in Figure 1 below, which was taken from [33].

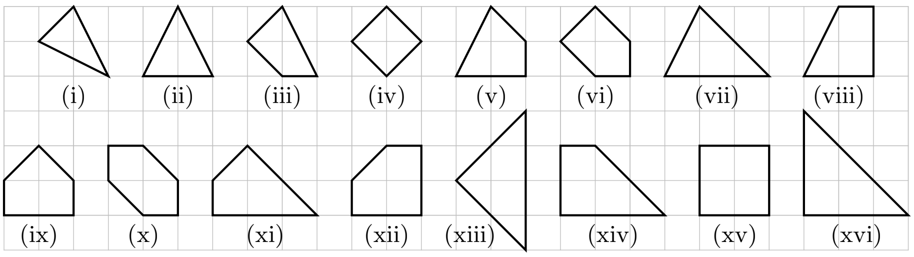


Figure 1: The 16 equivalence classes of lattice polytopes of genus 1

Using the above theory, we can prove the following simple observation, which seems new.

Lemma 5: Let k be a field of characteristic 0. Any geometric genus one curve over k which is defined by a bivariate trinomial has j-invariant 0 or 1728.

Proof. After rescaling and applying a suitable $\mathbb{Z}$-affine map to the exponent vectors, one sees that the locus in $\mathbb{T}^2$ is isomorphic to a curve defined by

$$1 + \alpha y^j + \beta x^k y^\ell \quad \in k[x, y],$$

with $\alpha, \beta \neq 0$ and $j, k > 0$. The $\overline{k}$-isomorphism

$$x \leftarrow \alpha^{\frac{\ell}{jk}} \beta^{-\frac{1}{k}} x, \qquad y \leftarrow \alpha^{-\frac{1}{j}} y$$

transforms the defining polynomial into

$$1 + y^j + x^k y^\ell,$$

which is nondegenerate with respect to its Newton polytope. This Newton polytope therefore contains exactly one interior lattice point. According to Figure 1, up to a $\mathbb{Z}$-affine map we remain with one of the following forms:

$$y^2 + y + x^3, \quad y^2 + x^3 + x, \quad y^2 + x^3 + 1, \quad y^3 + x^3 + 1, \quad y^2 + x^4 + 1.$$

Their j-invariants are 0, 1728, 0, 0 and 1728 respectively. ■

It follows from the proof that the lemma is still true if k is of sufficiently large finite characteristic (when compared to the degree of the trinomial).

3.2 The non-singular model of a toric form

From now on, we fix an irreducible polynomial $C(x, y) \in k[x, y]$, and we assume that it is nondegenerate with respect to its Newton polytope Δ. We also assume that $C(x, y)$ defines a curve $E \subset \mathbb{P}^2$ of geometric genus one, although most of the statements below are true for arbitrary genus. By Theorem 4, Δ has exactly one $\mathbb{Z}^2$-point in its interior. Hence up to equivalence, it is one of the polytopes listed in Figure 1.

We will follow the notation introduced in Section 2.1. The desingularization map

$$\lambda : \widetilde{E} \to E$$

can be described very explicitly. To each point $(i,j) \in \Delta \cap \mathbb{Z}^2$, associate a variable z_{ij}. These will be considered as coordinate functions on $\mathbb{P}^N$, where $N = \#(\Delta \cap \mathbb{Z}^2) - 1$. The combinatorics of Δ gives rise to a set of *binomial relations* in $k[z_{ij}]$: for $\alpha_1(i_1,j_1) + \alpha_2(i_2,j_2) = \beta_1(k_1,\ell_1) + \beta_2(k_2,\ell_2)$, where the α_i and β_i are integers satisfying $\alpha_1 + \alpha_2 = \beta_1 + \beta_2$, we have the relation

$$z_{i_1 j_1}^{\alpha_1} z_{i_2 j_2}^{\alpha_2} - z_{k_1 \ell_1}^{\beta_1} z_{k_2 \ell_2}^{\beta_2} = 0. \tag{2}$$

These relations can be shown to define a projective surface $X(\Delta) \subset \mathbb{P}^N$, which is called the *toric surface associated to* Δ. The torus $\mathbb{T}^2$ can be canonically embedded in $X(\Delta)$ by

$$\mathbb{T}^2 \hookrightarrow X(\Delta) : (x,y) \mapsto (x^i y^j)_{(i,j) \in \Delta \cap \mathbb{Z}^2}. \tag{3}$$

One can prove that it suffices to restrict to $\alpha_1 + \alpha_2 \leq 3$ and even to $\alpha_1 + \alpha_2 \leq 2$ whenever $\#(\partial\Delta \cap \mathbb{Z}^2) > 3$ [26].

The faces $\tau \subset \Delta$ naturally partition $X(\Delta)$ into sets of the form

$$O(\tau) = \left\{ (\alpha_{ij})_{(i,j) \in \Delta \cap \mathbb{Z}^2} \in X(\Delta) \,\middle|\, \alpha_{ij} \neq 0 \iff (i,j) \in \tau \right\},$$

which are called the *toric orbits* of $X(\Delta)$. Note that $O(\Delta)$ is precisely the image of the above map (3), hence it has the structure of a torus $\mathbb{T}^2$. More generally, each orbit $O(\tau)$ is canonically isomorphic to a torus $\mathbb{T}^{\dim \tau}$. Points in $X(\Delta) \setminus O(\Delta)$ are said to lie at *toric infinity.*

Now $C(x,y)$ itself defines one additional, linear relation in $\mathbb{P}^N$: if

$$C(x,y) = \sum_{(i,j) \in \Delta \cap \mathbb{Z}^2} c_{ij} x^i y^j, \quad \text{then it is} \quad \sum_{(i,j) \in \Delta \cap \mathbb{Z}^2} c_{ij} z_{ij} = 0.$$

This cuts out a curve $\widetilde{E}$ in $X(\Delta)$ which is birationally equivalent to E: this is easy to see using (3). More generally, for a positive integer n, we define the *Minkowski multiple* $n\Delta$ of a lattice polytope Δ as the lattice polytope obtained by 'dilating Δ with a factor n', i.e. by taking the convex hull in $\mathbb{R}^2$ of all points (na, nb) for which $(a,b) \in \Delta$. A straightforward calculation shows that there is a natural isomorphism $X(\Delta) \to X(n\Delta)$ such that the torus embedding $\mathbb{T}^2 \hookrightarrow X(n\Delta)$ is in fact the composition of the torus embedding $\mathbb{T}^2 \hookrightarrow X(\Delta)$ with this isomorphism. Now suppose the Newton polytope of $C(x,y)$ is $n\Delta$. Then its image in $X(\Delta)$ is cut out by a hypersurface of degree n. We still denote this image by $\widetilde{E}$.

The following theorem is the main statement on nondegenerate polynomials.

Theorem 6: The curve $\widetilde{E}$ is non-singular and intersects the 1-dimensional orbits $O(\tau)$ (corresponding to the edges τ of Δ) transversally in $\#(\tau \cap \mathbb{Z}^2) - 1$ points. It does not contain the 0-dimensional orbits (corresponding to the vertices of Δ). In particular, the number of points at toric infinity equals $\#(\partial\Delta \cap \mathbb{Z}^2)$. Moreover, these properties fully characterize the nondegeneracy of $C(x,y)$.

We can now describe the desingularization map $\lambda : \widetilde{E} \to E$. The restriction map $\lambda\,|_{O(\Delta)}$ is an isomorphism onto $E \cap \mathbb{T}^2 \subset U_{ns}$ whose inverse is given by the embedding (3). Now suppose $\mathcal{P} \in O(\tau) \cap \widetilde{E}$ for an edge $\tau \subset \Delta$. Let $\theta \in [0, 2\pi[$ be such that $(\cos\theta, \sin\theta)$ is a normal vector on τ that points towards the interior of Δ. Write $\mathcal{P} = (\alpha_{ij})_{(i,j) \in \Delta \cap \mathbb{Z}^2}$. Then

1. if $\theta = 0$, then $\lambda(P) = (0, \alpha_{0,k}, \alpha_{0,k-1})$ where $(0,k), (0,k-1) \in \tau \cap \mathbb{Z}^2$;
2. if $\theta \in]0, \pi/2[$, then $\lambda(P) = (0,0,1)$;
3. if $\theta = \pi/2$, then $\lambda(P) = (\alpha_{k,0}, 0, \alpha_{k-1,0})$ where $(k,0), (k-1,0) \in \tau \cap \mathbb{Z}^2$;
4. if $\theta \in]\pi/2, 5\pi/4[$, then $\lambda(P) = (1,0,0)$;
5. if $\theta = 5\pi/4$, then $\lambda(P) = (\alpha_{k+1,\ell}, \alpha_{k,\ell+1}, 0)$ where $(k+1,\ell), (k,\ell+1) \in \tau \cap \mathbb{Z}^2$;
6. if $\theta \in]5\pi/4, 2\pi[$, then $\lambda(P) = (0,1,0)$.

In cases (a), (c) and (e), the restriction map $\lambda\,|_{O(\tau)}$ is one-to-one.

3.3 The toric framework of well-known forms

The reader can verify that the basic forms **(1-5)** of Section 2.2 all fit in the above setting, i.e. they are all defined by a nondegenerate polynomial whose Newton polytope is therefore contained in Figure 1 (up to $\mathbb{Z}$-affine equivalence): Weierstrass curves are represented by (vii), Hessian curves by (xvi), Edwards curves by (xv), Jacobi quartics by (xiii) and Montgomery curves by (v). These five polytope classes seem to be the only cases that have been addressed in the literature so far, with the recent exception of binary Edwards curves [7], represented by (xii), that were designed for usage over fields of characteristic two only. Note that the nondegeneracy of Weierstrass polynomials proves that every elliptic curve can be shaped into a nondegenerate form, at least if $\operatorname{char} k > 3$. This is true for any k [12, Section 4]. Also note that in all cases, the base point $\mathcal{O}$ lies at toric infinity. Let us have look at the toric picture of these forms in closer detail.

(1) The Newton polytope Δ_W of a Weierstrass curve E defines the toric surface

$$X(\Delta_W): \quad z_{00}z_{20} = z_{10}^2, \quad z_{10}z_{20} = z_{00}z_{30}, \quad z_{11}z_{00} = z_{10}z_{0,1}, \quad z_{02}z_{00} = z_{01}^2$$

in $\mathbb{P}^6$. The hyperplane $z_{02} = z_{30} + Az_{10} + Bz_{00}$ cuts out a nonsingular model $\widetilde{E}$ of E, which of course was itself already nonsingular. The unique place dominating $\mathcal{O} = (0,1,0)$ is $(0,0,0,1,0,0,1)$, where the 1's correspond to the variables z_{30} and z_{02}. Now for practical applications, we do not suggest to work with coordinates in $\mathbb{P}^6$. But we remark that $X(\Delta_W)$ is exactly how the weighted projective space $\mathbb{P}(2;3;1)$ is canonically realized as a projective surface: it is the image of

$$\varphi : \mathbb{P}(2;3;1) \hookrightarrow \mathbb{P}^6 : (x,y,z) \mapsto (z^6, xz^4, x^2z^2, x^3, yz^3, xyz, y^2)$$

and under this map, the natural embedding of E in $\mathbb{P}(2;3;1)$ is precisely sent to $\widetilde{E}$. Thus, the toric picture of a Weierstrass curve is its natural embedding in $\mathbb{P}(2;3;1)$.

(2) The Newton polytope of a Hessian curve E is $\Delta_H = \operatorname{Conv}\{(0,0),(3,0),(0,3)\}$, which is 3Σ, where Σ is the standard 2-simplex in $\mathbb{R}^2$. The toric surface $X(\Sigma)$ is simply $\mathbb{P}^2$, and the toric model $\widetilde{E}$ is cut out by the cubic relation

$$z_{00}^3 + z_{10}^3 + z_{01}^3 = 3dz_{10}z_{01}z_{00},$$

hence the toric picture of a Hessian curve is the curve itself. One can verify that $X(\Delta_H)$ is the 3-uple embedding of $\mathbb{P}^2$ in $\mathbb{P}^9$, where the Hessian curve becomes a hyperplane section.

(3) The Newton polytope of an Edwards curve E is $\Delta_E = \operatorname{Conv}\{(0,0),(2,0),(0,2),(2,2)\}$, which is the Minkowski double of $\square = \operatorname{Conv}\{(0,0),(1,0),(0,1),(1,1)\}$. The toric surface $X(\square)$ is the surface in $\mathbb{P}^3$ defined by $z_{01}z_{10} = z_{00}z_{11}$, that is: it is $\mathbb{P}^1 \times \mathbb{P}^1$. The toric model $\widetilde{E}$ of E is cut out by an additional quadratic relation

$$z_{10}^2 + z_{01}^2 = z_{00}^2 + dz_{11}^2.$$

Renaming $z_{00} \leftarrow z$, $z_{10} \leftarrow x$, $z_{10} \leftarrow y$, $z_{11} \leftarrow w$ reveals a complete match with the description of Edwards curves given in Section 2.2, to which we refer for the rest of the story.

(4) Similarly, one can verify that the toric surface associated to the Newton Polytope Δ_J of a Jacobi quartic E is the conic $z_{10}^2 = z_{00}z_{10}$ (which is in fact $\mathbb{P}(1;2;1)$) and that $\widetilde{E}$ is cut out by $z_{01}^2 = z_{20}^2 + Az_{10}^2 + z_{00}^2$. Again compare this with the description given Section 2.2.

(5) The toric surface associated to the Newton polytope of a Montgomery curve is the blow-up of $\mathbb{P}(2;3;1)$ in $(0,0,0)$. The blow-up is only necessary to ensure that the Montgomery curve does not contain the 0-dimensional toric orbit $(0,0,0)$; cf. Theorem 6.

Toric surfaces are generalizations of projective space, and can serve as an inspiration for the choice of a coordinate system in which to perform efficient arithmetic (see also Section 2.5). Weighted projective coordinates for Weierstrass curves ($X(\Delta_W) \cong \mathbb{P}(2;3;1)$) and Jacobi quartics ($X(\Delta_J) \cong \mathbb{P}(1;2;1)$) have proven to be useful [14]. It is probably not a coincidence that Hisil et al. established their speed-records for Edwards curve arithmetic [24] using hyperboloidal coordinates ($X(\Delta_E) \cong \mathbb{P}^1 \times \mathbb{P}^1$) and that ordinary projective coordinates remain in many aspects the better system for Hessian curves ($X(\Delta_H) \cong \mathbb{P}^2$).

3.4 A vast class of toric forms

We will now algorithmically describe a large class of families of elliptic curves, that will be scanned for efficient arithmetic in Section 4. For sake of simplicity, all families depend on one parameter, which appears as the coefficient of a certain fixed monomial. Especially the latter property is a severe constraint. Fix an integer $d \geq 3$.

(i) First enumerate all lattice polytopes that

1. have exactly one interior lattice point (genus one);
2. have at least one vertex on the X-axis and one vertex on the Y-axis, not necessarily distinct (irreducibility);
3. are contained in $\Sigma_d = \mathrm{Conv}\{(0,0),(d,0),(0,d)\}$ (degree at most d).

This can be done fairly naively, by iteratively adjoining a vertex (note that all forms have at most six vertices due to Figure 1). The numbers of such lattice polytopes for $d = 3, \ldots, 8$ are 79, 208, 433, 650, 884, 1244.

(ii) For each such polytope Δ and each edge $\tau_b \subset \Delta$ (called the *base edge*), we label the interior lattice points of τ_b with 0, the most clockwise oriented vertex of τ_b with 1, and the most counter-clockwise oriented vertex of τ_b with -1.

(iii) For each such partially labeled pair (Δ, τ_b), complete the labeling in all possible ways in accordance with the following rules.

1. One lattice point v_C of $\Delta \setminus \tau_b$ gets the label 'A';
2. The vertices of Δ that were not labeled so far, become equipped with a '1'.
3. The lattice points of Δ that were not labeled so far, get a '0' or a '1'.

(iv) Finally, to each completely labeled pair (Δ, τ_b), associate a polynomial

$$C_A(x,y) = \sum_{(i,j) \in \Delta \cap \mathbb{Z}^2} (\text{label of } (i,j)) \cdot x^i y^j \quad \in \mathbb{Q}(A)[x,y].$$

In the spirit of Lemma 5, erase all trinomials from this list: such 'families' will define the same elliptic curve for each specialized choice of A.

All pairs (C_A, τ_b) are collected in the output set S_d. The numbers of elements of S_d for $d = 3, \ldots, 8$ are 5292, 14553, 32643, 55758, 73332, 103908. In practice we will take $d = 6$.

Lemma 7: Every $(C_A, \tau_b) \in S_d$ defines a smooth genus one curve in $X(\Delta)$ over $\mathbb{Q}(A)$, where $\Delta = \Delta(C_A)$. It contains the point $\mathcal{O} = (\alpha_{ij})_{(i,j)\in\Delta\cap\mathbb{Z}^2}$, where $\alpha_{ij} = 1$ if $(i,j) \in \tau_b$ and $\alpha_{ij} = 0$ if $(i,j) \notin \tau_b$.

Proof. It suffices to verify the statement up to $\mathbb{Z}$-affine equivalence. Hence, since all polytopes of Figure 1 have a representative in Σ_4, it suffices to prove the statement for $d = 4$. A finite computation then shows that all $(C_A, \tau_b) \in S_4$ define a curve of geometric genus one. In particular, the curve defined by C_A can have no singularities in $\mathbb{T}^2$. Indeed, modulo a $\mathbb{Z}$-affine equivalence we may assume that the Newton polytope of C_A is contained in $\mathrm{Conv}\{(0,0),(2,0),(0,2),(2,2)\}$, in Conv$\{(0,0),(4,0),(0,2)\}$ or in Σ_3. Suppose $(x_0, y_0) \in \mathbb{T}^2$ is a singular point. Then the Newton polytope of $C(x - x_0, y - y_0)$ can have no lattice points in its interior. By Baker's inequality we run into a contradiction.

In particular, for C_A the nondegeneracy condition with respect to Δ itself is fulfilled. The nondegeneracy conditions with respect to the vertices are immediate. Verifying the nondegeneracy conditions with respect to the edges boils down to verifying the squarefreeness of polynomials of the form

$$x^k - 1 \quad \text{and} \quad \alpha_4 x^4 + \alpha_3 x^3 + \alpha_2 x^2 + \alpha_1 x + 1$$

with the $\alpha_i \in \{0, 1, A\}$ (at most one α_i equalling A). In the former case, this is immediate. In the latter case, if $\alpha_4 = 0$ then a finite computation proves squarefreeness. This proves full nondegeneracy as soon as Δ is not of type (xiii) in Figure 1, and the lemma follows in that case.

It remains to deal with the subtle case where Δ is of type (xiii). Then C_A may accidentally *fail* to be nondegenerate: $C_A(x, y) = -y^2 + Ay + x^4 + x^3 + x + 1$ is an example. However, again following a reasoning using Baker's inequality one can prove that the nondegeneracy failure can only be due to tangency with toric infinity. In particular, the curve defined by $C_A(x, y) = 0$ still embeds smoothly in $X(\Delta)$. ■

The point $\mathcal{O}$ will be called the *base point* of (C_A, τ_b), and when speaking about arithmetic on (C_A, τ_b) it will always be with respect to this base point. In Section 6, we will report on an exhaustive scan of all $(C_A, \tau_b) \in S_6$ for efficient arithmetic over $\mathbb{Q}(A)$. Note that doubling and/or addition formulas over $\mathbb{Q}(A)$ suit for arithmetic over any finite field k of sufficiently large characteristic, for almost all specializations of A in k.

Of course, the finite class S_d has its limitations. But we tried to make a choice that is both practical and natural. We remark that:

- The methods for finding efficient arithmetic, explained in Section 4, do not depend on the particular construction of S_d.
- In the search for efficient arithmetic, it is a priori sufficient to consider 1-parameter families only. If a family depending on 2 parameters has some remarkable arithmetical properties, then specializing one of the parameters will result in a 1-parameter family having the same remarkable arithmetical properties.
- The fact that all constants are '0', '1' or '-1' is less restrictive than it seems at first sight. The efficiency of doubling and/or addition formulas is hardly affected by substitutions of the type $x \leftarrow \alpha x$, $y \leftarrow \beta y$, for small $\alpha, \beta \in k$.

  Conversely, as in the proof of Lemma 5, up to three non-zero coefficients (whose corresponding exponent vectors are not collinear) can always be transformed to '1' for some suitable choice of $\alpha, \beta \in \overline{k}$. In general however, this might involve the introduction of large constants that are defined over an extension field only.

With these remarks in mind, S_d essentially contains all our working examples (if $d \geq 4$).

Form	$C_A(x,y)$	τ_b	Fig. 1
(1) Weierstrass (with $B = 1$)	$-y^2 + x^3 + Ax + 1$	$[<3,0>, <0,2>]$	(vii)
(2) Hessian (modulo $y \leftarrow -y$)	$x^3 - y^3 + 1 + Axy$	$[<3,0>, <0,3>]$	(xvi)
(3) Edwards	$x^2 + y^2 - 1 + Ax^2y^2$	$[<0,2>, <0,0>]$	(xv)
(4) Jacobi quartic	$-y^2 + x^4 + Ax^2 + 1$	$[<4,0>, <0,2>]$	(xiii)
(5) Montgomery (with $B = 1$)	$-y^2 + x^3 + Ax^2 + x$	$[<3,0>, <0,2>]$	(v)

Of course, we also indirectly cover the doubly parameterized twisted Edwards curves [4] and twisted Hessian curves [5]. Thus, despite its apparent narrowness, S_d contains most of the prominent known forms whose arithmetical properties have been studied in the literature so far (over fields of large characteristic). The Doche/Icart/Kohel forms [17] are the most important absentees.

3.5 Efficient preliminary arithmetic on toric forms

The main prerequisite of the algorithm described in Section 4 is a relatively efficient method to perform arithmetic on the above toric forms. Very general algorithms based on Riemann-Roch computations are currently too slow for this purpose. Let (C_A, τ_b) be one of the above forms. Then our method consists of, using an appropriate $\mathbb{Z}$-affine map, transforming the curve to either an intersection of quadrics in $\mathbb{P}^3$ (corresponding to the cases (xiii) and (xv) of Figure 1), or a plane cubic in $\mathbb{P}^2$ (corresponding to the other cases), and perform arithmetic there. This can be done very quickly: in the plane cubic case tangent-chord applies, whereas in the quadric intersection case a projection from the base point takes us to the plane cubic case [22, Example 18.16]. We go into more detail for two exemplary situations:

Example. $C_A(x,y) = A + y + x^2y - x^2y^2, \tau_b = [< 2,1 >, < 2,2 >]$

Let $\mathcal{P}$ be a $\mathbb{T}^2$-point of the curve defined by $C(x,y)$, and let n be a positive integer. Suppose we wish to compute the sequence $\mathcal{P}$, $[2]\mathcal{P}$, $[4]\mathcal{P}$, $\ldots$, $[2^n]\mathcal{P}$, in order to interpolate doubling formulas (see Section 4). Note that $\Delta(C_A)$ has 4 lattice points $v_1, \ldots, v_4$ (enumerated counterclockwise) on the boundary, all of which are vertices. Together with the relation $v_1 + v_3 = v_2 + v_4$, this implies that $\Delta(C_A)$ is of type (iv). The transformation

$$f : \mathbb{R}^2 \to \mathbb{R}^2 : \begin{bmatrix} X \\ Y \end{bmatrix} \mapsto \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

is a corresponding $\mathbb{Z}$-affine map, taking $C_A(x,y)$ to $C'_A(x,y) = Ax + y + x^2y - xy^2$. The point $\mathcal{P} = (a,b)$ is sent to

$$\mathcal{P}' = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} a \\ b \end{bmatrix}.$$

The base edge becomes $[< 2,1 >, < 1,2 >]$ and, following the explicit description of λ given in Section 3.2, this corresponds to taking $\mathcal{O}' = (1,1,0)$ as neutral element. Now using tangent-chord arithmetic, one can compute $\mathcal{P}', [2]\mathcal{P}', [4]\mathcal{P}', \ldots, [2^n]\mathcal{P}'$. Transforming back gives the requested answer.

Example. $C_A(x,y) = x^4 - x^4y^2 + x^3y^2 + y^2 + Ax^3y, \tau_b = [< 4,0 >, < 4,2 >]$

Let $\mathcal{P}$ be a $\mathbb{T}^2$-point of the curve defined by $C_A(x,y)$, and let n be a positive integer. Suppose we wish to compute the sequence $\mathcal{P}$, $[2]\mathcal{P}$, $[3]\mathcal{P}$, $\ldots$, $[n]\mathcal{P}$, in order to interpolate addition formulas (see Section 4). Note that $\Delta(C_A)$ has an edge containing 5 lattice points, hence it is of type (xiii). For this type, we use the representative $\mathrm{Conv}\{(0,0),(4,0),(0,2)\}$, the Newton polytope of a Jacobi quartic form. Then under an appropriate $\mathbb{Z}$-affine transformation, our polynomial is sent to $C'_A(x,y) = y^2 - 1 + Axy + x + x^4$. Note that $\Delta(C'_A) = 2\Gamma$ for the smaller triangle $\Gamma = \mathrm{Conv}\{(0,0),(2,0),(1,0)\}$: then the non-singular model of $C'_A(x,y)$ can be realized by a quadratic relation

$$z_{10}^2 - z_{00}^2 + Az_{10}z_{01} + z_{00}z_{10} + z_{20}^2$$

in $X(\Gamma) : z_{10}^2 - z_{00}z_{20}$ in $\mathbb{P}^3$. The desingularization map λ can again be described explicitly, which allows one to trace back the neutral element $\mathcal{O}'$ and the place of interest $\mathcal{P}'$. Now projecting from $\mathcal{O}$, one obtains a plane cubic in which it is possible to use tangent-chord arithmetic. As such, one can quickly compute $\mathcal{P}', [2]\mathcal{P}', [3]\mathcal{P}', \ldots, [n]\mathcal{P}'$. Transforming back gives the requested answer.

4 Efficient formulas via lattice reduction

In Section 3.4 we computed a vast list of families of elliptic curves of the form (C_A, σ_b). For each (C_A, σ_b) there exist many doubling and addition formulas, but we are typically only interested in the most efficient ones in affine or projective coordinates, or in efficient generalized Montgomery arithmetic or in efficient uniform addition formulas. Furthermore, we require that these formulas are valid over $\mathbb{Q}$ such that they can be used for all finite fields of large enough characteristic. In this section, we describe an extremely simple but very powerful algorithm that is capable of computing these formulas in parameterized form (i.e. depending on the parameter A) and also automatically selects the most efficient ones.

To simplify the exposition, we first describe the algorithm for computing efficient doubling formulas in affine coordinates. Other applications of this strategy will be discussed in Section 5. Given a family of curves (C_A, σ_b) defined over $\mathbb{Q}$, we need to compute a quartet of nonzero polynomials $f_1, g_1, f_2, g_2 \in \mathbb{Q}[A][x, y]$ such that

$$x \circ \varphi_2 = \frac{f_1}{g_1}, \qquad y \circ \varphi_2 = \frac{f_2}{g_2}$$

inside the function field $\mathbb{Q}(\widetilde{C}_A)$. To do this, we first select a support set S of monomials in A, x, y that are allowed to appear in the f_i, g_i. Note that the parameter A could appear non-linearly in the support set S. The polynomials f_i, g_i can then be written as $\mathbb{Q}$-linear combinations of the monomials in S, i.e.

$$x \circ \varphi_2 = \frac{f_1}{g_1} = \frac{\sum_{m_i \in S} f_{1,i} \cdot m_i}{\sum_{m_i \in S} g_{1,i} \cdot m_i}, \qquad y \circ \varphi_2 = \frac{f_2}{g_2} = \frac{\sum_{m_i \in S} f_{2,i} \cdot m_i}{\sum_{m_i \in S} g_{2,i} \cdot m_i}.$$

To obtain a description of all possible doubling laws, we would like to use an evaluation strategy to compute a linear system of equations in the unknown coefficients. For this we need to find a point P on C_A and need to be able to compute $P, [2]P, \cdots, [2^n]P$. Even if we specialize the family C_A in a value $\bar{A} \in \mathbb{Q}$, we still need to find a rational point on the elliptic curve $C_{\bar{A}}$ over $\mathbb{Q}$ which is known to be hard. To solve this and other related problems coming from working over $\mathbb{Q}$, we reduce the whole setup modulo a large prime p and work over the finite field $\mathbb{F}_p$.

Therefore, choose a large prime p and choose a large random $\bar{A} \in \mathbb{F}_p$ to obtain the curve $\bar{C}_{\bar{A}}$ over $\mathbb{F}_p$. Now it becomes trivial to find a point $\bar{P}$ on $\bar{C}_{\bar{A}}$ and using the algorithms described in Section 3.5, we obtain the sequence $\bar{P}, [2]\bar{P}, \cdots, [2^n]\bar{P}$. Let $\bar{m}_{i,j}$ denote the evaluation of the monomial m_i in $\bar{A}$ and the coordinates of $[2^j]\bar{P}$. Each tuple $([2^j]\bar{P}, [2^{j+1}]\bar{P})$ results in two linear equations

$$\begin{aligned} x([2^{j+1}]\bar{P}) \sum_{m_i \in S} g_{1,i} \cdot \bar{m}_{i,j} - \sum_{m_i \in S} f_{1,i} \cdot \bar{m}_{i,j} = 0 \\ y([2^{j+1}]\bar{P}) \sum_{m_i \in S} g_{2,i} \cdot \bar{m}_{i,j} - \sum_{m_i \in S} f_{2,i} \cdot \bar{m}_{i,j} = 0\,. \end{aligned}$$

Therefore if $n \gg 2\#S$ we obtain an overdetermined system M_x (resp. M_y) of linear equations over $\mathbb{F}_p$ such that all possible formulas for the x-coordinate (resp. y-coordinate) of the doubling for the curve $\bar{C}_{\bar{A}}$ are contained in $\mathrm{Ker}(M_x)$ (resp. $\mathrm{Ker}(M_y)$).

Two problems remain: how to find the most efficient parameterized doubling formulas in the kernel and how to lift the situation from $\mathbb{F}_p$ back to $\mathbb{Q}$. Both problems can be solved simultaneously by finding short vectors in the following lattice over $\mathbb{Z}$ spanned by the columns of

$$\left[b_1,\, b_2,\, \cdots,\, b_n,\, pI_{2|S|}\right] \,,$$

where $\{b_1, \cdots, b_n\}$ is a basis of $\mathrm{Ker}(M_x)$ (resp. $\mathrm{Ker}(M_y)$) with $b_i \in \mathbb{F}_p^{2|S|}$ and I_n denotes the $n \times n$ identity matrix. Finding short vectors indeed solves both problems: firstly, a formula with only a few monomials will lead to a shorter vector than a formula consisting of many monomials. Secondly, since $\bar{A}$ was chosen randomly and large, the lattice reduction will automatically make the correct choice between using a large coefficient in front of a monomial not involving A and using a small coefficient in front of the corresponding monomial with A included.

Example. To illustrate this behaviour, assume that in the final formula, there is monomial of the form $(A+2)xy$. Let $m_u = xy$ and $m_v = Axy$, then over the finite field $\mathbb{F}_p$ the monomial $(\bar{A}+2)xy$ can be written as any of the following linear combinations of $\bar{m}_u$ and $\bar{m}_v$, namely

$$(\bar{A}+2)xy = (\alpha+2)\bar{m}_u + (1-\alpha\bar{A}^{-1})\bar{m}_v\,, \quad \alpha \in \mathbb{F}_p\,.$$

However, it is easy to see that the shortest linear combination corresponds precisely to the choice $\alpha = 0$, since for $\alpha \neq 0$, either α or $\alpha\bar{A}^{-1}$ will be large.

If the set S contains all monomials appearing in the equation of the curve, the kernel of M_x (resp. M_y) will also contain short vectors that correspond to polynomials f_i and g_i that are zero in the function field, i.e. are multiples of the equation of the curve. Therefore, when f_1/g_1 (resp. f_2/g_2) is computed, a final verification is necessary to ensure that the formula is not one of these trivial cases.

Note that the length of the vectors b_i appearing in the lattice is $2|S|$, so when S contains many monomials, finding a short vector in the lattice becomes a major bottleneck of the algorithm. One solution to overcome this problem is to assign several different *small* values to $\bar{A}$ (since we want the corresponding vectors to be short), run the lattice reduction to obtain efficient formulas for the different curves $C_{\bar{A}}$ and then use interpolation to find expressions for the coefficients that depend on A.

As a result, we obtain parameterized negation, doubling and adding formulas for the family (C_A, σ_b) that are efficient and at the same time valid over $\mathbb{Q}$ and thus in particular, valid over any finite field of large enough characteristic.

5 Applications

5.1 Results for a new family

In this paragraph we provide a fully worked example for one family out of the many we have tested. The family corresponds to type (ii) in Figure 1 and is defined by the equation

$$C_A = Ax + x^2 - xy^2 + 1 \qquad \sigma_b = [<2,0>, <1,2>]\,.$$

Negation is simply given by $-(x,y) = (x,-y)$ and affine doubling is

$$[2](x,y) = \left(\frac{(x^2-1)^2}{(2xy)^2}, \frac{-(x^2-1)^2 + 2xy^2(x^2+1)}{2xy(x^2-1)}\right)\,.$$

Affine addition formulas are as follows:

$$(x_1,y_1) + (x_2,y_2) = \left(\frac{(x_1x_2-1)^2}{x_1x_2(y_1+y_2)^2}, \frac{x_1x_2(x_1y_2 - x_2y_1) + (x_1y_1 - x_2y_2)}{x_1x_2(y_1^2 - y_2^2)}\right)\,.$$

Note that the formula for the x-coordinate of addition is uniform, i.e. by setting $x_1 = x_2$ and $y_1 = y_2$ we obtain the x-coordinate of the double. The negation formula implies that $k(x)$ is the index 2 invariant subfield, so x-only arithmetic is possible. The resulting formulas are:

$$x_{2n} = \frac{(x_n^2-1)^2}{4x_n(x_n^2 + Ax_n + 1)} \qquad x_{m-n}x_{m+n} = \frac{(x_mx_n - 1)^2}{(x_m - x_n)^2}$$

5.2 Generalized Montgomery arithmetic

To provide a non-trivial example of generalized Montgomery arithmetic, we revisit the Jacobi quartic. In this case, the invariant index 2 subfield is generated by $t = (y+1)/x^2$, and t-only doubling and addition formulas are:

$$t_{2n} = \frac{t_n^4 - (2t_n - A)^2 + 2t_n^2 + 1}{(2t_n^2+2)(2t_n - A)}$$

$$t_{m+n}t_{m-n} = \frac{(t_mt_n - 1)^2 + 2A(t_m + t_n) - A^2}{(t_m - t_n)^2}.$$

5.3 Uniform addition formulas

Recall that addition formulas are called uniform if they can also be used for doubling. The algorithm in Section 4 can be easily adapted to return uniform addition formulas by generating half of the total number of linear equations using $\psi(P,Q)$ with $P \neq Q$ and half of them using $\psi(P,P)$. As such, the resulting addition formulas will automatically be uniform.

To illustrate this approach, we give uniform addition formulas for the Weierstrass curve $y^2 - x^3 - Ax - B$ over a field k of characteristic > 3, where as usual $\mathcal{O}$ is the point at infinity:

$$x \circ \psi = \frac{(x_1x_2 - 2A)x_1x_2 - 4B(x_1+x_2) + A^2}{(x_1x_2 + A)(x_1+x_2) + 2y_1y_2 + 2B}$$

$$y \circ \psi = \frac{x_1x_2(x_1+x_2) - x_3\left((x_1+x_2)^2 - x_1x_2 + A\right) - y_1y_2 - B}{y_1+y_2}$$

here x_3 abbreviates $x \circ \psi$. Note that here we used the technique described at the end of Section 4, i.e. we first derived the above formulas for different small values of A and B using lattice reduction, and then used interpolation to recover the coefficients that depend on A and B. We remark that similar uniform Weierstrass addition formulas were already devised by hand [8, Corollary 1].

6 Quasi-optimality of Edwards doubling

Using the equation of the curve, the doubling law on an Edwards curve $C(x,y) = x^2 + y^2 - 1 - dx^2y^2$ can be rewritten as (see Appendix A)

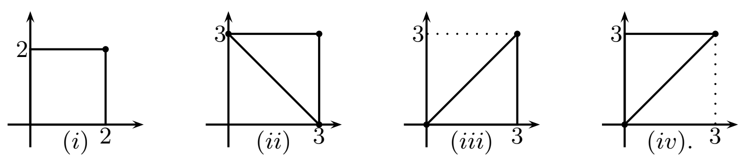
$$(x,y) \mapsto \left(\frac{2xy}{x^2+y^2}, \frac{(y^2-x^2)}{2-(x^2+y^2)} \right).$$

Thus Edwards curves allow for doubling formulas consisting of *quadratic polynomials*, which is an attractive property putting Edwards curves among the most efficient known-to-date models for point doubling in characteristic $\neq 2$. One can immediately deduce more families of curves having this property: translates $C(x-x_0, y-y_0)$ for $x_0, y_0 \in k$, flips of the type $y^2C(x,y^{-1})$ or $x^2C(x^{-1},y)$, and combinations of these. However, using the above machinery, we computationally proved:

Proposition 8: S_6 does not contain any non-Edwards-related families having quadratic doubling formulas.

The following proposition prudently suggests that this is not a coincidence.

Proposition 9: Let k be a field and let $C(x,y) \in k[x,y]$ be a nondegenerate polynomial defining a curve E of geometric genus one. Let $\mathcal{O}$ be a k-rational place of E, and suppose there are polynomials $f_1, g_1, f_2, g_2 \in k[x,y]$ of degree at most two defining a set of doubling formulas on $(E,\mathcal{O})$. Then the Newton polytope of $C(x,y)$ is contained in one of the following:



Moreover, in each case the bold-marked lattice points appear as vertices.

Proof. Write Δ for the Newton polytope of $C(x,y)$. By the irreducibility of $C(x,y)$, Δ has at least one vertex on the X-axis and at least one vertex on the Y-axis. Inheriting the notation of Section 2.1, we will denote the doubling morphism on $(E,\mathcal{O})$ by φ_2. This proof only makes use of the fact that φ_2 is a degree 4 morphism. In particular, $\mathcal{O}$ plays no role. We will write E' for $E \cap \mathbb{A}^2_k$.

We first prove that there is an $m \in \mathbb{Z}_{\geq 2}$ for which $(m;m)$ appears as a vertex of Δ, such that Δ is in its turn contained in

$$\operatorname{Conv}\{(0;0),(m;0),(0;m),(m;m)\}.$$

This property, as well as the property of having quadratic doubling formulas, is invariant under replacing C by $C(x-x_0, y-y_0)$ for any $x_0, y_0 \in \overline{k}$. The replacement might spoil the property of nondegeneracy, but we will not use this. We may therefore assume that E' satisfies the following generic conditions:

1. the x-axis intersects E' in $\deg_y C$ points (counting multiplicities),
2. the y-axis intersects E' in $\deg_x C$ points (counting multiplicities),
3. none of the coordinate axes contains singular points of E', or regular points $\mathcal{P} \in E'$ for which there is a place $\mathcal{Q}$ above a singularity or at infinity such that $\varphi_2(\mathcal{Q}) = \mathcal{P}$.

The first assumption implies that $\operatorname{div}(x) = D_0 - D_\infty$ for effective degree $\deg_y C$ divisors D_0 and D_∞ that are supported on E' and $E \setminus E'$ respectively. Then

$$\operatorname{div}(x \circ \varphi_2) = \varphi_2^* D_0 \ - \ \varphi_2^* D_\infty$$

is the difference of two effective degree $4\deg_y C$ divisors with disjoint support, so in particular $x \circ \varphi_2$ has $4\deg_y C$ zeroes. By our third assumption these are all in the affine part E' of E, hence they must be realized as zeroes of f_1. By Bezout's theorem, this number is bounded by $2\deg C$ and we conclude $2\deg_y C \leq \deg C$. Similarly, $2\deg_x C \leq \deg C$. When combined, these inequalities are seen to become equalities, and the statement follows with $m = \deg_x C = \deg_y C$. Since E is of geometric genus one, it is immediate that $m \geq 2$.

Now if $m = 2$ then we are in situation *(i)*. Therefore suppose $m \geq 3$. As mentioned above, Δ contains at least one point of the form $(a;0)$ and one point of the form $(0;b)$ (not necessarily distinct). Suppose $a, b \neq 0$, then a non-empty part of the line segment connecting $(0;0)$ and $(m;m)$ is contained in the interior of $\Delta(f)$. This part contains two or more lattice points unless $m = a = b = 3$, hence we are in situation (ii). Suppose on the other hand that Δ contains $(0;0)$. Since the line segment connecting $(0;0)$ and $(m;m)$ contains $m-1 \geq 2$ interior lattice points, it must be an edge. From Figure 1, we see that the maximal number of lattice points in the interior of an edge is 2. Therefore, $m = 3$ and we are in situation (iii) or (iv). ■

We were not able to eliminate the cases $(ii) - (iv)$, neither could we construct quadratic doubling laws corresponding to such a Newton polytope. Apart from that, it *is* possible to obtain alternative quadratic formulas if one allows the y-coordinate to depend on $x_3 = x \circ \varphi_2$. For example let $C_A(x,y) = A + x^3y - x^2y^2 + x^2$ and $\tau_b = [<3,1>, <2,2>]$, then

$$x_3 = x \circ \varphi_2 = \frac{-x^2 - 4xy + 4y^2 - 4}{2x - 4y}, \quad y \circ \varphi_2 = \frac{x^2 + xx_3 - 2yx_3}{2x_3}.$$

These and many similar formulas were computed using the methods described in Section 4.

7 Quasi-optimality of Montgomery arithmetic

A crucial observation in t-only arithmetic is that the specific form of the elliptic curve is actual of little importance. Indeed, let $(E', \mathcal{O}')$ be isomorphic to $(E, \mathcal{O})$. Then there is an isomorphism of function fields $\psi : k(\widetilde{E}) \to k(\widetilde{E}')$ such that $k(\psi(t))$ is exactly the set of functions $f \in k(\widetilde{E}')$ that satisfy $f = f \circ \chi'$ (where χ' is the negation morphism on $\widetilde{E}'$). Therefore, the $\psi(t)$-only doubling and addition formulas on $(E', \mathcal{O}')$ are exact copies of the t-only doubling and addition formulas on $(E, \mathcal{O})$.

What does matter, is the choice of the transcendental generator t. Every function generating $k(t)$ corresponds to an automorphism of $\mathbb{P}^1$ and is of the form

$$t' = \frac{at+b}{ct+d}, \quad a,b,c,d \in k, \quad ad-bc \neq 0, \quad \text{or conversely,} \quad t = \frac{dt'-b}{-ct'+a}.$$

With $x \leftrightarrow t$ and $x' \leftrightarrow t'$, one can verify that on the Montgomery curve (see Appendix A)

$$x' \circ \varphi_2 = \frac{\alpha_4 x'^4 + \alpha_3 x'^3 + \alpha_2 x'^2 + \alpha_1 x' + \alpha_0}{\beta_4 x'^4 + \beta_3 x'^3 + \beta_2 x'^2 + \beta_1 x' + \beta_0}, \tag{4}$$

where the α_i, β_i are long (but manageable) polynomial expressions in a,b,c,d and the curve parameter A. A Gröbner basis computation shows that the ideal generated by α_4 and β_4 contains $(c-d)^2(c+d)^2(ad-bc)$. If $c = d \neq 0$ then β_4 can only vanish if $A = 1$, and if $c = -d \neq 0$ then β_4 can only vanish if $A = 0$. Hence for generic A, it is impossible to let α_4 and β_4 vanish at the same time. Similarly, it is impossible that both α_0 and β_0 vanish and to get rid of the curve parameter A. An interesting corollary is the following (somewhat loosely stated) proposition:

Proposition 10: Let $\operatorname{char} k \neq 2$. On a randomly chosen elliptic curve in any nontrivial family over k, it is impossible to do t-only doubling using projective coordinates (t, z) in less than five field multiplications (which include squarings and multiplications with curve constants).

Proof. By the above discussion, $t \circ \varphi_2$ must be of the form (4), with $x' \leftrightarrow t$ (note that, although $t \circ \varphi_2 \in k(t)$, it might be possible that a,b,c,d live in an extension field, since an isomorphism with a Montgomery curve might not exist over k). Now $\alpha_4 \neq 0$ or $\beta_4 \neq 0$ and $\alpha_0 \neq 0$ or $\beta_0 \neq 0$, so one at least has to compute a term in x^4 and in t^4, already accounting for 4 multiplications. Since the curve is randomly chosen, the constant A will account for at least one additional multiplication. ■

Montgomery's doubling formula attains this bound, so in this sense it is optimal. But note that the above statement does not make a distinction between ordinary field multiplications, field squarings (considerably faster), and multiplications with curve constants (often chosen small, often to be multiplied with multiple times). And indeed, from a practical point of view, there is room for improvement over Montgomery arithmetic. For small curve parameters, the current way to go seems due to Gaudry and Lubicz [19, Section 6.2].

8 Conclusion

In this paper we introduced a theoretical framework based on toric geometry to study different forms of elliptic curves. Using this framework, we obtained a very large class of 50000 different elliptic curve forms. To compute the most compact group law for each of these forms, we described a simple but very powerful algorithm based on interpolation and lattice reduction and illustrated its use by computing addition/doubling laws, generalized Montgomery formulas and uniform group laws.

References

[1] V. Batyrev, *Variations of the mixed Hodge structure of affine hypersurfaces in algebraic tori*, Duke Math. J. **69**(2), pp. 349-409 (1993)

[2] P. Beelen, R. Pellikaan, *The Newton polygon of plane curves with many rational points*, Designs, Codes and Crypt. **21** no. 1-3, pp. 41-67 (2000)

[3] D. Bernstein, *Curve25519: new Diffie-Hellman speed records*, proceedings of PKC 2006, Lecture Notes in Computer Science **3958**, pp. 207-228 (2006)

[4] D. Bernstein, P. Birkner, M. Joye, T. Lange, C. Peters, *Twisted Edwards curves*, proceedings of AFRICACRYPT 2008, Lecture Notes in Computer Science **5023**, pp. 389-405 (2008)

[5] D. Bernstein, T. Lange, *Explicit Formulas Database*, `http://www.hyperelliptic.org/EFD/`

[6] D. Bernstein, T. Lange, *Faster addition and doubling on elliptic curves*, proc. ASIACRYPT 2007, Lecture Notes in Computer Science **4833**, pp. 29-50 (2007)

[7] D. Bernstein, T. Lange, R. Rezaeian Farashahi, *Binary Edwards curves*, proceedings of CHES 2008, Lecture Notes in Computer Science **5154**, pp. 244-265 (2008)

[8] É. Brier, M. Joye, *Weierstrass elliptic curves and side-channel attacks*, proceedings of PKC 2002, Lecture Notes in Computer Science **2274**, pp. 335-345 (2002)

[9] O. Billet, M. Joye, *The Jacobi model of an elliptic curve and side-channel analysis*, proceedings of AAECC-15, Lecture Notes in Computer Science **2643**, pp. 34-42 (2003)

[10] W. Castryck, *Forms of elliptic curves*, talk at BCRYPT ECC Day, 2008/03/20, Leuven (Belgium)

[11] W. Castryck, J. Denef, F. Vercauteren, *Computing zeta functions of nondegenerate curves*, Int. Math. Res. Papers, Vol. 2006, Article ID 72017 (2006)

[12] W. Castryck, J. Voight, *On nondegeneracy of curves*, Algebra & Number Theory **3**(3), pp. 255 - 281 (2009)

[13] W. Castryck, H. Hubrechts, *The distribution of the number of points modulo an integer on elliptic curves over finite fields*, preprint (2009)

[14] D. Chudnovsky, G. Chudnovsky, *Sequences of numbers generated by addition in formal groups and new primality and factorization tests*, Advances in Applied Mathematics **7**(4), pp. 385-434 (1986)

[15] H. Cohen, G. Frey (eds.), R. Avanzi, C. Doche, T. Lange, K. Nguyen, F. Vercauteren, *Handbook of elliptic and hyperelliptic curve cryptography*, Discrete Math. and its Appl. **34**, 848 pp. (2005)

[16] H. Cohen, A. Miyaji, T. Ono, *Efficient elliptic curve exponentiation using mixed coordinates*, proceedings of ASIACRYPT 1998, Lecture Notes in Computer Science **1514**, pp. 51-65 (1998)

[17] C. Doche, T. Icart, D. Kohel, *Efficient scalar multiplication by isogeny decompositions*, proceedings of PKC 2006, Lecture Notes in Computer Science **3958**, pp. 191206 (2006)

[18] W. Fulton, *Introduction to toric varieties*, Annals of Math. Studies **131**, William H. Roever Lectures in Geometry, Princeton University Press (1993)

[19] P. Gaudry, D. Lubicz, *The arithmetic of characteristic 2 Kummer surfaces and of elliptic Kummer lines*, Finite Fields and Their Applications **15**, pp. 246-260 (2009)

[20] J. Hirschfeld, *Projective geometries over finite fields*, Oxford Math. Monographs, Oxford Univ. Press, New York, xiv + 555 pp. (1988)

[21] M. Joye, J.-J. Quisquater, *Hessian elliptic curves and side-channel attacks*, proceedings of CHES 2001, Lecture Notes in Computer Science **2162**, pp. 402-410 (2001)

[22] J. Harris, *Algebraic geometry: a first course*, Graduate Texts in Mathematics **133**, Springer, xix + 328 pp. (1992)

[23] H. Hisil, G. Carter, E. Dawson, *New formulae for efficient elliptic curve arithmetic*, proceedings of INDOCRYPT 2007, Lecture Notes in Computer Science **4859**, pp. 138-151 (2007)

[24] H. Hisil, K. Koon-Ho Wong, G. Carter, E. Dawson, *Twisted Edwards curves revisited*, proceedings of ASIACRYPT 2008, Lecture Notes in Computer Science **5350**, pp. 326-343 (2008)

[25] N. Koblitz, *Elliptic curve cryptosystems*, Mathematics of Computation **48**, pp. 203209 (1987)

[26] R. Koelman, *A criterion for the ideal of a projectively embedded toric surface to be generated by quadrics*, Beitrage zur Algebra und Geometrie **34**, pp. 57-62 (1993).

[27] P.-Y. Liardet, N. Smart, *Preventing SPA/DPA in ECC systems using the Jacobi form*, proceedings of CHES 2001, Lecture Notes in Computer Science **2162**, pp. 391401 (2001)

[28] H. Lenstra, *Factoring integers with elliptic curves*, Annals of Math. **126** (2), pp. 649-673 (1987)

[29] V. Miller, *Use of elliptic curves in cryptography*, proceedings of CRYPTO '85, Lecture Notes in Computer Science **218**, pp. 417-426 (1986)

[30] M. Monagan, R. Pearce, *Rational simplicifation modulo a polynomial ideal*, proceedings of ISSAC 06, ACM press, pp. 239-245 (2006)

[31] P. Montgomery, *Speeding the Pollard and elliptic curve methods of factorization*, Math. Comp. **48** (177), pp. 243-264 (1987)

[32] K. Okeya, H. Kurumatani, K. Sakurai, *Elliptic curves with the Montgomery-form and their cryptographic applications*, proceedings of PKC 2004, Lecture Notes in Computer Science **1751**, pp. 238-257 (2004)

[33] B. Poonen, F. Rodriguez-Villegas, *Lattice polygons and the number* 12, Amer. Math. Monthly **107** (3), pp. 238-250 (2000)

[34] R. Rezaeian Farashahi, I. Shparlinski, *On the number of distinct elliptic curves in some families*, Designs, Codes and Cryptography, online publication (2009)

[35] N. Smart, *The Hessian form of an elliptic curve*, proceedings of CHES 2001, Lecture Notes in Computer Science **2162**, pp. 118-125 (2001)

[36] M. Tsfasman, S. Vlăduţ, D. Nogin, *Algebraic geometric codes: basic notions*, Math. Surveys and Monographs 139, Am. Math. Soc., Providence, xx + 338 pp. (2007)