Three Wise Men: Models and Maps in Modern ECC

Joost Renes

NXP Semiconductors, Eindhoven, The Netherlands

September 11, 2021

Abstract

This text was written for the isogeny-based cryptography school (week 10) hosted by the University of Bristol in July–September 2021. The goal of this lecture is to familiarize the readers with the most popular (i.e., Weierstrass, Montgomery and (twisted) Edwards) elliptic-curve models used in cryptography. By the end the reader should understand the choices made in elliptic-curve cryptography, both in the past (including discrete-log-based crypto) and for the current post-quantum proposals CSIDH and SIKE. This is by no means a complete overview of the subject–References are included that contain further reading material.

1 Introduction

Recall that elliptic-curve cryptography is classically based on the *discrete logarithm problem*. That is, for a (large[1]) prime $p > 3$, an elliptic curve $E/\mathbb{F}_p$ and a point $P \in E(\mathbb{F}_p)$ of (large enough) prime order, for any $Q \in \langle P \rangle$ it is difficult to find $\lambda \in [0, \#\langle P \rangle - 1]$ such that $Q = [\lambda]P$. As a result, the most important arithmetic operation is *scalar multiplication*

$$\lambda : P \mapsto [\lambda]P\,.$$

This is an especially simple to describe operation, yet much has been written about how to perform it. There are several reasons for this: firstly, the scalar λ typically corresponds to a *secret key*, meaning that the scalar multiplication has to be performed as independently from λ as possible (e.g., no time or power dependence, which will be discussed in upcoming lectures). Secondly, there are many ways to implement scalar multiplication (or exponentiation). For example, there are left-to-right and right-to-left bit-scanning techniques, or the clever Montgomery ladder algorithm [10]. However,

[1] We always assume that finite field characteristics are cryptographically large (say $>$ 256-bit) primes.

these two choices are mostly independent of the underlying group in which the multiplication is performed.

We shall rather concern ourselves with yet another dimension: the choice of representation of the elliptic curve. While E, and the operations on it required for cryptography, can be described completely independently from its embedding into projective space, the embedding becomes crucial once we actually want to compute a scalar multiplication. That is, to even store a point P on a computer we cannot treat it as a purely abstract element, but have to decide on a representation. Many introductory texts on elliptic curves choose the (short) *Weierstrass* model $y^2 = x^3 + ax + b$, which applies to any elliptic curve, that allows representing $P = (x, y) \in \mathbb{A}^2$ as two elements in $\mathbb{F}_p$ (or three in $\mathbb{P}^2$). However, this choice is not a canonical one and there exist many choices of models (i.e., embeddings into projective space). The natural question is therefore: which representation is *best*? To look ahead: in this lecture we shall not answer this question, as there are as many opinions as there are embeddings (in particular depending on your definition of *good* and *best*). Instead, we introduce several choices that appear in cryptography nowadays and leave it to the reader to form their own opinion.

Of course, this is the *isogeny* school and not the *scalar multiplication* school. Analogous statements hold for isogeny-based algorithms: not only is scalar multiplication an important part of such protocols, the isogeny computations themselves are also strongly dependent on the choice of model.

Exercise 1. The choice of representation not only impacts the computations, but also the cryptographic key sizes. Identify for elliptic-curve Diffie–Hellman, CSIDH and SIKE which properties of models might benefit the public-key sizes.

In this lecture we focus on the Weierstrass, Montgomery and (twisted) Edwards models. It is of course worth noting that many of the formulas that we describe are the culmination of many years of development by various authors. Moreover, many more interesting models have been proposed over the years but in the interest of space we do not treat them in much detail here. Examples include the twisted Hessian [3], Huff [9] and the Jacobi quartic models. As there is much literature on elliptic-curve models and their efficiency for cryptography, we do not attempt to summarize it all in this lecture. Rather we try to create an intuition on what differentiates the Weierstrass, Montgomery and (twisted) Edwards curves and, just as importantly, what connects them. This should set up the reader to start working out formulas for themselves and see if they can improve on the state-of-the-art.

Remark 1. As the treatment of elliptic-curve formulas is typically very computational (i.e., tedious), it can be helpful to make use of computer algebra package. Typical choices are `magma` [6] or `sage` [13].

2 Group Operations

In this section we first focus on operations that are performed on a single curve, i.e., in the group of points. Although we ultimately want to compute isogenies to other curves, group operations still play an important role. For example, recall that computing 2^e-isogenies for large e in SIKE involves many group doublings. Similarly, any $p_1^{e_1} \cdots p_k^{e_k}$-isogeny for distinct primes p_i and small integers e_i for $1 \leq i \leq k$ involves (scalar) multiplications by the various p_i. As the group arithmetic was chronologically first introduced into cryptography, we do the same here. Moreover, a lot of work has been done over the last decades to make the group arithmetic as efficient as possible. This will help build intuition about what it means to be efficient in the first place, and the strategies that have been used to improve this.

2.1 Weierstrass form

As mentioned in the introduction, we no longer want to think of an elliptic curve abstractly, but rather with an explicit embedding into affine or projective space. From this point onwards we write $E/\mathbb{F}_p : y^2 = x^3 + ax + b \subset \mathbb{A}^2$ for an elliptic curve in *Weierstrass form* defined over $\mathbb{F}_p$, with neutral element at infinity (using the homogeneous embedding into $\mathbb{P}^2$). Through an elementary geometric description one shows that for affine points $P = (s,t)$ and $Q = (u,v)$ such that $P \neq \pm Q$, then $R = (w,z)$ is given by

$$w = \lambda^2 - s - u\,, \quad z = -\lambda w - \mu\,,$$

where $\lambda = (v-t)/(u-s)$ and $\mu = (tu - sv)/(u-s)$. Having written down this group law, we can analyze its cost. For this we tyically distinguish finite field operations:

- Field *addition/subtraction*, denoted **A**;
- Field *multiplication*, denoted **M**;
- Field *inversion*, denoted **I**.

Exercise 2. For a prime of 256 bits and a platform with 32-bit words, argue why the cost of field multiplications is higher than that of additions or subtractions. (Note: a 32-bit platform represents 256-bit integers as $n = \sum_{i=0}^{7} n_i 2^{32i}$ where $n_i \in [0, 2^{32}-1]$.)

Exercise 3. Show that the Weierstrass group law above for affine points $P \neq \pm Q$ can be computed with fewer than $6\mathbf{M} + 6\mathbf{A} + \mathbf{I}$ operations.

From Exercise 2 we learned that the cost of multiplication can be considered to be much higher than additions or subtractions. On the other hand, the cost of inversion is even much higher still. One of the simplest ways to implement this is by using the fact that the multiplicative subgroup of $\mathbb{F}_p$ is cyclic of order $p-1$, and hence that $x^{p-2} \equiv x^{-1} \bmod p$ for any $x \in \mathbb{F}_p$. The inversion can therefore be computed by an exponentiation with $p-2$. Although there are much more involved algorithms for inversions (e.g., similar to gcd computations), the exponentiation has the advantage of having runtime independent of the base element. As the base elements can be related to the secret key, this is a very desirable property.

Exercise 4. Argue that for large primes the cost of a field inversion is higher than that of a multiplication.

Exercise 5. Using the homogeneous projective embedding $x = X/Z$ and $y = Y/Z$ into $\mathbb{P}^2$, show that the group law can be implemented without inversions in projective space. How many multiplications are required? Do you expect this to be more efficient than using affine formulas?

Exercise 6. Instead of the homogeneous embedding into projective space, one could also use the *Jacobian* coordinate embedding $x = X/Z^2$ and $y = Y/Z^2$. How many multiplications are required for the group operation in this case?

Remark 2. The exercises above argue that $\mathbf{I} > \mathbf{M} > \mathbf{A}$. Although this is generally true on a given platform, the actual ratio between them depends strongly on the platform. We refer to the next lecture by Daniel Bernstein for several integer multiplication methods and their associated costs. It is often possible to save a multiplication by spending a number of extra additions: whether this is worth it, depends on the implementation. Moreover, it is for example possible to further distinguish multiplications into generic multiplications and squarings, where the latter can often be implemented more cheaply.

2.2 Montgomery form

We now move towards the *Montgomery* form. First consider the curve in *long* Weierstrass form

$$E_0/\mathbb{F}_p : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6\,.$$

We show that a Montgomery curve can be defined as an elliptic curve with the additional requirement that $Q = (0,0)$ is a point of order two whose tangent line is defined by $x = 0$.

Exercise 7. Assume that $Q = (0,0) \in E_0$. Show that $a_6 = 0$.

Exercise 8. Assume that $Q = (0,0)$ has order 2. Show that $a_3 = 0$.

Exercise 9. Assume that $x = 0$ is the tangent line to $Q = (0,0)$. Show that $a_4 \neq 0$.

Exercise 10. Transform E_0 into $E_1 : y^2 = x^3 + b_2x^2 + b_4x$ for $b_2, b_4 \in \mathbb{F}_p$ with $b_4 \neq 0$.

Exercise 11. Transform E_1 into $M : By^2 = x^3 + Ax^2 + x$ for $A, B \in \overline{\mathbb{F}}_p$ with $B \neq 0$.

The form in Exercise 11 is what we call the *Montgomery model*, named after Peter Montgomery who first introduced it [11]. The definition of the Montgomery model additionally assumes that $A, B \in \mathbb{F}_p$, which is not necessarily true in Exercise 11. More precisely, E_0 might only admit a Montgomery model over a *quadratic* extension of $\mathbb{F}_p$. Note that there exist similar models where Q is a point of higher order, which leads to the *Tate Normal Form*. These models are at the core of *radical isogenies*, presented also this week by Fre Vercauteren.

At this point the Montgomery form might not look particularly different from the short Weierstrass form. However, the action of Q on x-coordinates is especially nice. We shall see that this strongly influences the simplicity of the isogeny formulas (see Exercise 27).

Exercise 12. Let $P = (s, t) \in M(\mathbb{F}_p)$ and $Q = (0, 0)$. Show that $x_{P+Q} = 1/s$.

We observe in Exercise 12 that the action of the origin is very simple on x-coordinates, but this is not necessarily true for y-coordinates. Similar behavior happens for the group operation and isogeny formulas. Luckily, as mentioned for example by Craig Costello last week, it suffices for cryptographic protocols to work on the *Kummer line* of a curve. For the Weierstrass and Montgomery models this corresponds to the x-line, hence it suffices to provide nice formulas for this.

Exercise 13. Let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ be points in $M(\mathbb{F}_p)$ such that $P_1 \neq \pm P_2$ and $x_1 x_2 \neq 0$. Let $P_1 + P_2 = (x_3, y_3)$ and $P_1 - P_2 = (x_4, y_4)$. Show that

$$x_3 x_4 = \frac{(x_1 x_2 - 1)^2}{(x_1 - x_2)^2}.$$

Exercise 14. Let notation be as in Exercise 13 but assume $P_1 = P_2$ instead. Show that

$$x_3 = \frac{(x_1^2 - 1)}{4x_1(x_1^2 + Ax_1 + 1)}.$$

Exercise 15. Show that addition and doubling can be implemented with 8 and 6 multiplications respectively in projective space, and without inversions. How does this compare to Weierstrass form?

The Montgomery model is extremely popular: it is used for Curve25519 [1], one of the most widely adopted curves for classical key exchange and adopted by NIST into their SP 800-186 standard. It is also used throughout SIKE as it leads to the most efficient arithmetic for (powers of) small degree isogenies. It was also originally proposed for CSIDH, but in that case (twisted) Edwards curves have also proven to be highly efficient. We discuss them next.

2.3 (Twisted) Edwards form

We invert the chronological order in this section. We define a *twisted Edwards curve* [2] as the (projective closure of the) elliptic curve defined by

$$\mathcal{E}^t : \alpha x^2 + y^2 = 1 + \delta x^2 y^2 \,.$$

for $\alpha, \delta \in \mathbb{F}_p$ such that $\alpha\delta(\alpha - \delta) \neq 0$. The base point is the affine point $(0, 1)$. Note that the homogeneous embedding into $\mathbb{P}^2$ does not work in this case, as it leads to singularities at infinity. Instead, the projective closure can be (other options exist) defined by

$$\alpha X^2 T^2 + Y^2 Z^2 = Z^2 T^2 + \delta X^2 Y^2 \subset \mathbb{P}^1 \times \mathbb{P}^1$$

by setting $x = X/Z$ and $y = Y/T$.

Exercise 16. Let $M : y^2 = x^3 + Ax^2 + x$ be a Montgomery curve with $A = (\alpha + \delta)/2$. Show that $\phi_0 : (x, y) \mapsto (x/y, (x+1)/(x-1))$ is an isomorphism of elliptic curves defined over $\mathbb{F}_p$ from M to $\mathcal{E}^t$.

Exercise 17. Let $P = (s, t) \in \mathcal{E}^t$. Show that $-P = (-s, t)$. Find $\mathcal{E}^t[2]$ and show that the points at infinity all lie in $\mathcal{E}^t[4]$.

Exercise 17 demonstrates a significant difference between the twisted Edwards form and Weierstrass or Montgomery models. Whereas the neutral point of the latter models lies at infinity and often has to be treated separately in addition formulas (though not always, see [12]), the neutral point for twisted Edwards curves is affine. Moreover, any point of odd order is affine. In a classical cryptographic context where we work in an $\mathbb{F}_p$-rational subgroup of prime order, we are guaranteed that all points defined over $\mathbb{F}_p$ lie in the affine subset and we can obtain *complete* formulas on this set (assuming that δ is non-square). In contrast to the well-known Weierstrass formulas, it is no longer necessary to distinguish between doubling and addition, or between the neutral element and any of the other points. A side-effect of the twisted Edwards model is that its order is always divisible by 4. For example, it contains the point $(0, -1)$ of order 2.

Exercise 18. Write down the projection of $\mathcal{E}^t$ onto the Kummer line. Show that the Kummer lines of M and $\mathcal{E}^t$ are connected by an *involution*. (Would you change anything in the definition of $\mathcal{E}^t$, if you had the chance?)

We define an *Edwards curve* [5] as the (projective closure of the) elliptic curve defined by

$$\mathcal{E} : x^2 + y^2 = c^2(1 + dx^2 y^2) \subset \mathbb{A}^2$$

for $c, d \in \mathbb{F}_p$ with $cd(1 - dc^4) \neq 0$, with *affine* neutral element $(0, c)$. This is slightly more general than the model originally defined by Harold Edwards [7], but the main ideas remain. Again, we embed it into $\mathbb{P}^1 \times \mathbb{P}^1$.

Exercise 19. Show that $\phi_1 : (x,y) \mapsto (x\sqrt{\delta}, y\sqrt{\delta/\alpha})$ is an isomorphism of elliptic curves defined from $\mathcal{E}^t$ to $\mathcal{E}$, for $c^2 = \delta/\alpha$ and $d = \alpha/\delta$. Show that ϕ_1 is defined over $\mathbb{F}_p$ if $\mathcal{E}^t[4] \subset \mathcal{E}^t(\mathbb{F}_p)$.

Exercise 20. Show that

$$\phi_2 : (x,y) \mapsto \left(\frac{4 + y^2\alpha - \delta}{(y^2\alpha/\delta - 1)\sqrt{\alpha\delta}}, \frac{4y}{(y^2\alpha/\delta - 1)x\sqrt{\alpha}} \right)$$

is a 2-isogeny from $\mathcal{E}$ to $\hat{M} : \hat{B}y^2 = x^3 + \hat{A}x^2 + x$ where $\hat{B} = 1/\sqrt{\alpha\delta}$ and $\hat{A} = -(\alpha + \delta)/\sqrt{\alpha\delta}$.

Exercise 21. Write down the x-line of a 2-isogeny $M \to \hat{M}$ with kernel $\langle(0,0)\rangle$.

Exercise 22. Write down doubling formulas on the twisted Edwards curve $\mathcal{E}^t$. How does the cost compare to Weierstrass and Montgomery form?

In general the addition formulas on a twisted Edwards curve are given by

$$(x_1, y_2) + (x_2, y_2) \mapsto \left(\frac{x_1y_2 + y_1x_2}{1 + \delta x_1x_2y_1y_2}, \frac{y_1y_2 - \alpha x_1x_2}{1 - \delta x_1x_2y_1y_2} \right),$$

which can be easily (with the help of `sage/magma`) be verified by using its isomorphic Montgomery form. Besides the fact that these formulas are *complete* on the affine patch, they are also symmetric on both the x- and y-line. That makes them much more suitable for generic additions, rather than Kummer line computations for the Montgomery model works well. For example, the digital signature scheme EdDSA [4] is defined with respect to the Ed25519 curve

$$\mathcal{E}^t/\mathbb{F}_{2^{255}-19} : -x^2 + y^2 = 1 + \frac{121665}{121666}x^2y^2,$$

which is isomorphic to Curve25519.

3 Isogenies

We now work out some explicit formulas for isogenies. Here we focus only on the *separable* case as it is all that is required for for SIKE and CSIDH. Further, for simplicity we focus on the *point evaluation* step of an isogeny, and not the computation of the codomain curve (which is also important!). We do not make any assumptions on fields of definition, though in practice it is typically set up so that all computations on the Kummer line and with curve coefficients can be performed over the base field.

3.1 Weierstrass form

We assume that the reader is familiar with Vélu's formulas, which for a subgroup

$$G \subset E : y^2 = x^3 + ax + b$$

define an isogeny ϕ such that $\ker \phi = G$ as

$$\phi : P \mapsto \big(x_P + \sum_{T \in G \setminus \mathcal{O}} (x_{P+T} - x_T), y_P + \sum_{T \in G \setminus \mathcal{O}} (y_{P+T} - y_T) \big) .$$

More concretely, Galbraith [8, Theorem 25.1.6] splits up G into its points of order 2 G_2 and G_1 such that

$$G = \{\mathcal{O}\} \cup G_1 \cup G_2 \cup \{-T : T \in G_1\}$$

and shows that

$$x_{\phi(P)} = x_P + \sum_{T \in G_1 \cup G_2} \frac{t(T)}{x_P - x_T} + \frac{u(T)}{(x_P - x_T)^2} , \tag{1}$$

where $u(T) = 4y_T^2$ and

$$t(T) = \begin{cases} 3x_T^2 + a & \text{if } T \in G_2 \\ 6x_T^2 + 2a & \text{if } T \in G_1 \end{cases} .$$

The codomain curve is given by

$$E' : y^2 = x^3 + (a - 5 \cdot \sum_{T \in G_1 \cup G_2} t(T))x + b - 7 \cdot \sum_{T \in G_1 \cup G_2} (u(T) + x_T t(T))$$

Exercise 23. Write down 2- and 3-isogenies on the Kummer line of a short Weierstrass curve. Assuming the inputs are in (homogeneous) projective coordinates, how many operations does it take to compute the codomain curve and to evaluate a point at ϕ?

3.2 Montgomery form

As we know, the Montgomery model M is a special case of the (long) Weierstrass form in which we have a point of order two at the origin (with vertical tangent line). Therefore we could simply apply Vélu's formulas to compute isogenies. However, although Vélu's formulas preserve long Weierstrass form, they do *not* necessarily preserve Montgomery form. Therefore additional computation is necessary to transform it to its Montgomery model, essentially moving a point of order 2 to the origin. In this section we take a different route: we show how the additional structure of the model leads to nicely structured isogenies as well.

Exercise 24. Equation (1) still holds true for long Weierstrass form, though t and u need to be slightly extended. Show that $x_{\phi(P)} = w(x_P)/v(x_P)$ for polynomials v, w where $v(x) = \prod_{T \in G \setminus \mathcal{O}}(x - x_T)$.

Exercise 25. Let $Q \in M$ such that $Q \notin G$. Show that

$$x_{\phi(P)} = x_{\phi(Q)} + c_1(x_P - x_Q) \prod_{T \in G \setminus \mathcal{O}} \frac{x_P - x_{Q+T}}{x_P - x_T}$$

for some unit c_1.

Exercise 26. Let ϕ be an isogeny M from a Montgomery model. Show that without loss of generationality we can assume that the co-domain curve is also in Montgomery form, and that ϕ fixes 0 and 1 on the Kummer line.

Exercise 27. Let $G \subset M$ be such that $(0, 0) \notin G$. Show that

$$x_{\phi(P)} = \pm x_P \prod_{T \in G \setminus \mathcal{O}} \frac{x_P x_T - 1}{x_P - x_T}.$$

is the x-line of an isogeny ϕ between Montgomery models with kernel G.

Exercise 28. Write down 2- and 3-isogenies on the Kummer line of a Montgomery curve. Assuming the inputs are in (homogeneous) projective coordinates, how many field operations does it take to evaluate a point at ϕ?

3.3 Twisted Edwards form

As the twisted Edwards form is quite different from the Weierstrass form, applying Vélu's formulas is not as trivial. Here we can instead use the fact that its Kummer line is connected to that of the Montgomery model via an involution.

Exercise 29. Let ψ be an isogeny from a twisted Edwards curve with kernel G such that $(0, -1) \notin G$ and that $\psi : (0, -1) \mapsto (0, -1)$. Show that

$$\psi : y \mapsto \frac{\prod_{T \in G}(y_T + y) - \prod_{T \in G}(y_T - y)}{\prod_{T \in G}(y_T + y) + \prod_{T \in G}(y_T - y)}.$$

Exercise 30. Write down 2- and 3-isogenies on the Kummer line of a twisted Edwards curve. Assuming the inputs are in (homogeneous) projective coordinates, how many field operations does it take to evaluate a point at ψ?

From Exercise 30 we see that the cost of evaluating an isogeny at a point is linear in $|G|$. In the case of cyclic isogenies, it is therefore linear in the order ℓ of the generator. We refer to the next session by Daniel Bernstein for formulas that require only $\tilde{O}(\sqrt{\ell})$ field operations.

References

[1] Daniel J. Bernstein, *Curve25519: New Diffie-Hellman Speed Records*, Public Key Cryptography - PKC 2006 (Berlin, Heidelberg) (Moti Yung, Yevgeniy Dodis, Aggelos Kiayias, and Tal Malkin, eds.), Springer Berlin Heidelberg, 2006, pp. 207–228.

[2] Daniel J. Bernstein, Peter Birkner, Marc Joye, Tanja Lange, and Christiane Peters, *Twisted Edwards Curves*, Progress in Cryptology – AFRICACRYPT 2008 (Berlin, Heidelberg) (Serge Vaudenay, ed.), Springer Berlin Heidelberg, 2008, pp. 389–405.

[3] Daniel J. Bernstein, Chitchanok Chuengsatiansup, David Kohel, and Tanja Lange, *Twisted hessian curves*, Proceedings of the 4th International Conference on Progress in Cryptology – LATINCRYPT 2015 - Volume 9230 (Berlin, Heidelberg), Springer-Verlag, 2015, p. 269–294.

[4] Daniel J. Bernstein, Niels Duif, Tanja Lange, Peter Schwabe, and Bo-Yin Yang, *High-speed high-security signatures*, Journal of Cryptographic Engineering **2** (2012), no. 2, 77–89.

[5] Daniel J. Bernstein and Tanja Lange, *Faster Addition and Doubling on Elliptic Curves*, Advances in Cryptology – ASIACRYPT 2007 (Berlin, Heidelberg) (Kaoru Kurosawa, ed.), Springer Berlin Heidelberg, 2007, pp. 29–50.

[6] Wieb Bosma, John J. Cannon, and Catherine Playoust, *The Magma Algebra System I: The User Language*, Journal of Symbolic Computation **24** (1997), no. 3/4, 235–265.

[7] Harold Edwards, *A normal form for elliptic curves*, Bulletin of the American Mathematical Society **44** (2007), no. 3, 393–422.

[8] Steven D. Galbraith, *Mathematics of public key cryptography*, Cambridge University Press, 2012.

[9] G. B. Huff, *Diophantine problems in geometry and elliptic ternary forms*, Duke Mathematical Journal **15** (1948), 443–453.

[10] Peter L. Montgomery, *Modular Multiplication without Trial Division*, Mathematics of Computation **44** (1985), no. 170, 519–521.

[11] ______, *Speeding the Pollard and elliptic curve methods of factorization*, Mathematics of Computation **48** (1987), no. 177, 243–264.

[12] Joost Renes, Craig Costello, and Lejla Batina, *Complete Addition Formulas for Prime Order Elliptic Curves*, Advances in Cryptology – EUROCRYPT 2016 (Berlin, Heidelberg) (Marc Fischlin and Jean-Sébastien Coron, eds.), Springer Berlin Heidelberg, 2016, pp. 403–428.

[13] The Sage Developers, *Sagemath, the Sage Mathematics Software System*, 2018, `https://sagemath.org`.