ECM Factorization with QRT Maps

Andrew N. W. Hone

1 Introduction

Elliptic curves are a fundamental tool in modern cryptography. The abelian group structure on an elliptic curve makes it suitable for versions of Diffie–Hellman key exchange and ElGamal key encryption, as well as providing techniques for primality testing and integer factorization, among many other applications relevant to network security [4, 22, 32, 36]. In this chapter, we consider an approach to integer factorization using elliptic curves.

The elliptic curve method (ECM) due to Lenstra [24] is one of the most effective methods known for finding medium-sized prime factors of large integers, in contrast to trial division, Pollard's rho method, or the $p-1$ method, which quickly find small factors, or sieve methods, which are capable of finding very large prime factors. For factoring an integer N, the basic idea of the ECM is to pick (at random) an elliptic curve E and a point $\mathrm{P} \in E$, then compute the scalar multiple $s\mathrm{P} = \mathrm{P} + \cdots + \mathrm{P}$ (s times) in the group law of the curve, using arithmetic in the ring $\mathbb{Z}/N\mathbb{Z}$, take a rational function f on E with a pole at the point O corresponding to the identity in the group E, and evaluate $f(s\mathrm{P})$ for some s chosen as the largest prime power less than some fixed bound B_1 or as the product of all such prime powers. For certain choices of E and P, this computation may lead to an attempt to divide by a non-unit in the ring, resulting in a factor of N being found.

This work begun on leave at School of Mathematics & Statistics, University of New South Wales, NSW 2052, Australia.

A. N. W. Hone (✉)
School of Mathematics, Statistics & Actuarial Science, University of Kent, Canterbury, UK
e-mail: A.N.W.Hone@kent.ac.uk

H. R. Arabnia et al. (eds.), *Advances in Software Engineering, Education, and e-Learning*, Transactions on Computational Science and Computational Intelligence, https://doi.org/10.1007/978-3-030-70873-3_28

To be more precise, traditionally, one starts with a Weierstrass cubic defined over $\mathbb{Q}$, which can be taken with integer coefficients as

$$y^2 = x^3 + Ax + B, \qquad A, B \in \mathbb{Z},$$

so that arithmetic mod N corresponds to working with the pseudocurve (or group scheme) $E(\mathbb{Z}/N\mathbb{Z})$ consisting of all $(x, y) \in (\mathbb{Z}/N\mathbb{Z})^2$ that satisfy the cubic equation together with O, the point at infinity; but, when N is composite, the group addition $\mathrm{P}_1 + \mathrm{P}_2$ is not defined for all pairs of points $\mathrm{P}_1, \mathrm{P}_2 \in E(\mathbb{Z}/N\mathbb{Z})$. Typically, f is taken to be the coordinate function x, and the method is successful if computing the scalar multiple $s\mathrm{P}$ leads to an x-coordinate with a denominator D which is not a unit in $\mathbb{Z}/N\mathbb{Z}$, such that $\gcd(D, N) > 1$ is a non-trivial factor of N. When this fortunate occurrence arises, it indicates that there is a prime factor $p|N$ for which $s\mathrm{P} = \mathrm{O}$ in the group law of the bona fide elliptic curve $E(\mathbb{F}_p)$, which is guaranteed if s is a multiple of the order $\#E(\mathbb{F}_p)$.

The original description of the ECM was based on computations with affine coordinates for a Weierstrass cubic; computing the scalar multiple $s\mathrm{P}$ is now known as "stage 1" of the ECM, and there is a further "stage 2", due to Brent, involving computing multiples $\ell s\mathrm{P}$ for small primes ℓ less than some bound $B_2 > B_1$, but here we only focus on stage 1. Improvements in efficiency can be made by using various types of projective coordinates and/or Montgomery curves (see chapter 7 in [4]). However, all of these approaches share an inconvenient feature of the addition law for $\mathrm{P}_1 + \mathrm{P}_2$ on a Weierstrass cubic, namely that the formulae for $\mathrm{P}_2 = \pm\mathrm{P}_1$ or $\mathrm{P}_2 = \mathrm{O}$ are different from the generic case.

An important new development was the proposal of Bernstein and Lange [1] to consider a different model for E, namely the Edwards curve [6]

$$E_d: \qquad x^2 + y^2 = 1 + dx^2y^2 \tag{1}$$

(d is a parameter), for which the addition law

$$(x_1, y_1) + (x_2, y_2) = \left(\frac{x_1y_2 + y_1x_2}{1 + dx_1x_2y_1y_2}, \frac{y_1y_2 - x_1x_2}{1 - dx_1x_2y_1y_2} \right) \tag{2}$$

has the advantage that it is also valid for a generic pair of points $\mathrm{P}_1, \mathrm{P}_2 \in E_d$, even when $\mathrm{P}_1 = \mathrm{P}_2$, so it can be used for doubling (following [1], we have used a rescaled curve compared with the original version in [6]). The fact that the addition law (2) on E_d is unified in this sense is implicit in the classical addition formula for the Jacobi sine function (see chapter XXII in [35], or chapter 22 in [28]), for we have been

$$\mathrm{sn}(z + w) = \frac{\mathrm{sn}(z)\mathrm{cd}(w) + \mathrm{cd}(z)\mathrm{sn}(w)}{1 + k^2\mathrm{sn}(z)\mathrm{sn}(w)\mathrm{cd}(z)\mathrm{cd}(w)},$$

$$\mathrm{cd}(z+w)=\frac{\mathrm{cd}(z)\mathrm{cd}(w)-\mathrm{sn}(z)\mathrm{sn}(w)}{1-k^2\mathrm{sn}(z)\mathrm{sn}(w)\mathrm{cd}(z)\mathrm{cd}(w)},$$

using Glaisher's notation for the quotient $\mathrm{cd}(z)=\mathrm{cn}(z)/\mathrm{dn}(z)=\mathrm{sn}(z+K)$, with the complete elliptic integral $K=K(k)$ being a quarter period of the Jacobi sine, which yields (2) when we parametrize the points on E_d by

$$(x,y)=\big(\mathrm{sn}(z),\mathrm{cd}(z)\big)=\big(\mathrm{sn}(z),\mathrm{sn}(z+K)\big) \tag{3}$$

and identify $d=k^2$.

It was shown in [1] that, compared with the Weierstrass representation and its variants, the Edwards addition law gives more efficient formulae for computing an addition step $(\mathrm{P}_1,\mathrm{P}_2)\mapsto \mathrm{P}_1+\mathrm{P}_2$ or a doubling step $\mathrm{P}_1\mapsto 2\mathrm{P}_1$, both of which are required to obtain the scalar multiple $s\mathrm{P}$ in subexponential time $O(\log s)$ via an addition chain. The implementation EECM-MPFQ introduced in [2] gains even greater efficiency by using twisted Edwards curves, with an extra parameter a in front of the term x^2 on the left-hand side of (1), and further optimizing the ECM in other ways, including the use of projective coordinates in $\mathbb{P}^2$, extended Edwards coordinates in $\mathbb{P}^3$, and choosing curves with large torsion.

In this chapter, we explore implementations of the ECM using other models of elliptic curves, which arise in the context of QRT maps, an 18-parameter family of birational maps of the plane introduced by Quispel, Roberts, and Thompson [30] to unify diverse examples of maps and functional relations appearing in dynamical systems, statistical mechanics, and soliton theory. A QRT map is one of the simplest examples of a discrete integrable system, being a discrete avatar of a Hamiltonian system with one degree of freedom, with an invariant function (conserved quantity) and an invariant measure (symplectic form) [5].

Each orbit of a QRT map corresponds to a sequence of points $\mathrm{P}_0+n\mathrm{P}$ on a curve of genus one, and in the special case $\mathrm{P}_0=\mathrm{O}$, the orbit consists of the scalar multiples $n\mathrm{P}$, being closely related to an elliptic divisibility sequence (EDS) [34]. Thus, we can implement the ECM by iterating a QRT map with a special choice of initial data and performing all the arithmetic in $\mathbb{Z}/N\mathbb{Z}$.

A terse overview of QRT maps is provided in the next section; see [5, 20, 21, 33] for further details. Section 3 briefly introduces Somos sequences and related EDS, showing how three particular examples of QRT maps arise in this context, namely the Somos-4 QRT map, the Somos-5 QRT map, and the Lyness map. Each of the subsequent Sects. 4–6 is devoted to one of these three types of QRT map, including the doubling map that sends any point $\mathrm{P}_1\mapsto 2\mathrm{P}_1$ and a corresponding version of the ECM. In Sect. 7, we analyse the complexity of scalar multiplication, concentrating on the Lyness case in projective coordinates, and the final section contains some conclusions.

2 A Brief Review of QRT Maps

A QRT map can be constructed from a biquadratic curve of the general form

$$F(x,y) := \sum_{i,j=0}^{2} a_{ij} x^i y^j = 0. \tag{4}$$

For generic coefficients a_{ij}, this is a smooth affine curve, and with the inclusion of additional points at infinity, it lifts to a smooth curve in $\mathbb{P}^1 \times \mathbb{P}^1$, by introducing homogeneous coordinates $\big((X:W),(Y:Z)\big)$ and setting $x = X/W$, $y = Y/Z$ to obtain a homogeneous equation of bidegree (2, 2), that is,

$$\hat{F}(X,W,Y,Z) = W^2 Z^2 F(X/W, Y/Z) = 0;$$

this curve is a double cover of $\mathbb{P}^1$ with four branch points and so has genus one by Riemann-Hurwitz. A biquadratic curve admits two simple involutions, namely the horizontal/vertical switches given by

$$\iota_h : (x,y) \mapsto (x^\dagger, y), \qquad \iota_v : (x,y) \mapsto (x, y^\dagger),$$

where $x^\dagger$ is the conjugate root of (4), viewed as a quadratic in x, and similarly for $y^\dagger$; the Vieta formulae for the sum/product of the roots of a quadratic allow explicit birational expressions to be given for these two involutions. On a given biquadratic curve, the QRT map is defined to be the composition of the two switches,

$$\varphi_{\mathrm{QRT}} = \iota_v \circ \iota_h,$$

which acts as a translation in the group law of the curve, $\varphi_{\mathrm{QRT}} : \mathrm{P}_0 \mapsto \mathrm{P}_0 + \mathrm{P}$, where the shift P is independent of the choice of initial point P_0 on the curve.

So far, the map φ_{QRT} is restricted to a single curve, but to define a map on the plane, one should allow each coefficient $a_{ij} = a_{ij}(\lambda)$ to be a linear function of a parameter λ, so that (4) becomes a biquadratic pencil,

$$E_\lambda : \qquad F(x,y) \equiv F_1(x,y) + \lambda\, F_2(x,y) = 0. \tag{5}$$

The map $(x,y) \mapsto \lambda = -F_1(x,y)/F_2(x,y)$, obtained by solving (5) for λ, defines an elliptic fibration of the plane over $\mathbb{P}^1$ (except at finitely many base points where $F_1 = F_2 = 0$). Each value of λ corresponds to a unique curve in the pencil, where the map φ_{QRT} is defined, and on each such curve, a suitable combination of Vieta formulae yields a birational expression, which is independent of λ, so defines a birational map on the (x,y) plane, also denoted φ_{QRT}. By construction, the function $-F_1/F_2$ is constant on each orbit and so is a conserved quantity for the map φ_{QRT} in the plane.

Henceforth, we restrict to the symmetric case $F(x, y) = F(y, x)$, so that each curve in the pencil also admits the involution

$$\iota : \ (x, y) \mapsto (y, x),$$

making the horizontal/vertical switches conjugate to one another; thus, φ_{QRT} is a perfect square: $\iota_v = \iota \circ \iota_h \circ \iota$, hence $\varphi_{\mathrm{QRT}} = (\iota \circ \iota_h)^2 = \varphi \circ \varphi$, where the "square root" of φ_{QRT} is the symmetric QRT map

$$\varphi = \iota \circ \iota_h .$$

As a simple example, note that the Edwards curve (1) is a symmetric biquadratic, and we can identify $d = \lambda$ as the parameter of the pencil. Then, the Vieta formula for the sum of the roots gives an expression that is independent of this parameter, and the symmetric QRT map $\varphi = \varphi_{Edwards}$ associated with this pencil has the very simple form

$$\varphi_{Edwards} : \qquad (x, y) \mapsto (y, -x),$$

which is periodic with period four, i.e. $(\varphi_{Edwards})^4 = \mathrm{id}$. This is another manifestation of the well-known fact that Edwards curves have 4-torsion or of the fact that the complete elliptic integral K in (3) is a quarter period of the Jacobi sine.

A generic symmetric QRT map is far from being so simple: starting from an initial point P_0 in the plane, each orbit is a sequence of points $\mathrm{P}_n = \mathrm{P}_0 + n\mathrm{P}$ on a particular curve E_λ, and in general (at least, over an infinite field), the shift P need not be a torsion point. Even over a finite field $\mathbb{F}_p$, where every point is torsion, the order of P typically varies with the choice of curve in the pencil, i.e. with the value of λ.

In the cases of interest for the rest of the chapter, the symmetric QRT map φ can be written in multiplicative form, so that the sequence of points P_n has coordinates $(x, y) = (u_n, u_{n+1})$, where u_n satisfies a recurrence of second order,

$$u_{n+2}\, u_n = \mathrm{R}(u_{n+1}), \tag{6}$$

for a certain rational function R of degree at most two, with coefficients that are independent of λ (cf. Proposition 2.5 in [15], or [20, 21], for more details).

3 Somos and Elliptic Divisibility Sequences

A Somos-k sequence satisfies a quadratic recurrence of the form

$$\tau_{n+k}\tau_n = \sum_{j=1}^{\lfloor k/2 \rfloor} \alpha_j\, \tau_{n+k-j}\tau_{n+j}, \tag{7}$$

where (to avoid elementary cases) it is assumed that $k \geq 4$ with at least two parameters $\alpha_j \neq 0$. It was a surprising empirical observation of Somos [31] that such rational recurrences can sometimes generate integer sequences, which was proved by Malouf [26] for the Somos-4 recurrence

$$\tau_{n+4}\tau_n = \alpha\, \tau_{n+3}\tau_{n+1} + \beta\, (\tau_{n+2})^2, \tag{8}$$

in the particular case that the coefficients are $\alpha = \beta = 1$ and the initial values are $\tau_0 = \tau_1 = \tau_2 = \tau_3 = 1$. A broader understanding came from the further observation that the recurrence (8) has the Laurent property [10], that is, $\tau_n \in \mathbb{Z}[\alpha, \beta, \tau_0^{\pm 1}, \tau_1^{\pm 1}, \tau_2^{\pm 1}, \tau_3^{\pm 1}]\ \forall n \in \mathbb{Z}$. Somos sequences arise from mutations in cluster algebras [9] or LP algebras [23] and as reductions of the bilinear discrete KP/BKP equations, being associated with translations on Jacobian/Prym varieties for the spectral curve of a corresponding Lax matrix [7, 17].

The three simplest non-trivial examples of Somos recurrences, with two terms on the right-hand side, are the Somos-4 recurrence (8), the Somos-5 recurrence

$$\tau_{n+5}\tau_n = \tilde{\alpha}\, \tau_{n+4}\tau_{n+1} + \tilde{\beta}\, \tau_{n+3}\tau_{n+2}, \tag{9}$$

and the special Somos-7 recurrence

$$\tau_{n+7}\tau_n = a\, \tau_{n+6}\tau_{n+1} + b\, \tau_{n+4}\tau_{n+3}. \tag{10}$$

All three of them can be reduced to two-dimensional maps of QRT type, and hence their orbits correspond to sequences of points $\mathrm{P}_0 + n\mathrm{P}$ on curves of genus one. (In contrast, generic Somos-6 sequences and Somos-7 sequences are associated with points on Jacobians of genus 2 curves [7].)

To see the connection with QRT maps, in (8), one should substitute

$$u_n = \frac{\tau_{n-1}\tau_{n+1}}{\tau_n^2} \implies u_{n+2}\, u_n = \frac{\alpha\, u_{n+1} + \beta}{(u_{n+1})^2}, \tag{11}$$

yielding a second-order recurrence that can be reinterpreted as the map

$$(u_n, u_{n+1}) \mapsto (u_{n+1}, u_{n+2})$$

in the plane, and it turns out to be a symmetric QRT map; for the associated biquadratic pencil and other details, see Sect. 4. Similarly, for the Somos-5 recurrence (9), one can make the substitution

$$u_n = \frac{\tau_{n-2}\tau_{n+1}}{\tau_{n-1}\tau_n} \implies u_{n+2}\,u_n = \frac{\tilde{\alpha}\,u_{n+1} + \tilde{\beta}}{u_{n+1}}, \tag{12}$$

where the latter recurrence for u_n is equivalent to the QRT map described in Sect. 5. Finally, for the special Somos-7 recurrence (10), one should substitute

$$u_n = \frac{\tau_{n-3}\tau_{n+2}}{\tau_{n-1}\tau_n} \implies u_{n+2}\,u_n = a\,u_{n+1} + b, \tag{13}$$

reducing the order from seven to two. The recurrence for u_n in (13) is known in the literature as the Lyness map, after the particular periodic case $b = a^2$ found in [25]; for details, see Sect. 6. The first two of these substitutions were derived in an ad hoc way in [14] and [15], but they all have a very natural interpretation in the theory of cluster algebras [8], which implies that these are the only Somos-k recurrences that can be reduced to two-dimensional maps.

Morgan Ward's elliptic divisibility sequences (EDSs) [34] are sequences of integers τ_n with $\tau_0 = 0$, $\tau_1 = 1$, $\tau_2, \tau_3, \tau_4 \in \mathbb{Z}$, and $\tau_2|\tau_4$, subject to the relations

$$\tau_{n+m}\tau_{n-m} = (\tau_m)^2\tau_{n+1}\tau_{n-1} - \tau_{m+1}\tau_{m-1}(\tau_n)^2, \tag{14}$$

$$\tau_2\tau_{n+m+1}\tau_{n-m} = \tau_{m+1}\tau_m\tau_{n+2}\tau_{n-1} - \tau_{m+2}\tau_{m-1}\tau_{n+1}\tau_n \tag{15}$$

for all $m, n \in \mathbb{Z}$. An EDS corresponds to a sequence of points nP on an elliptic curve over $\mathbb{Q}$. The relation (14) for $m = 2$ is a special case of the Somos-4 recurrence (8), with $\alpha = (\tau_2)^2$ and $\beta = -\tau_3$; similarly, (15) with $m = 2$ gives a special case of (9), and a linear combination of this relation for $m = 2$ and $m = 3$ yields (10) with the coefficients/initial values related in a particular way. The fact that the same EDS satisfies these higher Somos relations [29] provides one way to derive the isomorphisms between the associated biquadratic curves and a Weierstrass cubic in Theorem 1 below, which can also be deduced from results in [18].

4 Somos-4 QRT Map

Here, we give further details of the QRT map defined by (11) and the associated family of curves.

QRT map : $$\varphi : (x, y) \mapsto \Big(y, \big(\alpha\, y + \beta\big)/(xy^2)\Big). \tag{16}$$

Pencil of curves : $$x^2y^2 + \alpha\,(x + y) + \beta - J\,xy = 0. \tag{17}$$

Elliptic involution : $$\iota_E : (x, y) \mapsto \big(x, (\alpha\, x + \beta)/(x^2y)\big). \tag{18}$$

Identity element and shift : $$\mathrm{O} = (\infty, 0), \quad \mathrm{P} = (0, -\beta/\alpha). \tag{19}$$

$$\textbf{Doubling map :}\qquad \psi\,:\,(x,y)\mapsto$$

$$\left(\frac{\alpha\,(x-y)y\,(\alpha\,x+\beta-x^3y)}{(\alpha\,x+\beta-x^2y^2)^2},-\frac{(\alpha\,x+\beta-x^2y^2)(\alpha\,y+\beta-x^2y^2)}{\alpha\,xy(x-y)^2}\right). \tag{20}$$

The map (16) preserves the symplectic form $\omega=(xy)^{-1}\mathrm{d}x\wedge\mathrm{d}y$, that is, $\varphi^*(\omega)=\omega$, and the doubling map ψ gives $\psi^*(\omega)=2\omega$; the same is true for the Somos-5/Lyness maps. Each orbit of φ lies on a fixed biquadratic curve of the form (17), with $\lambda=-J$ being the parameter of the pencil (5); equivalently, solving (17) for $J=J(x,y)$ gives a conserved quantity for the map. On any curve (17), the elliptic involution (18) sends any point $\mathrm{P}\mapsto-\mathrm{P}$. A special sequence of points (u_n,u_{n+1}) on the curve is generated by iterating (16) with a suitable starting point, corresponding to the scalar multiples $n\mathrm{P}$ of a particular point P (the shift). To have both coordinates finite and non-zero, one should start with

$$2\mathrm{P}=(-\beta/\alpha,-\alpha(\alpha^2+\beta J)/\beta^2)=(u_2,u_3). \tag{21}$$

However, in order to calculate a particular scalar multiple $s\mathrm{P}$ in time $O(\log s)$, rather than $O(s)$, one must employ the doubling map on the curve, using some variant of the "double-and-add" method (an addition chain).

We can now present a version of the ECM based on the QRT map (16).

Algorithm 1 ECM with Somos-4 QRT *To factorize N, pick $\alpha,\beta,J\in\mathbb{Z}/N\mathbb{Z}$ at random and some integer $s>2$. Then, starting from the point $2\mathrm{P}=(u_2,u_3)$ on the curve (17), given by (21), use the QRT map (16) to perform addition steps and (20) to perform doubling steps, working in $\mathbb{Z}/N\mathbb{Z}$, to compute $s\mathrm{P}=(u_s,u_{s+1})$. Stop if, for some denominator D, $g=\gcd(D,N)>1$ appears at any stage; when $g<N$, the algorithm has been successful, but if $g=N$ or no forbidden divisions appear, then restart with new α,β,J, and/or a larger value of s.*

Example 1 Given $N=1{,}950{,}153{,}409$, we pick $\alpha=\beta=1$ and $J=4$ to find $(u_2,u_3)=(-1,-5)$, take $s=12$, and compute the sequence $(u_n \bmod N)$, that is,

$$\infty,0,-1,-5,1482116591,121884579,452175879,1062558798,154165861,$$
$$1566968710,1329544730,56956778,$$

where the last term is u_{11}; then, $g=\gcd(u_{11},N)=16{,}433$, so the algorithm terminates. Of course, not all the above terms are necessary, since by writing $12=2^2\times(2+1)$, it is more efficient to compute the addition chain $2\mathrm{P}\mapsto3\mathrm{P}\mapsto6\mathrm{P}\mapsto12\mathrm{P}$ using (16) and (20) as

$$(u_2,u_3)\overset{\varphi}{\mapsto}(u_3,u_4)\overset{\psi}{\mapsto}(u_6,u_7)\overset{\psi}{\mapsto}???$$

and then observe that the denominator $\alpha x+\beta-x^2y^2$ in (20) has common factor $g>1$ with N when $(x,y)=(u_6,u_7)$.

5 Somos-5 QRT Map

Here, we describe the features of the QRT map corresponding to recurrence (12).

QRT map : $$\varphi : (x, y) \mapsto \Big(y, \big(\tilde{\alpha}\, y + \tilde{\beta}\big)/(xy)\Big). \tag{22}$$

Pencil of curves : $$xy(x + y) + \tilde{\alpha}\,(x + y) + \tilde{\beta} - \tilde{J}\,xy = 0. \tag{23}$$

Elliptic involution : $\iota_E : \quad (x, y) \mapsto (y, x).$

Identity element and shift : $\mathrm{O} = (\infty, \infty), \qquad \mathrm{P} = (\infty, 0).$

Doubling map : $\psi : (x, y) \mapsto$

$$\left(\frac{(x^2y - \tilde{\alpha}x - \tilde{\beta})(x^2y - \tilde{\alpha}y - \tilde{\beta})}{x(x - y)(xy^2 - \tilde{\alpha}x - \tilde{\beta})}, \frac{(xy^2 - \tilde{\alpha}x - \tilde{\beta})(xy^2 - \tilde{\alpha}y - \tilde{\beta})}{y(y - x)(x^2y - \tilde{\alpha}y - \tilde{\beta})}\right). \tag{24}$$

The double of the translation point (shift) is $2\mathrm{P} = (0, -\tilde{\beta}/\tilde{\alpha}) = (u_2, u_3)$, so to obtain the sequence of multiples $n\mathrm{P}$, one must start with

$$3\mathrm{P} = (-\tilde{\beta}/\tilde{\alpha}, \tilde{J} + \tilde{\alpha}^2/\tilde{\beta} + \tilde{\beta}/\tilde{\alpha}) = (u_3, u_4). \tag{25}$$

We can paraphrase Algorithm 1 to get another version of the ECM.

Algorithm 2 ECM with Somos-5 QRT *To factorize N, pick $\tilde{\alpha}, \tilde{\beta}, \tilde{J} \in \mathbb{Z}/N\mathbb{Z}$ at random and some integer $s > 3$. Then, starting from $3\mathrm{P} = (u_3, u_4)$ on the curve (23), given by (25), use (22) to perform addition steps and (24) to perform doubling steps, working in $\mathbb{Z}/N\mathbb{Z}$, to compute $s\mathrm{P} = (u_s, u_{s+1})$. Stop if, for some denominator D, $g = \gcd(D, N)$ with $1 < g < N$ appears at any stage.*

6 Lyness Map

The real and complex dynamics of the recurrence (13), known as the Lyness map, has been studied by many authors, with a very detailed account in [5].

QRT map : $$\varphi : (x, y) \mapsto \left(y, \frac{a\,y + b}{x}\right). \tag{26}$$

Pencil of curves :

$$xy(x + y) + a\,(x + y)^2 + (a^2 + b)\,(x + y) + ab - K\,xy = 0. \tag{27}$$

Elliptic involution : $\iota_E : \quad (x, y) \mapsto (y, x).$

Identity element and shift : $$\mathrm{O} = (\infty, \infty), \quad \mathrm{P} = (\infty, -a). \tag{28}$$

$$\textbf{Doubling map :}\quad \psi:\ (x,y)\mapsto \Big(R(x,y),\,R(y,x)\Big),$$

$$R(x,y)=\frac{(xy-ay-b)(x^2y-a^2x-by-ab)}{x(x-y)(y^2-ax-b)}. \tag{29}$$

Doubling and tripling P give 2P $=(-a,0)$, 3P $=(0,-b/a)$, so to obtain the multiples $n\mathrm{P}=(u_n,u_{n+1})$ by iteration of (26) and (29), one should begin with

$$4\mathrm{P}=\left(-\frac{b}{a},-a-\frac{b(Ka+b)}{a(a^2-b)}\right)=(u_4,u_5). \tag{30}$$

Henceforth, it will be assumed that $b\neq a^2$, since otherwise all orbits of (13) are periodic with period five, meaning that P is a 5-torsion point on every curve in the pencil. This special case is the famous Lyness 5-cycle [25], related to the associahedron K_4 via the A_2 cluster algebra and to the Abel pentagon identity for the dilogarithm [27], among many other things.

The above formulae (and those for Somos-4/5) can all be obtained via the birational equivalence of curves described in the following theorem (cf. [18]).

Theorem 1 *Given a fixed choice of rational point* $\mathrm{P}=(\nu,\xi)\in\mathbb{Q}^2$ *on a Weierstrass cubic*

$$E(\mathbb{Q}):\ (y')^2=(x')^3+Ax'+B$$

over $\mathbb{Q}$*, a point* (x,y) *on a Lyness curve (27) is given in terms of* $(x',y')\in E(\mathbb{Q})$ *by*

$$x=-\frac{\beta(\alpha u+\beta)}{uv}-a,\qquad y=-\beta uv-a,$$

where

$$(u,v)=\left(\nu-x',\,\frac{4\xi y'+Ju-\alpha}{2u^2}\right)$$

are the coordinates of a point on the Somos-4 curve (17), and the parameters are related by

$$a=-\alpha^2-\beta J,\qquad b=2a^2+a\beta J-\beta^3,\qquad K=-2a-\beta J, \tag{31}$$

with

$$\alpha=4\xi^2,\qquad J=6\nu^2+2A,\qquad \beta=\frac{1}{4}J^2-12\nu\xi^2.$$

Also,

$$\left(-\frac{x+a}{\beta}, -\frac{y+a}{\beta}\right)$$

is a point on the Somos-5 curve (23) with parameters

$$\tilde{\alpha} = -\beta, \quad \tilde{\beta} = \alpha^2 + \beta J, \quad \tilde{J} = J.$$

Conversely, given $a, b, K \in \mathbb{Q}$, a point (x, y) on (27) corresponds to $(\bar{x}, \bar{y}) \in \bar{E}(\mathbb{Q})$, a twist of $E(\mathbb{Q})$ with coefficients $\bar{A} = \alpha^2\beta^4 A$ and $\bar{B} = \alpha^3\beta^6 B$, and P *in (28) corresponds to the point $(\bar{\nu}, \bar{\xi}) = \left(\frac{1}{12}(\beta J)^2 - \frac{1}{3}\beta^3, \frac{1}{2}\alpha^2\beta^3\right) \in \bar{E}(\mathbb{Q})$.*

Algorithm 3 ECM with Lyness *To factorize N, pick $a, b, K \in \mathbb{Z}/N\mathbb{Z}$ at random and some integer $s > 4$. Then, starting from $4\mathrm{P} = (u_4, u_5)$ on the curve (27), given by (30), use (26) to perform addition steps and (29) to perform doubling steps, working in $\mathbb{Z}/N\mathbb{Z}$, to compute $s\mathrm{P} = (u_s, u_{s+1})$. Stop if, for some denominator D, $g = \gcd(D, N)$ with $1 < g < N$ appears at any stage.*

7 Complexity of Scalar Multiplication

Of the three symmetric QRT maps above, the Lyness map (26) is the simplest, so we focus on that for our analysis. Before proceeding, we can make the simplification $a \to 1$ without loss of generality, since over $\mathbb{Q}$ we can always rescale $(x, y) \to (ax, ay)$ and redefine b and K. To have an efficient version of Algorithm 3, it is necessary to work in projective coordinates, to avoid costly modular inversions; then, only a single gcd needs to be calculated at the end. For cubic curves, it is most common to work in the projective plane $\mathbb{P}^2$ (or sometimes, Jacobian coordinates in the weighted projective space $\mathbb{P}(1, 2, 3)$ are used for Weierstrass cubics [11]). However, for the biquadratic cubics (27), $\mathbb{P}^1 \times \mathbb{P}^1$ is better, since doubling with (29) is of higher degree in $\mathbb{P}^2$.

In terms of projective coordinates in $\mathbb{P}^1 \times \mathbb{P}^1$, the Lyness map (26) becomes

$$\Big((X : W), (Y : Z)\Big) \mapsto \Big((Y : Z), ((aY + bZ)W : XZ)\Big). \tag{32}$$

Then, taking $a \to 1$, each addition step using (32) requires $2\mathbf{M} + 1\mathbf{B}$, i.e. two multiplications and one multiplication by parameter b.

The affine doubling map (29) for the Lyness case lifts to the projective version

$$\Big((X : W), (Y : Z)\Big) \mapsto \Big((A_1 B_1 : C_1 D_1), (A_2 B_2 : C_2 D_2)\Big), \tag{33}$$

where

$$X^* = A_1 B_1, \quad W^* = C_1 D_1, \quad Y^* = A_2 B_2, \quad Z^* = C_2 D_2,$$

$$A_1 = A_+ + A_-, \quad A_2 = A_+ - A_-, \quad B_1 = B_+ + B_-, \quad B_2 = B_+ - B_-,$$

$$C_1 = 2XT, \quad C_2 = -2YT, \quad D_1 = ZA_2 + C_2, \quad D_2 = WA_1 + C_1,$$

with $A_- = aT$ and

$$A_+ = 2G - aS - 2H', \; B_+ = S(G - a^2H - H') - 2aHH', \; B_- = T(G - a^2H + H'),$$

$$S = E + F, \; T = E - F, \; E = XZ, \; F = YW, \; G = XY, \; H = WZ, \; H' = bH.$$

Setting $a \to 1$ once again for convenience and using the above formulae, we see that doubling can be achieved with 15**M** + 1**B**. (To multiply by 2, use addition: $2X = X + X$.)

This should be compared with EECM-MPFQ [2]: using twisted Edwards curves $ax^2 + y^2 = 1 + dx^2y^2$ in $\mathbb{P}^2$, the projective addition formula requires 10**M** + 1**S** + 1**A** + 1**D** (**S**, **A**, *and* **D** denote squaring and multiplication by the parameters a and d, respectively), while doubling only takes 3**M** + 4**S** + 1**A**. So, the Lyness addition step (32) is much more efficient than for twisted Edwards, but doubling requires twice as many multiplications. For any addition chain, the number of doublings will be $O(\log s)$, so employing Algorithm 3 to carry out the ECM with the Lyness map in projective coordinates should require on average roughly twice as many multiplications per bit as for EECM-MPFQ.

8 Conclusions

Due to the complexity of doubling, it appears that scalar multiplication with Lyness curves is not competitive with the state of the art using twisted Edwards curves. However, in a follow-up study [19], we have shown that the projective doubling map (33) for Lyness curves can be made efficient by distributing it over four processors in parallel, dropping the effective cost to 4**M** + 1**B**. On the other hand, this is still roughly twice the cost of the best known algorithm for doubling with four processors on twisted Edwards curves in the special case $a = -1$ [13].

However, by Theorem 1, any elliptic curve over $\mathbb{Q}$ is isomorphic to a Lyness curve, while twisted Edwards curves only correspond to a subset of such curves. Thus, there may be other circumstances, whether for the ECM or for alternative cryptographic applications, where Lyness curves and QRT maps will prove to be useful. For instance, one could use families of Lyness curves with torsion subgroups that are impossible with twisted Edwards curves in EECM-MPFQ. Also, bitcoin uses the curve $y^2 = x^3 + 7$, known as secp256k1, which cannot be expressed in twisted Edwards form.

The remarkable simplicity of the addition step (32) means that it might also be suitable for pseudorandom number generation. In that context, it would be worth exploring non-autonomous versions of QRT maps mod N. For example, the recurrence

$$u_{n+2}u_n = u_{n+1} + b_n q^n, \qquad b_{n+6} = b_n \tag{34}$$

is a q-difference Painlevé version of the Lyness map (13) (see [16]), and over $\mathbb{Q}$, the arithmetic behaviour of such equations appears to be analogous to the autonomous case [12], with polynomial growth of logarithmic heights; although for (34), the growth is cubic rather than quadratic as in the elliptic curve case. It is interesting to compare this with the case where $q = 1$ and the coefficient b_n is periodic with a period that does not divide 6, when generically (34) appears to display chaotic dynamics [3], e.g. the period 5 example $u_{n+2}u_n = u_{n+1} + b_n$, $b_{n+5} = b_n$, for which the logarithmic height along orbits in $\mathbb{Q}$ grows exponentially with n. Working mod N, it would be worth carrying out a comparative study of the pseudorandom sequences generated by (34) to see how the behaviour for $q \neq 1$ differs from the Lyness case (13) and the effect of changing the period of b_n.

Acknowledgments This research was supported by Fellowship EP/M004333/1 from the Engineering & Physical Sciences Research Council, UK. The author thanks the School of Mathematics and Statistics, University of New South Wales, for hosting him twice during 2017–2019 as a Visiting Professorial Fellow, with funding from the Distinguished Researcher Visitor Scheme. He is also grateful to John Roberts and Wolfgang Schief for providing additional support during his stay in Sydney, where the idea behind this work originated, and to Reinout Quispel for useful discussions and hospitality during his visit to Melbourne in May 2019.

References

1. D.J. Bernstein, T. Lange, Faster addition and doubling on elliptic curves, in ed. by K. Kurosawa, *Advances in Cryptology – ASIACRYPT 2007* (Springer, Berlin, 2007), pp. 29–50. https://doi.org/10.1007/978-3-540-76900-2_3
2. D.J. Bernstein, P. Birkner, T. Lange, C. Peters, ECM using Edwards curves. Math. Comput. **82**, 1139–1179 (2013). https://doi.org/10.1090/S0025-5718-2012-02633-0
3. A. Cima, A. Gasull, V. Mañosa, Integrability and non-integrability of periodic non-autonomous Lyness recurrences. Dyn. Syst. **28**, 518–538 (2013). https://doi.org/10.1080/14689367.2013.821103
4. R. Crandall, C. Pomerance, *Prime Numbers - A Computational Perspective*, 2nd edn (Springer, New York, 2005)
5. J.J. Duistermaat, *Discrete Integrable Systems: QRT Maps and Elliptic Surfaces* (Springer, New York, 2010)
6. H.M. Edwards, A normal form for elliptic curves. Bull. Amer. Math. Soc. **44**, 393–422 (2007). https://doi.org/10.1090/S0273-0979-07-01153-6
7. Y.N. Fedorov, A.N.W. Hone, Sigma-function solution to the general Somos-6 recurrence via hyperelliptic Prym varieties. J. Integrable Syst. **1**, xyw012 (2016). https://doi.org/10.1093/integr/xyw012

8. A.P. Fordy, A.N.W. Hone, Discrete integrable systems and Poisson algebras from cluster maps. Commun. Math. Phys. **325**, 527–584 (2014). https://doi.org/10.1007/s00220-013-1867-y
9. A.P. Fordy, R.J. Marsh, Cluster mutation-periodic quivers and associated Laurent sequences. J. Algebraic Combin. **34**, 19–66 (2011). https://doi.org/10.1007/s10801-010-0262-4
10. D. Gale, The strange and surprising saga of the Somos sequences. Math. Intell. **13**(1), 40–42 (1991); Somos sequence update, Math. Intell. **13**(4), 49–50 (1991). Reprinted in D. Gale, *Tracking the Automatic Ant* (Springer, New York, 1998)
11. R.R. Goundar, M. Joye, A. Miyaji, Co-Z addition formulae and binary ladders on elliptic curves, in ed. by S. Mangard, F.-X. Standaert, *Cryptographic Hardware and Embedded Systems, CHES 2010*. Lecture Notes in Computer Science, vol. 6225. (Springer, Berlin, 2010), pp. 65–79. https://doi.org/10.1007/978-3-642-15031-9_5
12. R.G. Halburd, Diophantine integrability. J. Phys. A Math. Gen. **38**, L1–L7 (2005). https://doi.org/10.1088/0305-4470/38/16/L01
13. H. Huseyin, K.K.-H. Wong, G. Carter, E. Dawson, Twisted Edwards curves revisited, in ed. by J. Pieprzyk, *Advances in Cryptology - ASIACRYPT 2008*. Lecture Notes in Computer Science, vol. 5350 (2008), pp. 326–343. https://doi.org/10.1007/978-3-540-89255-7_20
14. A.N.W. Hone, Elliptic curves and quadratic recurrence sequences. Bull. Lond. Math. Soc. **37**, 161–171 (2005). https://doi.org/10.1112/S0024609304004163. Corrigendum. Bull. Lond. Math. Soc. **38**, 741–742 (2006). https://doi.org/10.1112/S0024609306018844
15. A.N.W. Hone, Sigma function solution of the initial value problem for Somos 5 sequences. Trans. Amer. Math. Soc. **359**, 5019–5034 (2007). https://doi.org/10.1090/S0002-9947-07-04215-8
16. A.N.W. Hone, R. Inoue, Discrete Painlevé equations from Y-systems. J. Phys. A: Math. Theor. **47**, 474007 (2014). https://doi.org/10.1088/1751-8113/47/47/474007
17. A.N.W. Hone, T.E. Kouloukas, C. Ward, On reductions of the Hirota-Miwa equation. SIGMA **13**, 057 (2017). https://doi.org/10.3842/SIGMA.2017.057
18. A.N.W. Hone, C.S. Swart, Integrality and the Laurent phenomenon for Somos 4 and Somos 5 sequences. Math. Proc. Camb. Phil. Soc. **145**, 65–85 (2008). https://doi.org/10.1017/S030500410800114X
19. A.N.W. Hone, Efficient ECM factorization in parallel with the Lyness map (2020). arXiv:2002:03811
20. A. Iatrou, J.A.G. Roberts, Integrable mappings of the plane preserving biquadratic invariant curves. J. Phys. A: Math. Gen. **34**, 6617–6636 (2001). https://doi.org/10.1088/0305-4470/34/34/308
21. A. Iatrou, J.A.G. Roberts, Integrable mappings of the plane preserving biquadratic invariant curves II. Nonlinearity **15**, 459–489 (2002). https://doi.org/10.1088/0951-7715/15/2/313
22. N. Koblitz, *Algebraic Aspects of Cryptography* (Springer, Berlin, 1998)
23. T. Lam, P. Pylyavskyy, Laurent phenomenon algebras. Cam. J. Math. **4**, 121–162 (2012). https://doi.org/10.4310/CJM.2016.v4.n1.a2
24. H.W. Lenstra, Jr., Factoring integers with elliptic curves. Ann. Math. **126**, 649–673 (1987). https://doi.org/10.2307/1971363
25. R.C. Lyness, Cycles. Math. Gaz. **26**, 62 (1942)
26. J.L. Malouf, An integer sequence from a rational recursion. Discrete Math. **110**, 257–261 (1992). https://doi.org/10.1016/0012-365X(92)90714-Q
27. T. Nakanishi, Periodicities in cluster algebras and dilogarithm identities, in ed. by A. Skowronski, K. Yamagata, *Representations of Algebras and Related Topics, EMS Series of Congress Reports* (European Mathematical Society, Zurich, 2011), pp. 407–444
28. F.W.J. Olver, A.B. Olde Daalhuis, D.W. Lozier, B.I. Schneider, R.F. Boisvert, C.W. Clark, B.R. Miller, B.V. Saunders, H.S. Cohl, M.A. McClain, (Eds.), *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/. Release 1.0.25 of 2019-12-15
29. A.J. van der Poorten, C.S. Swart, Recurrence relations for elliptic sequences: every Somos 4 is a Somos k. Bull. London Math. Soc. **38**, 546–554 (2006). https://doi.org/10.1112/S0024609306018534

30. G.R.W. Quispel, J.A.G. Roberts, C.J. Thompson, Integrable mappings and soliton equations. Phys. Lett. A **126**, 419–421 (1988)
31. M. Somos, Problem 1470. Crux Math. **15**, 208 (1989)
32. D.R. Stinson, *Cryptography Theory and Practice*, 3rd edn (Chapman & Hall/CRC, Boca Raton, 2006)
33. T. Tsuda, Integrable mappings via rational elliptic surfaces. J. Phys. A: Math. Gen. **37**, 2721–2730 (2004). https://doi.org/10.1088/0305-4470/37/7/014
34. M. Ward, Memoir on elliptic divisibility sequences. Amer. J. Math. **70**, 31–74 (1948). https://doi.org/10.2307/2371930
35. E.T. Whittaker, G.N. Watson, *A Course of Modern Analysis*, 4th edn. (Cambridge University Press, Cambridge, 1927)
36. S.Y. Yan, *Primality Testing and Integer Factorization in Public-Key Cryptography* (Kluwer Academic Publishers, Boston, 2004)