Explicit towers of curves with many points

Peter Bruin

25 April 2008

1. Introduction

The Drinfeld–Vlăduţ bound [2] states that for any finite field $\mathbf{F}_q$ of q elements, the limit

$$A(q) = \limsup_{C/\mathbf{F}_q} \frac{\#C(\mathbf{F}_q)}{g(C)},$$

where C runs over all smooth, projective, geometrically connected curves over $\mathbf{F}_q$ up to isomorphism, is at most $\sqrt{q}-1$. Ihara, Drinfeld and Vlăduţ also proved that this bound is attained if q is a square, using supersingular points on modular curves. Serre proved that it $A(q)$ is positive for all q.

Garcia and Stichtenoth started looking for explicit equations for families of curves $\{C_n\}_{n=1}^\infty$ in which $g(C_n) \to \infty$ as $n \to \infty$ and where $\lim_{n\to\infty} \#C_n(\mathbf{F}_q)/g(C)$ is positive. In the first part of this talk we explain one of their examples, based on the expositional article [5]. Elkies has given explicit equations for certain families of modular curves; see for example [3]. He also showed that all of the explicit towers which had so far been published were actually examples of towers of modular curves (of elliptic, Shimura or Drinfeld type.) The second part of this talk is devoted to an explanation of Elkies' method.

2. A first example of a tower with many points

This example is taken from Garcia and Stichtenoth [5], § 4. Consider the following tower of curves $\{C_n\}_{n\ge 1}$ over a field $\mathbf{F}_4$ with four elements: C_n is the normalisation of the curve in $(\mathbf{P}^1_{\mathbf{F}_4})^n$ defined by the $n-1$ equations

$$x_{i+1}^3 = (x_i+1)^3 - 1 \qquad (1 \le i \le n-1),$$

and the morphism $C_{n+1} \to C_n$ given by projection onto the first n coordinates. It is clear from the equations defining this tower that the place $x_1 = 0$ is totally ramified in each step $C_{n+1} \to C_n$; this implies that C_n is an integral curve.

The point $x_1 = \infty$ of $C_1 = \mathbf{P}^1_{\mathbf{F}_4}$ splits completely in C_2; this can be seen by rewriting the equation defining C_2 as

$$(x_2/x_1)^3 = 1 + x_1^{-1} + x_1^{-2}$$

and using the fact that $\mathbf{F}_4$ contains the third roots of unity. More precisely, the three points in C_2 with $x_1 = \infty$ are given by $x_2 = \infty$ and $x_2/x_1 = \zeta$ with ζ a third root of unity. (In the projective model that we have given, the last equation only makes sense after blowing up the model in the point $x_1 = x_2 = \infty$.) For the same reason, each of the three points x_1 in C_2 with $x_1 = \infty$ splits completely in C_3, and so on. This implies that C_n has at least 3^{n-1} rational points for each n.

The only points of C_1 above which the morphism $C_2 \to C_1$ is ramified are the zeros of $(x_1+1)^3 - 1$, i.e. $x_1 = 0$, $x_1 = \zeta_3 - 1$ and $x_1 = \zeta_3^2 - 1 = \zeta_3$, where ζ_3 is a primitive third root of unity. (These are precisely the elements of $\mathbf{F}_4$ except the unit element.)

If the morphism $C_3 \to C_2$ is ramified above a point P of C_2, then P has $x_2 \in \{0, \zeta_3 - 1, \zeta_3\}$. This set is contained in $\{0, 1, \zeta_3 - 1, \zeta_3\}$. For $x_2 = 0$ the equation defining C_2 implies $x_1 \in \{0, \zeta_3 - 1, \zeta_3\}$, and for $x_2 \in \{1, \zeta_3 - 1, \zeta_3\}$ it implies $x_1 = 1$. By induction, this means that the morphism $C_n \to C_1$ is only ramified for $x_1 \in \{0, 1, \zeta_3 - 1, \zeta_3\}$. Since furthermore this morphism is of degree 3^{n-1} and the total ramification index above each point is therefore at most $3^{n-1} - 1$, it follows from Hurwitz' genus formula that for all $n \ge 1$ we have

$$\begin{aligned} 2g(C_n) - 2 &\le 3^{n-1}(2g(C_1) - 2) + 4(3^{n-1} - 1) \\ &= 2 \cdot 3^{n-1} - 4. \end{aligned}$$

This implies that

$$g(C_n) \le 3^{n-1} - 1$$

and finally

$$\begin{aligned}\lim_{n\to\infty}\frac{\#C_n(\mathbf{F}_4)}{g(C_n)} &\geq \lim_{n\to\infty}\frac{3^{n-1}}{3^{n-1}-1}\\ &= 1.\end{aligned}$$

Since $\sqrt{4}-1=1$, this means that the tower $\{C_n\}_{n=1}^{\infty}$ attains the Drinfeld–Vlăduţ bound.

In the above we have implicitly used the fact that all morphisms $C_{n+1}\to C_n$ are tamely ramified; this is necessary for Hurwitz' genus formula to hold. In situations where wild ramification occurs, a little more care is needed.

3. Decomposition of cyclic isogenies

In the remainder of this talk, we will concentrate on towers of modular curves. which have first been studied by Ihara, Drinfeld and Vlăduţ; Elkies (see e.g. [3]) gave explicit equations for several families of such curves and showed that all explicit towers which had appeared in the literature were examples of towers of modular curves (of elliptic, Shimura or Drinfeld type).

We start with some generalities about isogenies between elliptic curves. Let S be a scheme and m an integer which is invertible on S. An isogeny of elliptic curves over S is called a *cyclic isogeny of degree* m if "locally for the étale topology on S" (i.e. after base extension by some surjective étale morphism $S'\to S$) its kernel is isomorphic to the constant group scheme $(\mathbf{Z}/m\mathbf{Z})_S$. Such isogenies are classified by an affine coarse moduli scheme over $\mathbf{Z}[1/m]$, which we denote by $Y_0(m)$.

Our goal is to prove that for $l=2,3,5$ and all $n\geq 1$ the affine curve $Y_0(l^n)$ (and hence also its compactification $X_0(l^n)$) is birationally equivalent to (i.e. has the same function field as) a closed subscheme of $(\mathbf{P}^1)^{n-1}$ given by $n-2$ equations which can be written down explicitly. These are the three simplest cases of the observation that for any integers $l\geq 2$ and $n\geq 2$ there is a birational map from $X_0(l^n)$ to a closed subscheme of $(X_0(l^2))$ (see below). The reason that we consider $l=2,3,5$ is that these are the prime numbers for which $X_0(l^2)$ has genus 0.

The observation referred to above is captured in the following two results:

Proposition 3.1. *Let l and n be positive integers, and let $\phi\colon E_0\to E_n$ be a cyclic isogeny of degree l^n of elliptic curves over some $\mathbf{Z}[1/l]$-scheme. Then ϕ has a unique decomposition*

$$\phi\colon E_0\xrightarrow{\phi_1}E_1\xrightarrow{\phi_2}E_2\xrightarrow{\phi_3}\cdots\xrightarrow{\phi_n}E_n$$

with each ϕ_i cyclic of degree l.

Proof. Let G denote the kernel of ϕ. Then the kernel of ϕ_1 in a decomposition as above is necessarily the unique subgroup $l^{n-1}G$ of order l in G, and $E_1=E_0/l^{n-1}G$. The image of G under ϕ_1 is then to $G/l^{n-1}G$, which is cyclic of order l^{n-1}. By induction we see that all the ϕ_i are defined uniquely. □

Proposition 3.2. *Let l and n be positive integers, and consider a sequence*

$$\phi\colon E_0\xrightarrow{\phi_1}E_1\xrightarrow{\phi_2}E_2\xrightarrow{\phi_3}\cdots\xrightarrow{\phi_n}E_n$$

of cyclic isogenies of degree l of elliptic curves over some $\mathbf{Z}[1/l]$-scheme. Then the composed isogeny $\phi\colon E_0\to E_n$ is cyclic (of degree l^n) if and only if each of the composed isogenies $\phi_{i+1}\circ\phi_i\colon E_{i-1}\to E_{i+1}$ with $1\leq i\leq n-1$ is cyclic (of degree l^2).

Proof. This is a special case of Theorem 6.7.15 of Katz and Mazur [6].

Remark. The assumption that l is invertible on the base scheme can be gotten rid of by requiring in addition that the decomposition is in so-called *standard order*. This property is defined in terms of *standard subgroups* of cyclic group schemes in the sense of Drinfeld. For details, see Katz and Mazur [6], [REF].

Using these two results, we can identify a cyclic isogeny $\phi\colon E\to E'$ of degree l^n with a sequence of $n-1$ isogenies of degree l^2 which overlap along an isogeny of degree l. A word of warning is appropriate: in general, given a sequence of $n-1$ isomorphism classes of such l^2-isogenies, the

l^n-isogeny *cannot* be reconstructed uniquely from this. Even over $\mathbf{C}$ there is a case where two overlapping 4-isogenies can be glued to an 8-isogeny in two non-isomorphic ways.

In terms of coarse moduli spaces, the situation can be described as follows. For $n \geq 2$, we have $n-1$ morphisms

$$p_i\colon Y_0(l^n) \to Y_0(l^2) \quad (1 \leq i \leq n-1)$$

where p_i sends $\phi\colon E_0 \to E_n$ to $\phi_{i+1} \circ \phi_i\colon E_{i-1} \to E_{i+1}$. Furthermore, there are two morphisms

$$q_1, q_2\colon Y_0(l^2) \to Y_0(l)$$

sending $E_0 \to E_1 \to E_2$ to $E_0 \to E_1$ and $E_1 \to E_2$, respectively. The curve $Y_0(2^n)$ can now be identified, at least up to birational equivalence, as

$$Y_0(2^n) \overset{(p_1,\ldots,p_{n-1})}{\longrightarrow} \left\{(x_1,\ldots,x_{n-1}) \in Y_0(4)^{n-1} \,\middle|\, q_1(x_{i+1}) = q_2(x_i) \text{ for } 1 \leq i \leq n-2\right\}.$$

4. Explicit equations for modular curves

In this section we give rational parameterisations of some curves $Y_0(n)$ for small n. As is well-known, $Y_0(1) = Y(1)$ is isomorphic to $\mathbf{A}^1_{\mathbf{Z}}$ and can be parameterised by the j-invariant. Because the formula for the j-invariant of a curve in general Weierstraß form is not very enlightening, we give the formula only for three special cases (note that in each case not every elliptic curve can be written in the indicated Weierstraß form):

base	Weierstraß equation	j-invariant
$\mathbf{Z}[1/2]$	$y^2 = x(x^2 + a_2x + a_4)$	$\dfrac{2^8(a_2^2 - 3a_4)^3}{a_4^2(a_2^2 - 4a_4)}$
$\mathbf{Z}[1/3]$	$y^2 = x^3 + a_4x + a_6$	$\dfrac{2^8 3^3 a_4^3}{4a_4^3 + 27a_6^2}$
$\mathbf{Z}$	$y^2 + a_1xy + a_3y = x^3$	$\dfrac{a_1^3(a_1^3 - 24a_3)^3}{a_3^3(a_1^3 - 27a_3)}$

Next we do $Y_0(2)$ over $\mathbf{Z}[1/2]$. Let E be an elliptic curve over a $\mathbf{Z}[1/2]$-scheme S together with a rational cyclic subgroup G of order 2. This G has the form $\{O, P\}$, where $O \in E(S)$ is the neutral element and $P \in E(S)$ is a point of order 2. Locally on S, we may choose coordinates (x, y) such that our equation has the form

$$y^2 = x(x^2 + ax + b)$$

as in the first line of the table above, where P has coordinates $(0, 0)$. This equation is unique up to substitutions of the form $y = u^3y'$ and $x = u^2x'$, where u is a unit on S; such a substitution results in a Weierstraß equation having coefficients $a' = u^{-2}a$ and $b' = u^{-4}b$. In particular, the rational function

$$h_2(E, G) = a^2/4b$$

on S is independent of the chosen coordinates and is therefore defined globally; it is called a *modular function.* We note that given values of a and b determine an elliptic curve if and only if both b and $a^2 - 4b$ are non-zero, i.e. if and only if $h_2 \notin \{1, \infty\}$. When S is the spectrum of a quadratically closed field K of characteristic different from 2, we can moreover choose coordinates uniquely such that the coefficient b in the above Weierstraß equation equals 1. This means that a given value $h_2 \in K \setminus \{1\}$ uniquely determines a pair (E, G) up to isomorphism. We summarise this by saying that the modular function h_2 is a *principal modulus* for $Y_0(2)$ (the German word *Hauptmodul* is more commonly used).

It follows directly from the definitions of j and h_2 that the composition of the quotient map

$$\begin{gathered} q_1\colon Y_0(2) \to Y(1) \\ (E, G) \mapsto E \end{gathered}$$

with the j-invariant $j\colon Y(1) \xrightarrow{\sim} \mathbf{A}^1$ is given in terms of the coordinate h_2 by

$$j \circ q_1 = 2^6 \frac{(4h_2 - 3)^3}{h_2 - 1}.$$

Equivalently, the following diagram of $\mathbf{Z}[1/2]$-schemes is commutative:

$$\begin{tikzcd}
Y_0(2) \arrow[r, "h_2"] \arrow[d, "q_1"'] & \mathbf{P}^1 \arrow[d, "2^6\frac{(4x-3)^3}{x-1}"] \\
Y(1) \arrow[r, "j"'] & \mathbf{P}^1.
\end{tikzcd}$$

To describe the map q_2 which sends (E, G) to E/G, we apply a well-known formula for the 2-isogeny with kernel $\{O, (0, 0)\}$ of the elliptic curve E given by $y^2 = x(x^2 + ax + b)$. Namely, this isogeny goes from E to the elliptic curve

$$E'\colon y'^2 = x'(x'^2 - 2ax' + a^2 - 4b)$$

and sends (u, v) to $\left(\frac{u^2+au+b}{u}, v\frac{u^2-b}{u^2}\right)$. The definitions of j and h_2 now imply that q_2 is given in coordinates by

$$j \circ q_2 = 2^6 \frac{(h_2 + 3)^3}{(h_2 - 1)^2},$$

or equivalenty by the following commutative diagram of $\mathbf{Z}[1/2]$-schemes:

$$\begin{tikzcd}
Y_0(2) \arrow[r, "h_2"] \arrow[d, "q_2"'] & \mathbf{P}^1 \arrow[d, "2^6\frac{(x+3)^3}{(x-1)^2}"] \\
Y(1) \arrow[r, "j"'] & \mathbf{P}^1.
\end{tikzcd}$$

In order to describe $Y_0(4)$, we have to parameterise elliptic curves E with a given cyclic subgroup G of order 4 over a $\mathbf{Z}[1/2]$-scheme S. As above, we can, locally on S, embed our curve into $\mathbf{P}^2$ via a Weierstraß equation

$$y^2 = x(x^2 + ax + b),$$

with $(0, 0)$ corresponding to the unique point of order 2 inside the given cyclic subgroup. It follows easily from the doubling formula for this elliptic curve that our subgroup has the form $\{x = c\}$, where c is a square root of b. Under a change of coordinates $x = u^2x'$ and $y = u^3y'$, both a and c are multiplied by u^{-2}, so that

$$h_4(E, G) = a/2c$$

is a well-defined function on S independent of the choice of coordinates. Moreover, the value of h_4 determines the isomorphism class of (E, G) (although this is not completely trivial if $h_4 = 0$), so that h_4 is a principal modulus for $Y_0(4)$. A given value of h_4 actually defines an elliptic curve with a cyclic subgroup of order 4 if and only if $h_4 \notin \{1, -1, \infty\}$.

The morphism $q_1\colon Y_0(4) \to Y_0(2)$ with moduli interpretation $(E, G) \mapsto (E, 2G)$ is given in coordinates by

$$h_2 \circ q_1 = h_4^2;$$

this follows directly from the definitions of h_2 and h_4. The morphism $q_2\colon Y_0(4) \to Y_0(2)$ with moduli interpretation $(E, G) \mapsto (E/2G, G/2G)$ is given by

$$h_2 \circ q_2 = \frac{(h_4 + 3)^2}{8(h_4 + 1)};$$

we leave it to the reader to check this using the formulas for 2-isogenies mentioned before.

Now we are ready to write down equations for $Y_0(2^n)$ (and $X_0(2^n)$) for all $n \geq 2$:

$$Y_0(2^n) \subset X_0(2^n) \longrightarrow \left\{ (x_1, \ldots, x_{n-1}) \in (\mathbf{P}^1)^{n-1} \;\middle|\; x_{i+1}^2 = \frac{(x_i+3)^2}{8(x_i+1)} \text{ for } 1 \leq i \leq n-2 \right\}.$$

The cusps are given by $x_i \in \{1, -1, \infty\}$ for any (hence all) x_i. As we remarked in § 3, these equations actually describe a singular model of $Y_0(2^n)$.

Now let p be an odd prime number, and consider the above family of modular curves over a field $\mathbf{F}_{p^2}$ of p^2 elements. By methods similar to the ones we saw last time in Jeroen's talk, one can show that the supersingular points on $Y_0(2^n)$ are $\mathbf{F}_{p^2}$-rational, and that there are precisely $(p-1)2^{n-3}$ of these points (Ihara–Vlăduţ). Using the well-known formulas for the genera of modular curves (see for example Diamond and Shurman [1], Chapter 3), it is straightforward to show that for $n \geq 2$

$$g(X_0(2^n)) = \begin{cases} 1 + 2^{n-3} - 2^{(n-1)/2}, & n \text{ odd;} \\ 1 + 2^{n-3} - 3 \cdot 2^{n/2-2}, & n \text{ even.} \end{cases}$$

This implies that the tower $\{X_0(2^n)\}$ over $\mathbf{F}_{p^2}$ attains the Drinfeld–Vlăduţ bound.

It remains to write down equations for the supersingular points. For this we use the following explicit description of the supersingular points on $Y(2)_{\mathbf{F}_p}$. First of all, an elliptic curve E over a field of characteristic different from 2, together with a basis for the 2-torsion, can be written in Legendre form as

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

where the given basis corresponds to $\{(0,0), (1,0)\}$. The x-coordinate λ of the third point of order 2 is then a principal modulus for $Y(2)$. Now a point on $Y(2)$ with a given value of λ is supersingular if and only if the so-called *Deuring polynomial*

$$H(X) = \sum_{m=0}^{(p-1)/2} \binom{(p-1)/2}{m} X^m$$

vanishes in λ. Furthermore, there is an isomorphism

$$Y_0(4) \xrightarrow{\sim} Y(2)$$

given by the commutative diagram

$$\begin{tikzcd}
Y_0(4) \arrow[r, "h_2"] \arrow[d, "\sim"'] & \mathbf{P}^1 \arrow[d, "\frac{x+1}{2}"] \\
Y(2) \arrow[r, "\lambda"] & \mathbf{P}^1.
\end{tikzcd}$$

Again, we leave it to the reader to show this using the formulas for 2-isogenies. This shows that a point on $Y_0(4)_{\mathbf{F}_p}$ with a given value of h_2 is supersingular if and only if

$$H\left(\frac{h_2+1}{2}\right) = 0.$$

It is possible to show, using properties of the Deuring polynomial $H(X)$, that for all points $(x_1, \ldots, x_{n-1})$ of $Y_0(2^n)$ over an algebraic closure of $\mathbf{F}_{p^2}$ the following holds: if $H(x_i) = 0$ for some i, then $H(x_i) = 0$ for all i, and in this case all the x_i are $\mathbf{F}_{p^2}$-rational. This was done by Garcia, Stichtenoth and Rück [4]; the proof is quite involved and uses the hypergeometric differential equation.

References

[1] F. Diamond and J. Shurman, *A First Course in Modular Forms.* Springer, New York, 2005.

[2] V. G. Drinfeld and S. G. Vlăduţ, Number of points of an algebraic curve. *Functional Anal. Appl.* **17** (1983), 53–54. (English translation.)

[3] N. D. Elkies, Explicit modular towers. In: T. Basar and A. Vardy (eds.), *Proceedings of the Thirty-Fifth Annual Allerton Conference on Communication, Control and Computing* (1997). University of Illinois at Urbana-Champaign, 1998.

[4] A. Garcia, H. Stichtenoth and H.-G. Rück, On tame towers over finite fields. *J. Reine Angew. Math.* **557** (2003), 53–80.

[5] A. Garcia and H. Stichtenoth, Explicit towers of function fields over finite fields. In: A. Garcia and H. Stichtenoth (eds.), *Topics in Geometry, Coding Theory and Cryptography*. Springer, Dordrecht, 2007.

[6] N. M. Katz and B. Mazur, *Arithmetic Moduli of Elliptic Curves.* Annals of Mathematics Studies **108**. Princeton University Press, Princeton, NJ, 1985.