CURVES OVER FINITE FIELDS WITH MANY POINTS: AN INTRODUCTION

JOHN VOIGHT

ABSTRACT. The number of points on a curve defined over a finite field is bounded as a function of its genus g. In this introductory article, we survey what is known about the maximum number of points on a curve of genus g defined over $\mathbb{F}_q$, including an exposition of upper bounds, lower bounds, known values of this maximum, and briefly indicate some methods of constructing curves with many points, providing many references to the literature.

By the Hasse-Weil bound (also known as the Riemann hypothesis for curves over finite fields), the number of points on a smooth, geometrically integral projective curve X of genus $g = g_X$ over a finite field $\mathbb{F}_q$ satisfies

$$\#X(\mathbb{F}_q) \leq q + 1 + 2g\sqrt{q}$$

and is therefore bounded as a function of g and q. It is natural to inquire about the sharpness of this upper bound, and so we consider the quantity

$$N_q(g) = \max_{\substack{X/\mathbb{F}_q \\ g_X = g}} \#X(\mathbb{F}_q)$$

which is the maximum number of points on a curve of genus g defined over $\mathbb{F}_q$, as well as the asymptotic quantity

$$A(q) = \limsup_{g \to \infty} \frac{N_q(g)}{g}.$$

Recently, these quantities have seen a great deal of study, motivated in part by applications in coding theory and cryptography and because it is an enticing problem. This article surveys the results in this area, including an exposition of upper bounds, lower bounds, known values of $N_q(g)$, and methods of constructing curves with many points.

1. Goppa codes

We begin with a brief motivation coming from coding theory. In 1981, Goppa introduced a way to associate an error-correcting code to a curve over a finite field [6, 7]—good codes require curves with a large number of

Date: May 5, 2005.

The author owes a great debt of gratitude to the Arizona Winter School and lectures by Kristin Lauter, René Schoof, and Michael Zieve for introducing him to the topic. The author wishes to further thank Michael Zieve, Bjorn Poonen, Hendrik Lenstra, and an anonymous reader for their helpful comments.

points, as we will see. For more information, see [22, §§II.2, VII.4] and [16, §5.5].

Messages are transmitted using a finite *alphabet* which we will take to be the finite field $\mathbb{F}_q$; the entire message is split into *words* which are blocks of fixed size and *encoded*: that is, provided with redundant letters to identify errors that were introduced during transmission. Therefore encoding is an injective map $E : \mathbb{F}_q^k \to \mathbb{F}_q^n$; the *code* is $C = \operatorname{img}(E)$, the *length* of the code is n.

If no errors occur, then the receiver can recover the original k letters using knowledge of E. If a word $v \notin C$ occurs, an error has occurred and assuming that the error made was small, the decoder looks for a word $\widetilde{v}$ in C that is as close to v as possible, i.e. such that the distance $|\widetilde{v} - v|$ is minimized, where the metric $|w|$ is defined to be the number of nonzero coordinates of w, called the *Hamming weight*. The *minimal distance* d is the minimal Hamming distance between distinct words in the code; since every ball of radius $< d/2$ contains at most one code word, one can correct up to $\lfloor (d-1)/2 \rfloor$ errors in each word. Therefore if possible we would like to simultaneously maximize k/n (the *transmission rate*) and d/n for a good code, and so we identify codes by the array $[k, d, n]$.

Let X be a curve, i.e. a smooth, geometrically integral projective variety of dimension 1, of genus g, defined over the finite field $\mathbb{F}_q$. Let D be a divisor on X defined over $\mathbb{F}_q$ and let $P_1, \dots, P_n$ be a collection of points in $X(\mathbb{F}_q)$. We assume that $P_1, \dots, P_n$ do not occur in D, but this assumption can be removed with a little extra effort. We define the map

$$\begin{aligned} \theta : L(D) &\to \mathbb{F}_q^n \\ f &\mapsto (f(P_1), \dots, f(P_n)), \end{aligned}$$

and in this way associate the code $C = \operatorname{img}\theta$ to the curve X. Note that the Goppa code defines a linear subspace of $\mathbb{F}_q^n$.

We can estimate the parameters of this code using standard facts about the geometry of curves (see e.g. [22]). The code has length n. If the image of $f \in L(D)$ has weight d, then f vanishes in $n-d$ points, so $\deg D - (n-d) \geq 0$ (a divisor has at least as many poles as zeros), i.e. $d \geq n - \deg D$, and hence we assume that $\deg D < n$ so that d is positive. Then $\ker\theta = L(D - \sum_{i=1}^n P_i)$ is trivial $(\deg(D - \sum_{i=1}^n P_i) < 0)$, and the dimension k of the code is by Riemann-Roch equal to $\dim L(D) \geq \deg D + (1-g)$. In sum,

$$d \geq n - \deg D \text{ and } k \geq \deg D + 1 - g.$$

In particular, using the Singleton bound $k + d \leq n + 1$ (proved as follows: a code of dimension k has a codeword with at least $k-1$ coordinates equal to 0 and hence has weight at most $n - (k-1)$, so $d \leq n + 1 - k$), we find

$$1 + \frac{1-g}{n} \leq \frac{k}{n} + \frac{d}{n} \leq 1 + \frac{1}{n};$$

Therefore a good code (one with k/n and d/n large) is one which arises from a curve of small genus with a large number of points.

Example 1.1. As an example, we can consider the Klein quartic X defined by the equation

$$x^3y + y^3z + xz^3 = 0$$

in the projective plane over $\mathbb{F}_2$. Since X is smooth and is defined by an equation of degree 4, the curve X has genus $g = (4-1)(4-2)/2 = 3$. One has the three points

$$(0:0:1), (0:1:0), (1:0:0) \in X(\mathbb{F}_2),$$

and these are the only points over any field $\mathbb{F}_{2^k}$ $(k \geq 1)$ with $xyz = 0$. Let us now consider the curve over $\mathbb{F}_8$. We can represent $\mathbb{F}_8 = \mathbb{F}_2(\alpha)$ where $\alpha^3 + \alpha + 1 = 0$; then α generates the multiplicative group $\mathbb{F}_8^*$. If $(x:y:z) \in X(\mathbb{F}_8)$ with $xyz \neq 0$, then we may assume $z = 1$, $y = \alpha^i$, and $x = \alpha^{3i}\beta$ for some $\beta \in \mathbb{F}_8$, so that

$$(\alpha^{3i}\beta)^3 + \alpha^{3i} + \alpha^{3i}\beta = \alpha^{3i}(\beta^3 + \beta + 1) = 0$$

hence $\beta^3 + \beta + 1 = 0$, so $\beta \in \{\alpha, \alpha^2, \alpha^2 + \alpha\}$. It follows that $\#X(\mathbb{F}_8^*) = 21$ and so

$$\#X(\mathbb{F}_8) = 24.$$

Let P be a $\mathbb{F}_8$-rational point and let D be the divisor $D = 10P$. If we take the sum of the other 23 points on X as the rational points divisor, then this code has length 23, dimension $10 - g + 1 = 8$ and minimum distance $\geq 23 - \deg(D) = 13$. If one adds a parity-check row and views $\mathbb{F}_8$ as a dimension 3 vector space over $\mathbb{F}_2$, one obtains a $[92, 24, \geq 26]$-code which is very nearly an optimal code with these parameters.

Asymptotically, the quality of these codes came as quite a surprise, exceeding the so-called Varshamov-Gilbert bound for $q \geq 7^2$—we refer the reader to [15].

2. Upper bounds

With this as our motivation, we begin by finding upper bounds on $N_q(g)$, the maximum number of points on a curve of genus g over $\mathbb{F}_q$.

2.1. **Hasse-Weil bound.** For a motivated introduction to the material in this section, see [8, Appendix C], and for an elementary proof of the Weil conjectures for curves, see [22, §V.1] or [16, Chapter 3].

To study the number of points $N_r = \#X(\mathbb{F}_{q^r})$, we form the zeta function of X:

$$Z(T) = Z_X(T) = \exp\left(\sum_{r=1}^{\infty} \#X(\mathbb{F}_{q^r})\frac{T^r}{r}\right) \in \mathbb{Q}[[t]].$$

Let a_d be the number of closed points of degree d of X. Then equivalently, we claim that we may write $Z(T)$ as the product

$$Z(T) = \prod_{d=1}^{\infty} \frac{1}{(1-T^d)^{a_d}};$$

since $N_r = \sum_{d|r} da_d$ (a closed point of degree d corresponds to a Galois-orbit of d points defined over $\mathbb{F}_{q^d}$ and hence $\mathbb{F}_{q^r}$ whenever $d \mid r$), taking the logarithmic derivative we find

$$\begin{aligned}\frac{d}{dT}\log\prod_{d=1}^{\infty}\frac{1}{(1-T^d)^{a_d}} &= \sum_{d=1}^{\infty}\frac{da_dT^{d-1}}{1-T^d} = \frac{1}{T}\sum_{d=1}^{\infty} da_d(T^d+T^{2d}+\dots) \\ &= \frac{1}{T}\sum_{r=1}^{\infty}\Bigl(\sum_{d|r} da_d\Bigr)T^r = \frac{1}{T}\sum_{r=1}^{\infty} N_rT^r = \frac{d}{dT}\log Z(T).\end{aligned}$$

The zeta function satisfies the following properties.

Theorem 2.1 (Weil conjectures for curves). *We have*

$$Z(T) = \frac{P(T)}{(1-T)(1-qT)}$$

where

$$P(T) = \prod_{i=1}^{2g}(1-\alpha_iT) \in \mathbb{Z}[T]$$

is a polynomial of degree $2g$ with reciprocal roots $\alpha_i \in \mathbb{C}$ that are algebraic integers satisfying $|\alpha_i| = \sqrt{q}$.

The zeta function $Z(T)$ satisfies the functional equation

$$Z\left(\frac{1}{qT}\right) = \frac{Z(T)}{(qT^2)^{g-1}}.$$

As a corollary, we have a bound on $\#X(\mathbb{F}_{q^r})$.

Corollary 2.2 (Hasse-Weil bound). *We have*

$$N_r = \#X(\mathbb{F}_{q^r}) = q^r + 1 - \sum_{i=1}^{2g}\alpha_i^r \le q^r + 1 + \lfloor 2gq^{r/2}\rfloor.$$

Proof. Taking the logarithmic derivative and multiplying by T, we have

$$\begin{aligned}\sum_{r=1}^{\infty} N_rT^{r-1} &= \frac{1}{1-T} + \frac{q}{1-qT} - \sum_{i=1}^{2g}\frac{\alpha_i}{1-\alpha_iT} \\ &= \sum_{r=0}^{\infty}(1+q^{r+1})T^r - \sum_{r=1}^{\infty}\left(\sum_{i=1}^{2g}\alpha_i^{r+1}\right)T^r.\end{aligned}$$

Hence

$$|N_r - (1-q^r)| \le \left|\textstyle\sum_{i=1}^{2g}\alpha_i^r\right| \le 2gq^{r/2}.$$

□

As a consequence, we know

$$A(q) = \limsup_{g\to\infty}\frac{N_q(g)}{g} \le 2\sqrt{q}.$$

We will see improvements on this bound in later sections of this article.

There is a large amount of literature which discusses the question of how one can efficiently compute the zeta function of a curve X given a set of equations which define X in an ambient projective space. We refer the reader to [27] and the references therein. We can always recover the zeta function in very simple cases by counting points over a sufficient number of extensions of the ground field, as indicated in the following example.

Example 2.3. Consider again the curve $X : x^3y+x^3z+y^3z = 0$ over $\mathbb{F}_2$. We write

$$Z(T) = \frac{P(T)}{(1-T)(1-2T)}$$

where $P(T) \in 1 + T\mathbb{Z}[T]$ is a polynomial of degree $2g = 6$ by the Weil conjectures. We count that X has number of points $N_i = 3, 5, 24, 17, 33, 38$ for $i = 1, \dots, 6$, hence

$$(1-T)(1-2T)\exp\left(3T + \frac{5}{2}T^2 + \cdots + \frac{38}{6}T^6 + \dots\right) = P(T)$$

and by matching coefficients we find that $P(T) = 1+5T^3+8T^6$, and therefore

$$Z(T) = \frac{1+5T^3+8T^6}{(1-T)(1-2T)}.$$

From this we can calculate the reciprocal roots α_i of $P(T)$; among them are $(1 \pm \sqrt{-7})/2$ which indeed have absolute value $\sqrt{2}$.

2.2. Serre bound.

For many years after Weil first proved these conjectures, there was essentially no investigation into how close the bound was to being sharp. In many cases, the bound can be improved greatly, and in the next few sections we will discuss the varied techniques which have been used to obtain improvements in certain cases. The first is due to Serre.

Proposition 2.4 (Serre [18, 19]). *We have*

$$N_q(g) \leq q + 1 + g\lfloor 2\sqrt{q}\rfloor.$$

Proof. Let $\beta_i = \lfloor 2\sqrt{q}+1\rfloor + \alpha_i + \overline{\alpha_i}$. Then $\beta_i \in \mathbb{R}$ are algebraic integers stable under $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ with $\beta_i > 0$ (since $|\alpha_i| = \sqrt{q}$). Therefore $\prod_{i=1}^{g} \beta_i > 0$ is an integer, and by the arithmetic-geometric mean

$$\frac{1}{g}\sum_{i=1}^{g} \beta_i \geq \left(\prod_{i=1}^{g} \beta_i\right)^{1/g} \geq 1$$

we deduce

$$g \leq \sum_{i=1}^{g} \beta_i = g\left(\lfloor 2\sqrt{q}\rfloor + 1\right) + \sum_{i=1}^{g} (\alpha_i + \overline{\alpha_i})$$

which gives the result. □

As a consequence, we immediately obtain $A(q) \leq \lfloor 2\sqrt{q}\rfloor$.

Example 2.5. Serre's modification already gives an essential improvement. For $g = 3$, $q = 8$, the Weil bound gives $N_8(3) \leq 25$, whereas the improved bound gives $N_8(3) \leq 24$; but this is exactly the number of points on the Klein quartic (1.1). We conclude that $N_8(3) = 24$.

2.3. **Ihara bound.** When the genus is large in comparison to the size of the field, a significant improvement can be made, due to Ihara.

Theorem 2.6 (Ihara [11]). *We have*

$$N_q(g) \leq \frac{1}{2}\left(\sqrt{(8q+1)g^2 + (4q^2-4q)g} - (g-2q-2)\right)$$

and asymptotically

$$A(q) \leq \frac{1}{2}\left(\sqrt{8q+1} - 1\right).$$

Proof. Let X be a curve of genus g over $\mathbb{F}_q$. Then by (2.1),

$$\#X(\mathbb{F}_{q^r}) = q + 1 - \sum_{i=1}^{g} (\alpha_i^r + \overline{\alpha_i}^r).$$

If we write $a_i = \alpha_i + \overline{\alpha_i}$, then

$$q + 1 - \sum_{i=1}^{g} a_i = \#X(\mathbb{F}_q) \leq \#X(\mathbb{F}_{q^2}) = q^2 + 1 + 2qg - \sum_{i=1}^{g} a_i^2$$

using the fact that $\alpha_i\overline{\alpha_i} = q$.

Let $N = \#X(\mathbb{F}_q)$. Then by the Cauchy-Schwarz inequality, we have

$$g\left(\textstyle\sum_{i=1}^{g} a_i^2\right) \leq \left(\textstyle\sum_{i=1}^{g} a_i\right)^2,$$

hence

$$N = \#X(\mathbb{F}_q) \leq q^2 + 1 + 2qg - \frac{1}{g}\Big(\sum_{i=1}^{g} a_i\Big)^2 = q^2 + 1 + 2qg - \frac{1}{g}(q+1-N)^2.$$

Simplifying, we have

$$N^2 + (g - 2 - 2q)N + (q+1)^2 - (q^2+1)g - 2qg^2 \leq 0;$$

solving for N we obtain

$$N \leq \frac{1}{2}\left(\sqrt{(8q+1)g^2 + (4q^2-4q)g} - g + (2q-2)\right)$$

as claimed. Finally,

$$\frac{N}{g} \leq \frac{1}{2}\left(\sqrt{(8q+1) + \frac{4q^2-4q}{g}} - 1 + \frac{2q-2}{g}\right) \to \frac{1}{2}\left(\sqrt{8q+1} - 1\right)$$

as $g \to \infty$. □

Remark 2.7. This bound is better than the Hasse-Weil bound when

$$\begin{aligned}2\,(q+1+2g\sqrt{q}) &> \sqrt{(8q+1)g^2+(4q^2-4q)g}-(g-2q-2)\\ g^2(1+4\sqrt{q})^2 &> (8q+1)g^2+(4q^2-4q)g\\ g &> \frac{4q^2-4q}{(1+4\sqrt{q})^2-8q-1}=\frac{4q(q-1)}{8\sqrt{q}(\sqrt{q}+1)}=\frac{\sqrt{q}(\sqrt{q}-1)}{2},\end{aligned}$$

i.e. when the genus is large in comparison to the size of the field. The bound is already better for $q=4$ and $g>1$.

Example 2.8. For example, the Weil bound tells us that $N_2(100)\le 285$, but the Ihara bound tells us that $N_2(100)\le 159$.

2.4. **Drinfel'd-Vlăduţ bound.**

Generalizing Ihara's argument, Drinfel'd-Vlăduţ obtained the following.

Theorem 2.9 (Drinfel'd-Vlăduţ bound [26]). *We have*

$$A(q)\le \sqrt{q}-1.$$

Proof. For every positive integer k, by the Weil conjectures we have

$$\#X(\mathbb{F}_{q^k})=q^k+1-\sum_{i=1}^{g}(\alpha_i^k+\overline{\alpha_i}^k)$$

with $\alpha_i=\sqrt{q}\omega_i$ where $|\omega_i|=1$. Then

$$0\le\sum_{i=1}^{g}|1+\omega_i+\cdots+\omega_i^k|^2=\sum_{i=1}^{g}\left(\sum_{r=0}^{k}\omega_i^r\right)\left(\sum_{r=0}^{k}\overline{\omega_i}^r\right);$$

But

$$(1+\omega_i+\cdots+\omega_i^k)(1+\overline{\omega_i}+\cdots+\overline{\omega_i}^k)=\sum_{r=0}^{k}(k+1-r)(\omega_i^r+\overline{\omega_i}^r)+(k+1),$$

hence

$$\begin{aligned}0&\le g(k+1)+\sum_{r=1}^{k}(k+1-r)\sum_{i=1}^{g}(\omega_i^r+\overline{\omega_i}^r)\\ &=g(k+1)+\sum_{r=1}^{k}(k+1-r)\frac{q^r+1-\#X(\mathbb{F}_{q^r})}{(\sqrt{q})^r}\\ &\le g(k+1)+\sum_{r=1}^{k}(k+1-r)\frac{q^r+1-N}{q^{r/2}}.\end{aligned}$$

where $N = \#X(\mathbb{F}_q)$. Solving for N gives

$$N\sum_{r=1}^{k}(k+1-r)\frac{1}{q^{r/2}} \leq g(k+1) + \sum_{r=1}^{k}(k+1-r)q^{r/2} + \sum_{r=1}^{k}(k+1-r)\frac{1}{q^{r/2}}$$

$$N \leq 1 + \frac{g(k+1) + \sum_{r=1}^{k}(k+1-r)q^{r/2}}{\sum_{r=1}^{k}(k+1-r)q^{-r/2}}.$$

Therefore as $g \to \infty$ we have

$$\begin{aligned} A(q) = \limsup_{g\to\infty}\frac{N_q(g)}{g} &\leq \frac{k+1}{\sum_{r=1}^{k}(k+1-r)q^{-r/2}} \to \frac{1}{\sum_{r=1}^{\infty}q^{-r/2}} \\ &= \left(\frac{q^{-1/2}}{1-q^{-1/2}}\right)^{-1} = \sqrt{q}-1 \end{aligned}$$

as claimed. □

Remark 2.10. The argument of Drinfel'd-Vlăduţ also gives upper bounds for $N_q(g)$ when g is much larger than q. For example, choosing $k = 8$ in the above proof we find that $N_2(100) \leq 77$.

2.5. **Oesterlé bound.**

We would like to modify the preceding argument to find the best possible minimal bound obtainable by such methods.

As an overview to motivate the method, suppose we start with a set of complex numbers α_i that might feasibly arise as *Frobenius eigenvalues* of a curve X, i.e. reciprocal roots of the numerator of the zeta function of X. Suppose we forget almost all properties of these numbers α_i, remembering only that they occur in complex conjugate pairs and have complex absolute value $|\alpha_i| = \sqrt{q}$. We form a formal collection of these numbers by allowing them to have multiplicities which are rational numbers. Then from any such formal collection, if we assume that they in fact arise from a curve X, we can write down the zeta function of the corresponding curve, and hence the genus and number of points. The Oesterlé bound we will exhibit below is the optimal upper bound on $N_q(g)$ for this general type of object.

We carry out this program following Serre [18]. If we let $\omega_j = e^{i\theta_j}$ for $\theta_j \in \mathbb{R}$ in the above notation, then

$$\begin{aligned} N_r = \#X(\mathbb{F}_{q^r}) = q^r + 1 + \sum_{j=1}^{g}(\alpha_j^r + \overline{\alpha_j}^r) &= q^r + 1 - q^{r/2}\sum_{j=1}^{g}(e^{ir\theta_j} + e^{-ir\theta_j}) \\ &= q^r + 1 - 2q^{r/2}\sum_{j=1}^{g}\cos(r\theta_j). \end{aligned}$$

Therefore if we are given real numbers $c_r \in \mathbb{R}$, we can multiply the above equation by c_r and divide by $q^{r/2}$ to obtain

$$2\sum_{j=1}^{g}c_r\cos(r\theta_j) + c_rN_rq^{-r/2} = c_rq^{r/2} + c_rq^{-r/2}.$$

Define

$$f(\theta) = 1 + 2\sum_{n=1}^{\infty} c_r \cos(r\theta), \qquad \Psi_d(t) = \sum_{r=1}^{\infty} c_{rd} t^{rd},$$

so after summing over r we obtain

$$\sum_{j=1}^{g} 2\sum_{r=1}^{\infty} c_r \cos(r\theta_j) + \sum_{r=1}^{\infty} c_r N_r q^{-r/2} = \sum_{r=1}^{\infty} c_r q^{r/2} + \sum_{r=1}^{\infty} c_r q^{-r/2}$$

$$\sum_{j=1}^{g} f(\theta_j) + \sum_{r=1}^{\infty} c_r N_r q^{-r/2} = g + \Psi_1(q^{1/2}) + \Psi_1(q^{-1/2}).$$

Now from $N_r = \sum_{d|r} da_d$ where a_d is the number of closed points of X of degree d, we have

$$\begin{aligned}\sum_{r=1}^{\infty} c_r N_r q^{-r/2} &= \sum_{r=1}^{\infty} c_r \left(\sum_{d|r} da_d\right) q^{-r/2} \\ &= \sum_{d=1}^{\infty} da_d \left(\sum_{m=1}^{\infty} c_{md} q^{-md/2}\right) = \sum_{d=1}^{\infty} da_d \Psi_d(q^{-1/2}).\end{aligned}$$

We have shown the following result.

Proposition 2.11 (Weil's explicit formulas)**.** *With the notation above, we have*

$$\sum_{j=1}^{g} f(\theta_j) + \sum_{d=1}^{\infty} da_d \Psi_d(1/\sqrt{q}) = g + \Psi_1(\sqrt{q}) + \Psi_1(1/\sqrt{q}).$$

Serre [18] used these formulas to obtain bounds on N_r as follows.

Corollary 2.12. *If the c_r are chosen such that $c_r \geq 0$ for all r and $f(\theta) \geq 0$ for all $\theta \in \mathbb{R}$, and only finitely many c_r are nonzero, then*

$$N_q(g) \leq \frac{g}{\Psi(1/\sqrt{q})} + 1 + \frac{\Psi(\sqrt{q})}{\Psi(1/\sqrt{q})},$$

where $\Psi = \Psi_1$.

Proof. We have

$$\begin{aligned} N\Psi_1(1/\sqrt{q}) &= g + \Psi_1(\sqrt{q}) + \Psi_1(1/\sqrt{q}) - \sum_{d=2}^{\infty} da_d \Psi_d(1/\sqrt{q}) - \sum_{j=1}^{g} f(\theta_j) \\ &\leq g + \Psi_1(\sqrt{q}) + \Psi_1(1/\sqrt{q}) \end{aligned}$$

since by assumption $f(\theta_j) \geq 0$ and $c_r \geq 0$. □

For every choice of c_r satisfying these conditions, we obtain an upper bound. In particular, the Weil bound corresponds to the choice $f(\theta) = 1 + \cos\theta$.

Example 2.13 (Serre [18]). We can, for example, take f of the form

$$f(\theta) = \frac{(1 + 2\sum_{r=1}^{\infty} u_r \cos r\theta)^2}{1 + 2\sum_{r=1}^{\infty} u_r^2}.$$

It has constant term 1 (just check the Taylor expansion), and using trigonometric identities we can write it in the desired form if we take $u_r \geq 0$. If we take $q = 2$, then the choice $u_1, u_2, u_3 = 1, 3/4, 1/2$ gives

$$f(\theta) = 1 + \frac{31}{17}\cos\theta + \frac{24}{17}\cos 2\theta + \frac{16}{17}\cos 3\theta + \frac{1}{2}\cos 4\theta + \frac{3}{17}\cos 5\theta + \frac{1}{34}\cos 6\theta$$

and thus

$$N_2(g) \leq \frac{544}{318\sqrt{2} + 227} g + \frac{16(75\sqrt{2} + 86)}{318\sqrt{2} + 227} \leq 0.8038g + 4.514.$$

We find $N_2(g) \leq 7, 8, 9$ for $g = 3, 4, 5$. Compare this to the Weil bound,

$$N_2(g) \leq 3 + 2\sqrt{2}g \leq 2.828g + 3,$$

which gives $N_2(g) \leq 11, 14, 17$.

In this case, these upper bounds are in fact best possible. For example, the curve

$$X : x^3y + y^3z + xz^3 + x^2y^2 + x^2z^2 + y^2z^2 + x^2yz + xy^2z = 0$$

passes through each of the 7 points of the projective plane $\mathbb{P}^2_{\mathbb{F}_2}$, and X is nonsingular and hence has genus 3. We will briefly mention some methods of constructing curves with many points below (§4).

Corollary 2.14 (Serre [20]). *We have*

$$N_q(g) \leq \begin{cases} q^2 + 1, & \text{if } g \leq \sqrt{2q}(q-1)/2; \\ q^3 + 1, & \text{if } g \leq \sqrt{q}\left(\sqrt{3}(q+1) + \sqrt{q}\right)(q-1)/2. \end{cases}$$

Proof. If we take

$$f(\theta) = \frac{1}{2}(1 + \sqrt{2}\cos\theta)^2 = 1 + \sqrt{2}\cos\theta + \frac{1}{2}\cos 2\theta$$

then

$$\Psi(1/\sqrt{q}) = \frac{1}{4q}(2\sqrt{2q} + 1), \quad \Psi(\sqrt{q}) = \frac{\sqrt{q}}{4}(2\sqrt{2} + \sqrt{q})$$

and hence

$$\begin{aligned} N &\leq \frac{4gq}{2\sqrt{2q} + 1} + 1 + q\sqrt{q}\left(\frac{2\sqrt{2} + \sqrt{q}}{2\sqrt{2q} + 1}\right) \\ &\leq \frac{2q\sqrt{2q}(q-1) + 2q\sqrt{2q} + q^2}{2\sqrt{2q} + 1} + 1 = q^2 + 1. \end{aligned}$$

The second statement follows from applying the same argument to

$$\begin{aligned} f(\theta) &= \frac{1}{3}\cos^2\theta\left(\sqrt{3}+2\cos\theta\right)^2 \\ &= 1+\sqrt{3}\cos\theta+\frac{7}{6}\cos 2\theta+\frac{\sqrt{3}}{3}\cos 3\theta+\frac{1}{6}\cos 4\theta. \end{aligned}$$

□

As can be seen from the above, the next problem is to find an optimal choice of the c_r which will minimize this upper bound. Equivalently, given a fixed number of points N and field q, by (2.12) we have

$$g \geq (N-1)\Psi(1/\sqrt{q}) - \Psi(\sqrt{q}) = (N-1)\sum_{r=1}^{\infty} c_r q^{-r/2} - \sum_{r=1}^{\infty} c_r q^{r/2}$$

and thus we want the c_r that give the best possible lower bound on g. Oesterlé was able to give an explicit recipe for finding the c_r using linear programming. It is really something of a miracle that one can give a formula for every q and g as follows.

Theorem 2.15 (Oesterlé). *Let $\lambda = N-1$ and m be the integer such that $q^{m/2} < \lambda \leq q^{(m+1)/2}$. Let*

$$u = \frac{q^{(m+1)/2}-\lambda}{\lambda\sqrt{q}-q^{m/2}}.$$

Then there is a unique solution $\vartheta \in [\pi/(m+1), \pi/m)$ to

$$\cos\left(\vartheta(m+1)\right) + u\cos\left(\vartheta(m-1)/2\right) = 0$$

and if we let

$$a_r = (m-r)\cos(r\vartheta)\sin\vartheta + \sin((m-r)\vartheta)$$

and $c_r = a_r/a_0$, then

$$g \geq \sum_{r=1}^{m-1} c_r(\lambda q^{-r/2} - q^{r/2}) = \frac{(\lambda-1)\sqrt{q}\cos(\vartheta)+q-\lambda}{q-2\sqrt{q}\vartheta+1}.$$

Ingredients of proof. The proof is given in [10], following Serre in an exposition of Oesterlé's argument in unpublished notes from a course at Harvard in 1985.

There exists a measure μ on $\mathbb{S}^1 = \{z : \mathbb{C} : |z| = 1\}$ such that

$$\frac{1}{2}\int_{\mathbb{S}^1} d\mu = (N-1)\sum_{r=1}^{\infty} c_r q^{-r/2} - \sum_{r=1}^{\infty} c_r q^{r/2}$$

with the property that equality occurs exactly when the c_r are optimal (if $q \geq 3$; one can treat the case $q = 2$ by different methods). This measure is given by $\mu = \sum_{t \in T} \nu_t \delta_t$ where δ_t is the Dirac measure ($\delta_t(x) = 1$ or 0 as $x = t$ or $x \neq t$), the ν_t are the zeros of the function $f(t) = 1 + \sum_{n=1}^{\infty} c_r(t^r + t^{-r})$

on $\mathbb{S}^1$, and T is a symmetric set with $\#T = m-1$. One then verifies that ϑ exists and gives rise to the solutions t in the above formula directly. □

Example 2.16. For $q=2$, $N=71$, we find $m=12$,

$$u = \frac{64\sqrt{2}-70}{70\sqrt{2}-64},$$

and $\vartheta = 0.2562$, and hence

$$c_1 = 0.9698, c_2 = 0.8868, \dots, c_{11} = 0.004227$$

hence

$$g \geq \left\lceil \frac{66.75\sqrt{2}-68}{3-1.935\sqrt{2}} \right\rceil = 101.$$

Hence a curve with genus $g \leq 100$ has $N_3(g) \leq 70 \leq 77$, which is an improvement on the Drinfel'd-Vlăduţ bound.

In sum, when the genus is small in comparison to the size of the field, the Serre-Weil bound is minimal, and as the genus becomes larger, one ascends from the Ihara bound to the Oesterlé bound (which is best possible in the sense that it minimizes the bound coming from Weil's explicit formula).

3. Lower bounds

One is also interested in knowing in general a lower bound on $N_q(g)$. There are two essentially different cases.

3.1. q a square.

In the case that q is a square, we can get quite precise asymptotic information.

Proposition 3.1 (Ihara [11], Tsfasman-Vlăduţ-Zink [24])**.** *If q is a square, then*

$$A(q) \geq \sqrt{q}-1.$$

Sketch of proof. The curves that give this bound are Shimura curves considered over $\mathbb{F}_q$. In the case $q=p^2$, the Shimura curves are the modular curves $X_0(\ell)$ and the proof runs as follows. We can consider the curves $X_0(\ell)$ for $\ell \neq p$, $\ell \equiv 11 \pmod{12}$ over $\mathbb{F}_{p^2}$, parametrizing elliptic curves equipped with an isogeny of degree ℓ.

The genus of this curve is $(\ell+1)/12$ (this can be computed using the Riemann-Hurwitz formula using the quotient map $X_0(\ell) \to X_0(1)$, since ramification happens only at i, ρ, and ∞). There are roughly $(p-1)/12$ supersingular j-invariants in characteristic p all defined over $\mathbb{F}_{p^2}$ (with a slight error due to the cusps and the j-values 0 and 1728). Since there are $\ell+1$ distinct cyclic subgroups of order ℓ of any elliptic curve defined over $\overline{\mathbb{F}}_p$, and when $j \neq 0, 1728$ the automorphism group of E is $\langle \pm 1 \rangle$, these give rise to distinct isogenies. Therefore there are approximately $(\ell+1)(p-1)/12$ rational points on $X_0(\ell)$, so the ratio to its genus is roughly $p-1$. One then verifies that it actually approaches $p-1$ as $\ell \to \infty$. □

Remark 3.2. For a different proof that counts points using Hecke operators, see [16, §5.6]. These curves exceed the Varshamov-Gilbert bound—see [24].

Theorem 3.3. *If q is a square, $A(q) = \sqrt{q} - 1$.*

This follows from the theorem and the upper bound (2.9) above.

Garcia and Stichtenoth constructed an explicit tower X_i over $\mathbb{F}_{q^2}$ which meets this bound.

Theorem 3.4 (Garcia and Stichtenoth [5]). *Start with $X_1 = \mathbb{P}^1_{\mathbb{F}_q^2}$ with coordinate x_1 and define Artin-Schrier covers X_i by*

$$y_{i+1}^q + y_{i+1} = (y_i/x_{i-1})^{q+1}.$$

Then

$$\#X_i(\mathbb{F}_{q^2})/g(X_i) \to q-1 \quad (g \to \infty).$$

Ingredients of proof. One uses the ramification theory of Artin-Schreier extensions (see [22, Proposition VI.4.1]) and the Hurwitz genus formula. The hard steps are determining precisely the places that ramify in each relative extension $K(X_i)/K(X_{i-1})$ and calculating the genus of X_i, which is found to be approximately $q^{i-1}(q-1)$. Then one calculates that

$$\#X_i(\mathbb{F}_{q^2}) \geq (q^2-1)q^{i-1} + 2q$$

by counting places that split completely and those that ramify. This implies the result. □

Such an infinite sequence of distinct curves X_i over $\mathbb{F}_q$ satisfying

$$\frac{\#X_i(\mathbb{F}_q)}{g(X_i)} \to \sqrt{q} - 1$$

as $i \to \infty$ is called an *optimal tower.* Remarkably, at the time of this writing, all such optimal towers have been shown to be *modular*, meaning they arise as parameter spaces for objects such as elliptic curves, Shimura curves, or Drinfel'd modules for sufficiently large i. We refer the reader to [3] for a proof that the tower above is modular and to [2] for more on this fascinating subject.

3.2. q not a square.

In the much more difficult case when q is not a square, Serre has given the following asymptotic result.

Proposition 3.5 (Serre [18]). *There exists a constant $c > 0$ such that*

$$A(q) \geq c \log q$$

for all q.

The proof uses an infinite class field tower of the function field of a hyperelliptic curve over $\mathbb{F}_q$. We also have the following lower bound.

Theorem 3.6 (Elkies-Kresch-Poonen-Wetherell-Zieve [28])**.** *For all q and g,*

$$N_q(g) > ga_q$$

where $a_q > 0$ is a constant that depends only on q (if q is a square, we can take $a_q = (\sqrt{q}-1)/6$).

Ingredients of proof. The proof combines the following. The first is due to Ihara-Serre: for any q, there exist curves X_i of strictly increasing genera g_i such that for all i, $\#X_i(\mathbb{F}_q) \geq b_q g_i$ where $b_q > 0$ and $\lim_{i\to\infty} g_{i+1}/g_i < \infty$. The second is: for any $X/\mathbb{F}_q$ and any $h > 4g_X$, there exists a curve $X'/\mathbb{F}_q$ such that $g_{X'} = h$ and $\#X'(\mathbb{F}_q) \geq \#X(\mathbb{F}_q)$. It is clear that these imply the theorem. To prove the second, one proves only that there exists a degree two cover $X' \to X$ with $h = g_{X'}$ for any $h > 4g_X$. □

4. Determination of $N_q(g)$

Actual values of $N_q(g)$ for small values of q and g are interesting. Lower bounds are given by the best curve we know, and with luck, this attains one of the upper bounds above; otherwise we obtain an interval in which $N_q(g)$ lies.

4.1. **Maximal curves.** First, we seek out curves that attain the Weil bound. A curve X is called *maximal* if it attains the Weil bound (2.2), i.e.

$$\#X(\mathbb{F}_q) = q + 1 + 2g\sqrt{q}.$$

It follows that if X is maximal, then q is a square (so that the right-hand side is an integer) and $g \leq (q - \sqrt{q})/2$ by Remark 2.7.

Example 4.1. The *Hermitian curve*

$$X : x^{q+1} + y^{q+1} = z^{q+1}$$

has genus $q(q-1)/2$ and has q^3+1 points over $\mathbb{F}_{q^2}$. (See [22, Lemma VI.4.4] for the properties of the function field of X.) This curve is also birational to $X : y^q z + yz^q = x^{q+1}$, which clarifies many of the properties of the curve; see [22, Example VI.4.3].

We may count points as follows: for any $a \in \mathbb{F}_{q^2}^*$, we have $a^{q+1} = N(a) \in \mathbb{F}_q$; the norm map $N : \mathbb{F}_{q^2}^* \to \mathbb{F}_q^*$ is surjective and hence

$$\#\ker N = (q^2-1)/(q-1) = q+1.$$

Therefore for a choice of $a \neq 0$, there are $q+1$ choices for b such that $N(a) + N(b) = a^{q+1} + b^{q+1} = 0$, and there is the unique value $c = 0$ such that $(a : b : c) \in X$. A similar count holds when $a = 0$. Accounting for scalar multiples, and we find

$$\#X(\mathbb{F}_{q^2}) = \frac{2q^2(q+1) + q^2\left(q^2 - (q+1)\right)(q+1)}{q^2-1} = q^3 + 1.$$

Since

$$q^2 + 1 + 2g\sqrt{q^2} = q^2 + 1 + q^2(q-1) = q^3 + 1,$$

this curve attains the Hasse-Weil bound.

It can be shown that maximal curves only occur for specific genera.

Theorem 4.2 (Fuhrmann-Garcia-Torres [4]). *If X is a maximal curve of genus g over $\mathbb{F}_q$ (where q is a square) then $g = (q - \sqrt{q})/2$ or $g \leq (\sqrt{q}-1)^2/4$. Moreover, if q is odd and*

$$(\sqrt{q}-1)(\sqrt{q}-2)/4 < g \leq (\sqrt{q}-1)^2/4$$

then $g = (\sqrt{q}-1)^2/4$.

Ingredients of proof. If X is a maximal curve and X' a curve dominated by X then X' is also a maximal curve, since the Jacobian of X' is an isogeny factor of the Jacobian of X. For a maximal curve X the action of Frobenius on the Jacobian is by $-\sqrt{q}$. Choose a $\mathbb{F}_{q^2}$-rational point P_0 on X and map X to the Jacobian $\mathrm{Jac}(X)$ by $P \mapsto [P - P_0]$. If $F(P)$ is the Frobenius image of P on X then

$$-\sqrt{q}[P - P_0] = [F(P) - P_0]$$

in the Jacobian, so for any point P the divisor $\sqrt{q}P + F(P)$ is linearly equivalent to $(\sqrt{q}+1)P_0$ and this gives a canonically defined linear system on such curves. The result then follows by looking at this linear system, as it gives a bound on g. □

We refer the reader [23] for an introduction to Stöhr-Voloch theory which concerns the geometric consequences of the arithmetic property of having many rational points.

4.2. **Fixed g.** Serre [18] began the systematic study of $N_q(g)$ for small q and g. For $g = 0$ (the projective line), we have $N_q(0) = q + 1$.

For $g = 1$ (elliptic curves), we have the following classical result.

Proposition 4.3 (Hasse-Deuring). *We have*

$$N_q(1) = \begin{cases} q + \lfloor 2\sqrt{q} \rfloor, & \textit{if } q = p^n,\ n \geq 3,\ n \textit{ odd, and } p \mid \lfloor 2\sqrt{q} \rfloor; \\ q + 1 + \lfloor 2\sqrt{q} \rfloor, & \textit{otherwise.} \end{cases}$$

The smallest exceptional case is $q = 128 = 2^7$; we have $N_{128}(1) = 150$, $\lfloor 2\sqrt{q} \rfloor = 22$; other examples are $q = 2^{11}, 2^{15}, 3^7, 5^9$.

For curves of genus $g = 2$, Serre proved the following result.

Theorem 4.4 (Serre [19]). *If q is a square and $q \neq 4, 9$ then*

$$N_q(2) = q + 1 + 4\sqrt{q}.$$

Furthermore, we have $N_4(2) = 10$ and $N_9(2) = 20$.

If q is not a square, then

$$N_q(2) = \begin{cases} q + 1 + 2\lfloor 2\sqrt{q} \rfloor, & \textit{if } q \textit{ is nonspecial;} \\ q + 2\lfloor 2\sqrt{q} \rfloor, & \textit{if } q \textit{ is special and, } 2\sqrt{q} - \lfloor 2\sqrt{q} \rfloor > (\sqrt{5}-1)/2; \\ q - 1 + 2\lfloor 2\sqrt{q} \rfloor, & \textit{otherwise.} \end{cases}$$

We deem q to be *special* if $q = p^e$ and $p \mid \lfloor 2\sqrt{q} \rfloor$ or if there exist solutions to one of the equations

$$q = x^2 + 1,\ q = x^2 + x + 1,\ q = x^2 + x + 2$$

with $x \in \mathbb{Z}$. For example, if $e = 1$, i.e. $q = p$, then p is special iff $p = x^2 + 1$ or $p = x^2 + x + 1$. When $e \geq 3$, it can be shown that the Diophantine condition is satisfied only for $q = 7^3$ or $q = 2^3, 2^5, 2^{13}$.

Ingredients of proof, [20]. We endeavor to construct a curve of genus 2 having many rational points. One begins with an elliptic curve E having $q + 1 + \lfloor 2\sqrt{q} \rfloor$ points with $\mathrm{End}(E) = R = \mathbb{Z}[\pi]$ with π the Frobenius map. The ring R is an order in a quadratic imaginary field of discriminant $D = \lfloor 2\sqrt{q} \rfloor^2 - 4q < 0$. Find a nondegenerate Hermitian form P over R, projective of rank 2, positive definite, and indecomposable (which exists when $D < -7$). Then $A = P \otimes_R E$ is an abelian variety of dimension 2, isogeneous to $E \times E$, equipped with an indecomposable, principal polarization by P. There is a correspondence between curves of genus 2 and abelian varieties of dimension 2 equipped with such a polarization, and therefore A is the Jacobian of a curve of genus 2 with $\#X(\mathbb{F}_q) = q + 1 + 2\lfloor 2\sqrt{q} \rfloor$.

The exceptional cases occur when the form is decomposable, which translates into an occurence of p satisfying one of the three given Diophantine conditions. □

4.3. **Other examples.** For small genera, certain values of $N_q(g)$ can be eliminated by listing all of the possibilities of the zeta function and showing that zeta functions in this list imply a decomposition of the Jacobian as a product of principally polarized abelian varieties, which contradicts the irreducibility of the theta divisor of the curve. Sometimes one can rule out that $N_q(g)$ meets the Serre bound by arguments like Galois descent, which works for $q = 27$, $g = 3$ and $q = 8$, $g = 4$. (For a complete list of references, see [25].)

Tables of the best lower and upper bounds for $g \leq 50$ and $q = 2^m, 3^m$ for small m have been collected by van der Geer and van der Vlugt [25]. The most updated tables can be found at `http://www.science.uva.nl/~geer/`. Once one has good upper bounds, one would like to show that these bounds are met by some curve, therefore one needs good methods of constructing curves; these methods include using class field theory, quotients and covers of classical curves (like the Hermitian, Fermat, and modular curves) and other curves with many rational points, fibre products of Artin-Schreier curves, rank 1 Drinfel'd modules, Kummer covers, exhaustive computer search, and many others. We refer the reader to [25] for an overview.

References

[1] Ian F. Blake, *Curves with many points and their applications*, Applied alebgra, algebraic algorithms and error correcting codes: 13th International Symposium,

AAEECC-13 (Honolulu, Hawaii), Lect. Notes in Comp. Sci., vol. 1719, 1999, 55–64.

[2] Noam D. Elkies, *Explicit modular towers*, Proceedings of the Thirty-Fifth Annual Allerton Conference on Communication, Control and Computing (1997), T. Basar, A. Vardy, eds., 1998, 23–32, also available at `arXiV:math.NT/0103107`.

[3] Noam D. Elkies, *Explicit towers of Drinfeld modular curves*, Progress in Math. **202** (2001), 189–198.

[4] Rainer Fuhrmann, Arnaldo Garcia, and Fernando Torres, *On maximal curves*, J. Num. Theory **67** (1997), 29–51.

[5] Arnolda Garcia and Henning Stichtenoth, *A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vladut bound*, Invent. Math. **121** (1995), 211–222.

[6] V.D. Goppa, *Codes on algebraic curves*, Sov. Math. Dokl. **24** (1981), 170–172.

[7] V.D. Goppa, *Geometry and codes*, Mathematics and its applications, Soviet series, vol. 24, Kluwer Academic Publishers: Dordrecht, Netherlands, 1988,

[8] Robin Hartshorne, *Algebraic geometry*, Springer-Verlag, New York, 1977.

[9] Johan P. Hansen, *Deligne-Lusztig varieties and group codes*, Coding theory and algebraic geometry, Springer: Berlin, 1992, 63–81.

[10] Søren Have Hansen, *Rational points on curves over finite fields*, Lect. Notes Ser., Aarhus Univ. Mat. Institute, 1995.

[11] Yasutaka Ihara, *Some remarks on the number of rational points of algebraic curves over finite fields*, J. Fac. Sci. Univ. Tokyo Sect. IA Math. **28** (1981), no. 3, 721–724.

[12] Kristin Lauter, *Deligne-Lusztig curves as ray class fields*, preprint available via search engine at `http://www.mpim-bonn.mpg.de/cgi-bin/preprint/preprint_search.pl`.

[13] Kristin Lauter and René Schoof, Using class field theory to construct curves with many points, Lectures given at Arizona Winter School: Arithmetic of function fields, February 2000.

[14] F.J. MacWilliams and N.J.A. Sloane, *Theory of error-correcting codes*, Amsterdam, North-Holland, 1977.

[15] Yu. I. Manin, *What is the maximum number of points on a curve over* $\mathbb{F}_2$*?*, J. Fac. Sci. Univ. Tokyo Sect. IA Math. **28** (1981), no. 3, 715–720.

[16] Carlos Moreno, *Algebraic curves over finite fields*, Cambridge tracts in mathematics, vol. 97, Cambridge University Press: Cambridge, 1991.

[17] R. Schoof, *Algebraic curves and coding theory*, UTM 336, Univ. of Trento, 1990.

[18] Jean-Pierre Serre, *Sur le nombre des points rationnels d'une courbe algébrique sur un corps fini*, C. R. Acad. Sci. Paris, I **296** (1983), 397–402. (*Oeuvres*, vol. III, no.128, 658–663.)

[19] Jean-Pierre Serre, *Nombres de points des courbes algébriques sur* $\mathbb{F}_q$, Sém. de Théorie des Nombres de Bordeaux (1982/1983) exp. no. 22. (*Oeuvres*, vol. III, no. 129, 664–668.)

[20] Jean-Pierre Serre, Résumé des cours de 1983–1984, Annuaire du Collège de France (1984), 79–83. (*Oeuvres*, vol. III, no. 132, 701–705.)

[21] J.H. Silverman, *The arithmetic of elliptic curves*, Berlin: Springer, 1994.

[22] Henning Stichtenoth, *Algebraic function fields and codes*, Springer: Berlin, 1993.

[23] Fernando Torres, *The approach of Stöhr-Voloch to the Hasse-Weil bound with applications to optimal curves and plane arcs*, preprint 2000, `arXiV:math.AG/0011091`.

[24] M.A. Tsfasman, S.G. Vlăduţ, and Th. Zink, *Modular curves, Shimura curves, and Goppa codes, better than Varshamov-Gilbert bound*, Math. Nachr. **109** (1982), 21–28.

[25] Gerard van der Geer and Marcel van der Vlugt, *Tables of curves with many points*, Math. Comp. **69** (2000), 797–810.

[26] S.G. Vlăduţ and V.G. Drinfel'd, *Number of points of an algebraic curve*, Functional Anal. Appl. **17** (1983), no. 1, 53–54.

[27] Daqing Wan, *Computing zeta functions over finite fields*, Contemporary Math. **225** (1999), 131–141.

[28] Michael Zieve, Curves of large genus with many points, Lectures given at Arizona Winter School: Arithmetic of function fields, February 2000.

E-mail address: jvoight@math.berkeley.edu

Department of Mathematics, University of California, Berkeley; Berkeley, California, 94720, USA