1. Introduction.

Let E be an elliptic curve over a finite field $\mathbf{F}_q$. Then E is a smooth cubic in $\mathbf{P}_2$. It can be given by a Weierstrass equation, an affine version of which is

$$Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6, \qquad \text{with } a_1, a_2, a_3, a_4, a_6 \in \mathbf{F}_q.$$

The unique point at infinity is the neutral element of the group law. It is denoted by ∞. We denote the affine curve itself by E^0. The zeta-function $Z_E(T)$ of E is the power series defined by

$$Z_E(T) \;=\; \sum_{D\geq 0} T^{\deg D} \qquad \text{in } \mathbf{Z}[[T]].$$

Here D runs over the effective divisors of E that are defined over $\mathbf{F}_q$. In this note we prove two theorems concerning $Z_E(T)$.

Theorem 1.1. *Let E be an elliptic curve over $\mathbf{F}_q$. Then the power series $Z_E(T)$ is equal to the rational function*

$$Z_E(T) \;=\; \frac{1 - \tau T + qT^2}{(1-T)(1-qT)},$$

where τ is given by the formula $\#E(\mathbf{F}_q) = q + 1 - \tau$.

And we prove Hasse's Theorem:

Theorem 1.2. *Let E be an elliptic curve over $\mathbf{F}_q$. Then the complex zeroes of the reciprocal $T^2 - \tau T + q$ of the numerator of its zeta function have absolute value $\sqrt{q}$.*

This means that the function $\zeta_E(s) = Z_E(q^{-s})$ of the complex variable s admits a meromorphic continuation to $\mathbf{C}$ and that its zeroes have real part equal to $\frac{1}{2}$. Therefore Theorem 1.2 is the analogue of the Riemann Hypothesis for the curve E. Since $\#E(\mathbf{F}_q) = q + 1 - \tau$, it implies $|\tau| \leq 2\sqrt{q}$ and hence the inequalities

$$q + 1 - 2\sqrt{q} \;\leq\; \#E(\mathbf{F}_q) \;\leq\; q + 1 + 2\sqrt{q}.$$

Theorem 1.2 was proved by H. Hasse in 1933. Our approach is elementary and follows a method invented by S.A. Stepanov around 1969. We only make use of the Weierstrass equation and the group law.

2. Rationality of the zeta function.

In this section we prove Theorem 1.1. First we review some properties of elliptic curves. Let E be an elliptic curve given by a Weierstrass equation as in the introduction. The ring R of functions on E without poles outside ∞ is the $\mathbf{F}_q$-algebra generated by the functions X and Y. So we have

$$R = \mathbf{F}_q[X,Y]/(Y^2 + a_1XY + a_3Y - X^3 - a_2X^2 - a_4X - a_6).$$

Every element $f \in R$ has the form $g(X)+Yh(X)$ for unique polynomials $g,h \in \mathbf{F}_q[X]$. For every non-zero $f \in R$, let $\deg f$ denote the order of the pole of f at ∞. We have $\deg X = 2$ and $\deg Y = 3$. In general, for $f = g(X)+Yh(X)$ with $g,h \in \mathbf{F}_q[X]$ polynomials of degrees d,e respectively, one has $\deg f = \max(2d, 3+2e)$. In particular, R contains no functions f with $\deg f = 1$. We call $f \in R$ *monic* if the coefficient of its highest degree term is equal to 1. Any $f \in R$ has, counting multiplicities, precisely $\deg f$ zeroes on E^0. Indeed, if $f = g(X)+Yh(X)$ as above, then the equation obtained by substituting $Y = -g(X)/h(X)$ in the Weierstrass equation has degree $\deg f$ in X.

A *divisor* is a formal sum of points of E that have coordinates in a fixed algebraic closure $\overline{\mathbf{F}}_q$. It is said to be defined over $\mathbf{F}_q$, if it is fixed by the Galois group of $\overline{\mathbf{F}}_q$ over $\mathbf{F}_q$. The principal divisor associated to an element g of the function field $\overline{\mathbf{F}}_q(E)$ is denoted by (g). If the divisor (g) is defined over $\mathbf{F}_q$, then there exists a function $g' \in \mathbf{F}_q(E)$ with $(g') = (g)$. For two divisors D, D' of E we write $D \sim D'$ if $D - D'$ is principal, i.e. if $D - D' = (g)$ for some function g on E.

Lemma 2.1. *Let E be an elliptic curve and let P, Q be two points on E. Then*

$$P + Q \sim (P+Q) + \infty.$$

Here the leftmost and rightmost plus signs indicate addition of divisors, while the one in the middle refers to the group law on E.

Proof. The quotient of the equations of the chords or tangents used to add the points P and Q is a function g on E whose divisor is precisely $P + Q - (P+Q) - \infty$. Moreover, g is defined over the same field as P and Q.

Proposition 2.2. *Let E be an elliptic curve over $\mathbf{F}_q$ and let D be a divisor of E of degree d. Then we have*

$$D \sim P + (d-1)\infty, \qquad \text{for a unique point } P \in E(\overline{\mathbf{F}}_q).$$

Moreover, if D is defined over $\mathbf{F}_q$, then so is P and the divisor $D - (d-1)\infty - P$ is the divisor of a function in $\mathbf{F}_q(E)$.

Proof. Let $D = \sum_Q n_Q Q$ for certain integers n_Q. Let P be the point on E that one obtains by adding the points Q, with multiplicities n_Q, on the curve E using the chord and tangent group law. The point P is defined over the field of definition of D. Applying Lemma 2.1 inductively gives the relation

$$D \sim P + (d-1)\infty,$$

as required.

Example 2.3. We first compute the zeta function of the projective line $\mathbf{P}_1$ over $\mathbf{F}_q$ and then deal in a similar way with zeta functions of elliptic curves E. The zeta function of $\mathbf{P}_1$ over $\mathbf{F}_q$ is defined by

$$Z_{\mathbf{P}_1}(T) = \sum_{D\geq 0} T^{\deg D} \qquad \text{in } \mathbf{Z}[[T]],$$

where D runs over the effective divisors of $\mathbf{P}_1$ that are defined over $\mathbf{F}_q$. Since every divisor is a sum of points, we have

$$Z_{\mathbf{P}_1}(T) = \prod_P \frac{1}{1-T^{\deg P}}.$$

Here P runs over the $\mathrm{Gal}(\overline{\mathbf{F}}_q/\mathbf{F}_q)$-conjugacy classes of points of $\mathbf{P}_1$. The zeta function of the affine line $\mathbf{A}^1$ is obtained by omitting the factor $1/(1-T)$ corresponding to the point at infinity. So we have

$$Z_{\mathbf{A}^1}(T) = \sum_{D\geq 0} T^{\deg D} \qquad \text{in } \mathbf{Z}[[T]],$$

where D runs over the effective divisors of $\mathbf{A}^1$. Since the ring $\mathbf{F}_q[X]$ is a principal ideal domain, every divisor $D \geq 0$ of $\mathbf{A}^1$ that is defined over $\mathbf{F}_q$ is the divisor of a unique monic polynomial g in $\mathbf{F}_q[X]$. Moreover, the degree of D is equal to the degree of g. We can therefore compute the zeta function of $\mathbf{A}^1$ by counting polynomials. We find

$$Z_{\mathbf{A}^1}(T) = \sum_{d\geq 0} c_d T^d = \sum_{d\geq 0} q^d T^d = \frac{1}{1-qT}.$$

Here c_d denotes the number of effective divisors of $\mathbf{A}^1$ of degree d. Since the number of monic degree d polynomials in $\mathbf{F}_q[X]$ is q^d, we have $c_d = q^d$. Going back to the projective line $\mathbf{P}_1$, we obtain the following fomula for the zeta function of $\mathbf{P}_1$ over $\mathbf{F}_q$.

$$Z_{\mathbf{P}_1}(T) = \frac{1}{(1-T)(1-qT)}.$$

This completes the computation of the zeta function of $\mathbf{P}_1$.

Proof of Theorem 1.1. We determine the zeta function of an elliptic curve E over $\mathbf{F}_q$ in a similar way. Recall that E^0 is the affine curve that is obtained by removing the point ∞ from E. We first determine the zeta-function of E^0. This means that we must count effective divisors on E^0 that are defined over $\mathbf{F}_q$. These are simply divisors on E of the form $\sum_P n_P P$ with $n_P \geq 0$ for all P in E for which $n_P = n_{P'}$ whenever P and P' are conjugate points. Moreover, we have $n_\infty = 0$.

The only effective divisor of E^0 of degree 0 is the divisor 0. The effective divisors over $\mathbf{F}_q$ of degree 1 are precisely the points in $E(\mathbf{F}_q) - \{\infty\}$. Denoting $\#E(\mathbf{F}_q)$ by h, there are $h-1$ of them.

Let D be an effective divisor on E^0 of degree $d > 1$. By Proposition 2.2 there exists a unique point $P \in E(\mathbf{F}_q)$ for which we have $-D \sim (-d-1)\infty + P$ on E. Equivalently, there exists a function $f \in E(\mathbf{F}_q)$ whose divisor on E is $D + P - (d+1)\infty$. The function f is unique up to a non-zero constant. Since D is effective, f is contained in the ring R.

There are two cases. If $P = \infty$, the function f has degree d. Conversely, for every $f \in R$ of degree d, the divisor $(f) - P + (d+1)\infty$ is effective. There are q^{d-1} monic functions f with this property.

If $P \neq \infty$, the function f has degree $d+1$ and vanishes in P. Conversely, for any f having these properties, the divisor $(f) - P + (d+1)\infty$ is effective. For each point $P \neq \infty$ there are q^{d-1} monic functions f with this property.

Counting all functions, we see that there are $q^{d-1} + (h-1)q^{d-1} = hq^{d-1}$ effective divisors on E^0 of degree d. This computation shows that

$$Z_{E^0}(T) \;=\; 1 + (h-1)T + \sum_{d\geq 2} hq^{d-1}T^d \;=\; \frac{1 + (h-q-1)T + qT^2}{1-qT}.$$

The zeta function of E is obtained from the one of E^0 in the same way the zeta function of $\mathbf{P}_1$ is obtained from the one of $\mathbf{A}^1$. In order to take into account the point at infinity, we multiply $Z_{E^0}(T)$ by the factor $1/(1-T)$. This gives

$$Z_E(T) \;=\; \frac{1 - \tau T + qT^2}{(1-T)(1-qT)},$$

where $\tau = q + 1 - h$. This proves Theorem 1.1.

3. An upper bound.

In this section we obtain an upper bound for the number of points of an elliptic curve E over a finite field. This is the key ingredient in the proof of Theorem 1.2. Our method is due to S.A. Stepanov.

We introduce some notation. Recall that E is given by a Weierstrass equation and that R is the $\mathbf{F}_q$-algebra generated by the functions X and Y. For $a \geq 0$ let L_a denote the $\mathbf{F}_q$-vector space

$$L_a = \{f \in R : \deg f \leq a\}.$$

Since R does not contain any functions $f \in R$ with $\deg f = 1$, the space L_a consists only of constant functions when $a = 0$ or 1 and therefore has dimension 1. In general we have the following. Put $e_1 = 1$ and

$$e_{2i} = X^i \quad \text{and} \quad e_{2i+1} = X^{i-1}Y \quad \text{for } i \geq 1.$$

Then e_i has degree i for $i \geq 1$.

Lemma 3.1. *For $a \geq 1$, the monomials e_i with $i \leq a$ are an $\mathbf{F}_q$-basis for L_a. In particular, L_a has $\mathbf{F}_q$-dimension a.*

Proof. The monomials e_i certainly generate L_a. On the other hand, the orders of their poles at ∞ are all distinct. Therefore they are linearly independent and hence form a basis of L_a. This proves the lemma. Note that the fact that $\dim L_a = a$ also easily follows from the Riemann-Roch Theorem.

For $a \geq 1$ the set $L_a^q = \{f^q : f \in L_a\}$ is an $\mathbf{F}_q$-vector space of dimension $a = \dim L_a$. Indeed, the map $f \mapsto f^q$ is an $\mathbf{F}_q$-linear bijection $L_a \leftrightarrow L_a^q$.

Lemma 3.2. *Let $a, b \geq 1$ and let $L_a^q L_b$ denote the $\mathbf{F}_q$-vector space generated by the functions $f^q g$ where $f \in L_a$ and $g \in L_b$. Then we have*
(a) $\dim L_a^q L_b \leq aq + b$*;*
(b) $\dim L_a^q L_b \leq ab$ *with equality if $b < q$.*

Proof. Part (a) follows from the fact that $L_a^q L_b \subset L_{aq+b}$ and Lemma 3.1. The inequality of part (b) follows from the fact that the functions $e_i^q e_j$ with $1 \leq i \leq a$ and $1 \leq j \leq b$ generate $L_a^q L_b$. To get equality when $b < q$, we observe that

$$\deg e_i^q e_j = q\deg e_i + \deg e_j = iq + j$$

Since we have $j \leq b < q$, the degrees $\deg e_i^q e_j$ are all distinct. So any $\mathbf{F}_q$-linear combination $\sum_{i,j} \lambda_{ij} e_i^q e_j$ that is zero, necessarily has $\lambda_{ij} = 0$ for every i, j. This proves that the functions $e_i^q e_j$ are independent. Therefore the dimension of $L_a^q L_b$ is equal to ab. This proves the lemma.

From now on we assume that $a, b \geq 1$ with $b < q$. Lemma 3.1 implies that the $\mathbf{F}_q$-linear map

$$\vartheta : L_a^q L_b \longrightarrow L_a L_b^q$$

given by

$$e_i^q e_j \mapsto e_i e_j^q, \qquad \text{for } 1 \leq i \leq a \text{ and } 1 \leq j \leq b,$$

is well defined.

The following proposition is the key ingredient in the proof of Theorem 3.4.

Proposition 3.3. *Let $a, b \geq 1$ with $b < q$. If the map ϑ is not injective, then*

$$\#E(\mathbf{F}_{q^2}) \leq aq + b + 1.$$

Proof. Every function $F \in \ker \vartheta$ vanishes on $E(\mathbf{F}_{q^2}) - \{\infty\}$. Indeed, let $F = \sum \lambda_{ij} e_i^q e_j$ for certain $\lambda_{ij} \in \mathbf{F}_q$ and let $P \in E(\mathbf{F}_{q^2}) - \{\infty\}$. Then

$$F(P)^q = \sum \lambda_{ij} e_i^{q^2}(P) e_j^q(P) = \sum \lambda_{ij} e_i(P) e_j^q(P) = (\sum \lambda_{ij} e_i e_j^q)(P) = \vartheta(F)(P) = 0,$$

which is zero when $F \in \ker \vartheta$. The second equality follows from the fact that $P \in E(\mathbf{F}_{q^2})$ so that $f^{q^2}(P) = f(P)$ for every function $f \in R$.

Since ϑ is not injective, there exists a non-zero F in $\ker \vartheta$. Therefore we obtain the following estimate.

$$\#E(\mathbf{F}_{q^2}) - 1 \leq \#\{\text{zeroes of } F\} = \deg(F) \leq aq + b.$$

The rightmost inequality follows from Lemma 3.2 (a). This proves the proposition.

Theorem 3.4. *Let E be an elliptic curve defined over $\mathbf{F}_q$ and suppose that $q \geq 5$. Then we have*

$$\#E(\mathbf{F}_{q^2}) \;\leq\; q^2 + 3q.$$

Proof. The map ϑ defined above cannot be injective if $a, b \geq 1$ have the property that

$$\dim L_a^q L_b \;>\; \dim L_a L_b^q.$$

Since $b < q$, Lemma 3.2 (b) implies that $L_a^q L_b$ has dimension ab. Lemma 3.2 (b) cannot be applied to $L_a L_b^q$. In some sense this is the point of the proof. But by Lemma 3.2 (a) we know that $L_a L_b^q$ has dimension $\leq a + bq$. Therefore the map ϑ is *not* injective when

$$ab \;>\; a + bq.$$

In order to deduce a sharp estimate from Proposition 3.3, we choose a as small as possible. Since the inequality $ab > a + bq$ must be satisfied, the minimal choice for a is $a = q + 2$. Once a is chosen, we can take $b = q - 1$, at least for $q \geq 5$. With these choices the quantity $aq + b + 1$ in Proposition 3.3 becomes $(q + 2)q + q - 1 + 1 = q^2 + 3q$, as required.

4. The Riemann Hypothesis.

Let E be an elliptic curve over $\mathbf{F}_q$. In this section we prove that the complex zeroes of the numerator of its zeta function have absolute value $1/\sqrt{q}$. The key ingredient is the inequality af Theorem 3.4. First we use the proof of Theorem 3.4 to obtain a lower bound for $\#E(\mathbf{F}_{q^2})$.

Proposition 4.1. *Let E be an elliptic curve over $\mathbf{F}_q$ and suppose that $q \geq 5$. Then we have*

$$\#E(\mathbf{F}_{q^2}) \;>\; q^2 - 3q$$

Proof. Let Ω denote the set of points (x, y) of $E^0(\overline{\mathbf{F}}_q)$ for which $x \in \mathbf{F}_{q^2}$. For every $x \in \mathbf{F}_{q^2}$ there are at most two points $(x, y) \in \Omega$. If (x, y) is one such point, then $(x, \overline{y})$ where $\overline{y} = -y - a_1 x - a_3$, is the other. We have

$$\#\Omega \;=\; 2q^2 - r.$$

where r is the number of values of x for which $y = \overline{y}$. We have $r \leq 3$.

The automorphism σ of $\overline{\mathbf{F}}_q$ given by $\sigma(t) = t^{q^2}$ also acts on Ω. It maps a point $(x, y) \in \Omega$ to $(\sigma(x), \sigma(y)) = (x^{q^2}, y^{q^2}) = (x, y^{q^2})$. It follows that either $\sigma(y) = y$ or $\sigma(y) = \overline{y}$. Therefore have

$$\Omega = \Omega^+ \cup \Omega^-,$$

where $\Omega^+ = \{(x, y) \in \Omega : \sigma(y) = y\}$ and $\Omega^- = \{(x, y) \in \Omega : \sigma(y) = \overline{y}\}$. The intersection $\Omega^+ \cap \Omega^-$ consists of the r points (x, y) for which $y = \overline{y}$.

Clearly Ω^+ is the set $E(\mathbf{F}_{q^2}) - \{\infty\}$. Theorem 3.4 provides an estimate for its size. In this section we use the method of section 3 to obtain an estimate of the size of the set Ω^-. Let a, b be as in the proof of Theorem 3.4. Note that the spaces L_a and L_b are preserved by the automorphism of R given by $f(X,Y) \mapsto f(X, -Y - a_1X - a_3)$. Consider the $\mathbf{F}_q$-linear map

$$\vartheta' : L_a^q L_b \longrightarrow L_a L_b^q$$

defined by

$$e_i^q e_j \ \mapsto \ \overline{e_i} e_j^q.$$

Every function $F \in \ker \vartheta'$ vanishes on the set W. Indeed, let $F = \sum \lambda_{ij} e_i^q e_j$ for certain $\lambda_{ij} \in \mathbf{F}_q$ and let $P \in W$.

$$F(P)^q = \sum \lambda_{ij} e_i^{q^2}(P) e_j^q(P) = \sum \lambda_{ij} \overline{e}_i(P) e_j^q(P) = (\sum \lambda_{ij} \overline{e}_i f_j^q)(P) = \vartheta'(F)(P) = 0,$$

and hence $F(P) = 0$. Therefore we can draw the same conclusion as in the previous section. We have

$$\#\Omega^- \ \leq \ q^2 + 3q.$$

and hence

$$\begin{aligned} \#E(\mathbf{F}_{q^2}) - 1 &= \#\Omega^+, \\ &= \#\Omega - \#\Omega^- + \#(\Omega + \cap \#\Omega^-), \\ &\geq (2q^2 - r) - (q^2 + 3q) + r, \\ &\geq q^2 - 3q. \end{aligned}$$

as required.

Let $1 - \tau T + qT^2$ be the numerator of the zeta function of E and let π and π' be the complex zeroes of the reciprocal polynomial $T^2 - \tau T + q$.

Lemma 4.2. *For every $d \geq 1$, we have*

$$\#E(\mathbf{F}_{q^d}) \ = \ q^d + 1 - \pi^d - \pi'^d.$$

Proof. By Theorem 1.1 we have

$$Z_E(T) = \frac{1 - \tau T + qT^2}{(1-T)(1-qT)}.$$

Combining this with the identity

$$Z_E(T) = \sum_{D \geq 0} T^{\deg D} \ = \prod_P \frac{1}{1 - T^{\deg P}},$$

we obtain

$$\frac{(1-\pi T)(1-\pi' T)}{(1-T)(1-qT)} = \prod_{d \geq 1} (1 - T^d)^{-a_d}.$$

For $d \geq 1$ we write here a_d for the number of points on E of degree d up to conjugacy. For every $e \geq 1$ we have $\#E(\mathbf{F}_{q^e}) = \sum_{d|e} da_d$. Taking the logarithmic derivative of this identity, expanding the geometric series and comparing coefficients shows that we have $q^e + 1 - \pi^e - \pi'^e = \sum_{d|e} da_d =$ for every $d \geq 1$. This proves the Lemma.

Theorem 4.3. *The complex zeroes π and π' of the polynomial $T^2 - \tau T + q$ have absolute value $\sqrt{q}$. In particular $\pi' = \overline{\pi}$.*

Proof. Lemma 4.2, Theorem 3.4 and Proposition 4.1 provide us with the inequalities

$$q^d - 3q^{d/2} \leq q^d + 1 - \pi^d - \pi'^d \leq q^d + 3q^{d/2}, \qquad \text{for even } d \geq 0.$$

Therefore we have

$$|\pi^d + \pi'^d| \leq 3q^{d/2}, \qquad \text{for even } d \geq 0.$$

Suppose $|\pi| > \sqrt{q}$. Since $\pi\pi' = q$, we have $|\pi'| < \sqrt{q}$. Then the absolute values of both $1 + (\pi'/\pi)^d$ and $(\pi'/\pi)^d$ go to zero as $d \to \infty$. This is impossible. Therefore we have $|\pi| \leq \sqrt{q}$. By symmetry also $|\pi'| \leq \sqrt{q}$. This implies $|\pi| = |\pi'| = \sqrt{q}$, as required.

The inequalities of Theorem 3.4 and Proposition 4.1 have only been proved for $q \geq 5$. However, when $q < 5$, we have $q^d > 5$ for $d \geq 3$. This implies that we still have the inequality for even degrees $d \geq 6$. Therefore the argument involving $d \to \infty$ is not affected and the conclusion is the same for $q < 5$. This proves the theorem.