Gauss: the Last Entry
Frans Oort
(1) **Introduction.** Carl Friedrich Gauss (1777-1855) kept a mathematical diary (from 1796). The last entry he wrote was on 7 July 1814. A remarkable short statement.

We present one of the shortest examples of a statement with a visionary impact: we discuss this expectation by Gauss. His idea preludes developments only started more than a century later.

I have two goals:
– explain the visionary insight by Gauss and show the connection with 20-th century mathematics,
– with an emphasis on "the flow of mathematics": early ideas becoming a theory once understood; this we will see in the sequence:
Tagebuch Gauss → The Riemann hypothesis in positive characteristic → the Weil conjectures.
In this talk I will explain the diary note by Gauss.

We give a short proof, using methods, developed by Hasse, Weil and many others. Of course this is history upside down: instead of seeing the Last Entry as a prelude to modern developments, we give a 20-th century proof of this 19-th century statement.

We will discuss also:
Lemniscate transcendental functions $\rightarrow$ Weierstrass $\wp$ functions $\rightarrow$ Koebe / Poincaré uniformization of algebraic curves over $\mathbb{C}$. 
The last entry Gauss wrote in his notebook was on 7 July 1814.
A remarkable short statement.

Here is the text:

Observatio per inductionem facta gravissima
theoriam residuorum biquadraticorum cum
functionibus lemniscaticis elegantissime
nectens. Puta, si $a + bi$ est numerus primus,
$a - 1 + bi$ per $2 + 2i$ divisibilis, multitudo
omnium solutionum congruentiae
$1 = xx + yy + xxyy \pmod{a + bi}$ inclusis
$x = \infty, \ y = \pm i, \ x = \pm i, \ y = \infty$ fit
$= (a - 1)^2 + bb$. 
The text of the “Tagebuch” was rediscovered in 1897 and edited and published by Felix Klein (Math. Annalen 1903). In the collected works by Gauss, in Volume X₁ (1917) we find a facsimile of the original text.

A most important observation made by induction which connects the theory of biquadratic residues most elegantly with the lemniscatic functions. Suppose, if \( a + bi \) is a prime number, \( a - 1 + bi \) divisible by \( 2 + 2i \), then the number of all solutions of the congruence

\[
1 = xx + yy + xxyy \pmod{a + bi} \text{ including } x = \infty, \ y = \pm i; \ x = \pm i, \ y = \infty, \text{ equals } (a - 1)^2 + bb.
\]
Remarks. In the original we see that Gauss indeed used the notation $xx$, as was usual in his time. In his *Disquisitiones Arithmeticae* for example we often see that $x^2$ and $x'x'$ are used in the same formula.

The terminology “Tagebuch” used, with subtitle “Notizenjournal”, is perhaps better translated by “Notebook” in this case. In the period 1796 - 1814 we see 146 entries, and, for example, the Last Entry is the only one in 1814. Gauss wrote down discoveries made. The first entry on 30 March 1796 is his famous result that a regular 17-gon can be constructed by ruler and compass.

André Weil remarks that “per inductionem” could be understood as, and translated by “emperically”. The word “induction” in our context does not stand for “mathematical induction”.

(2) We phrase the prediction by Gauss in other terms. We write $\mathbb{F}_p = \mathbb{Z}/p$ for (the set, the ring) the field of integers modulo a prime number $p$. Suppose $p \equiv 1 \pmod{4}$. Once $p$ is fixed we write

$$N = \# \left\{ (x, y) \in (\mathbb{F}_p)^2 \mid 1 = x^2 + y^2 + x^2y^2 \right\} + 4.$$ 

A prime number $p$ with $p \equiv 1 \pmod{4}$ can be written as a sum of two squares of integers (as Fermat proved). These integers are unique up to sign and up to permutation. Suppose we write

$$p = a^2 + b^2, \quad \text{with} \quad b \text{ even and } a - 1 \equiv b \pmod{4}.$$ 

In this case Gauss predicted

$$N = (a - 1)^2 + b^2$$ 

for every $p \equiv 1 \pmod{4}$. 
(3) **Some history.** Herglotz gave in 1921 the first proof for this expectation by Gauss.

We will see the attempt and precise formulation of Gauss of this problem as the pre-history and a prelude of the *Riemann hypothesis in positive characteristic* as developed by E. Artin, F. K. Schmidt, Hasse, Deuring, Weil and many others.
(4) Some examples. For $p = 5$ we obtain $a = -1$, and $b = \pm 2$ and $N = 8$. Indeed, $(x = \pm 1, y = \pm 1)$ are the only solutions for $Y^2 = X(X - 1)(X + 1)$ (see Proposition 2(a) for an explanation). $p = 5 = (-1)^2 + 2^2$, $N = 8$.

For $p = 13$ we have $(+3)^2 + (2)^2 = 13$ and indeed $(+3 - 1)^2 + 2^2 = 8$ is the number of $\mathbb{F}_{13}$-rational points as can be computed easily; $p = 13 = (+3)^2 + 2^2$, $N = 8$.

For $p = 17$ there are the 12 solutions: the point 0, the three 2-torsion points, and $x = 4, 5, 7, 10, 12, 13$; the points $(x = 4, y = \pm 3)$ and $(x = 13, y = \pm 5)$ we will encounter below as $(x = -e, y = 1 + e)$ with $e^2 = -1$; $p = 17 = (+1)^2 + 4^2$, $N = 16$. 


(5) This curve considered by Gauss. We see in the statement by Gauss four points “at infinity”. Here is his explanation. Consider the projective curve

\[ C = \mathcal{Z}(-Z^4 + X^2Z^2 + Y^2Z^2 + X^2Y^2) \subset \mathbb{P}_K^2 \]

over a field \( K \) of characteristic not equal to 2. For \( Z = 0 \) we have points \( P_2 = [x = 0 : y = 1 : z = 0] \) and \( P_1 = [x = 1 : y = 0 : z = 0] \). Around \( P_2 \) we can use a local chart given by \( Y = 1 \), and \( \mathcal{Z}(-Z^4 + X^2Z^2 + Z^2 + X^2) \); we see that the tangent cone is given by \( \mathcal{Z}(Z^2 + X^2) \) (the lowest degree part); hence we have a ordinary double point, rational over the base field \( K \) and the tangents to the two branches are conjugate if \(-1\) is not a square in \( K \), respectively given by \( X = \pm eZ \) with \( e^2 = -1 \) in \( L \). This is what Gauss meant by \( x = \infty, \quad y = \pm i \). Analogously for \( P_2 \) and \( y = \infty, \quad x = \pm i \).
Explanation. Any algebraic curve (an absolutely reduced, absolutely irreducible scheme of dimension one) \( C \) over a field \( K \) is birationally equivalent over \( K \) to a non-singular, projective curve \( C' \), and \( C' \) is uniquely determined by \( C \). The affine curve \( \mathcal{Z}(-1 + X^2 + Y^2 + X^2Y^2) \subset \mathbb{A}^2_K \), over a field \( K \) of characteristic not equal to 2 determines uniquely a curve, denoted by \( E_K \) in this note. This general fact will not be used: \textit{we will construct explicit equations for } \( E_K \) (over any field considered) \textit{and for } \( E_L \) over a field with an element \( e \in L \) satisfying \( e^2 = -1 \).
In the present case, we write $C \subset \mathbb{P}_K^2$ as above (the projective closure of the curve given by Gauss), $E$ for the normalization. We have a morphism $h : E \to C$ defined over $K$. On $E$ we have a set $S$ of 4 geometric points, rational over any field $L \supset K$ in which $-1$ is a square, such that the induced morphism

$$E \setminus S \longrightarrow \mathcal{Z}(-1 + X^2 + Y^2 + X^2Y^2) \subset \mathbb{A}_K^2$$

is an isomorphism.
(6) A normal form, 1.

Proposition 1. Suppose $K$ is a field of characteristic not equal to 2.

(a). The elliptic curve $E$ can be given by $T^2 = 1 - X^4$.

(b). The elliptic curve $E$ can be given by $U^2 = V^3 + 4V$.

(c). There is a subgroup $\mathbb{Z}/4 \hookrightarrow E(K)$.
(7) The case \( p \equiv 3 \pmod{4} \) (not mentioned by Gauss).

**Theorem 2.** The elliptic curve \( E \) over \( \mathbb{F}_p \) with \( p \equiv 3 \pmod{4} \) has:

\[
\#(E(\mathbb{F}_p)) = p + 1.
\]

**Proof.** The elliptic curve \( E \) can be given by the equation \( Y^2 = X^3 + 4X \). We define \( E' \) by the equation \( -Y^2 = X^3 + 4X \). We see:

\[
\#(E(K)) + \#(E'(K)) = 2p + 2; \quad E \cong_K E'.
\]

Indeed, any \( x \in \mathbb{P}^1(K) \) giving a 2-torsion point contributes +1 to both terms; any possible \((x, \pm y)\) with \( y \neq 0 \) contributes +2 to exactly one of the terms (we use the fact that \(-1\) is not a square in \( \mathbb{F}_p \) in this case). The substitution \( X \mapsto -X \) shows the second claim. Hence \( \#(E(K)) = (2p + 2)/2. \) \( \square \)
Proposition 3. Suppose \( L \) is a field of characteristic not equal to 2. Suppose there is an element \( e \in L \) with \( e^2 = -1 \).

(a). The elliptic curve \( E_L \) can be given by \( Y^2 = X(X - 1)(X + 1) \).

(b). There is a subgroup \((\mathbb{Z}/4 \times \mathbb{Z}/2) \hookrightarrow E(L)\).

Note: \( x = e \) gives \( e(e - 1)(e + 1) = (e - 1)^2 \).

We will study this in case either \( L = \mathbb{Q}(\sqrt{-1}) \) or \( L = \mathbb{F}_p \) with \( p \equiv 1 \pmod{4} \) (as Gauss did in his Last Entry).
(9) An aside: a historical line, uniformization of algebraic curves. We work over $\mathbb{C}$.
(a) The circle given by $X^2 + Y^2 = 1$ can be parametrized by $X = \sin(z)$, and $Y = \cosin(z)$ (a uniformization by complex functions). However, as any conic is a rational curve it can also be parametrized by rational functions, in this case as

$$
\left(\frac{t^2 - 1}{t^2 + 1}\right)^2 + \left(\frac{2t}{t^2 + 1}\right)^2 = 1.
$$
(b) The quartic equation given by Gauss in his Last Entry originates in the theory of the lemniscate functions. At the time of Gauss it was known that the curve given by \( x^2 + y^2 + x^2 y^2 = 1 \) can be uniformed by the lemniscate functions

\[
t \mapsto (x = cl(t), y = sl(t)).
\]

In the notation of Gauss: \( x = \text{sinlemn } u, \)
\( y = \text{coslemn } u. \) These functions are analogous of the usual \textit{sine} and \textit{cosine} functions, with the circle replaced by the lemniscate of Bernouilli. However for this curve (of genus one) there is no rational parametrization.
(c) Abel, Jacobi and Weierstrass proved that this parametrization of this particular elliptic curve can be generalized to any elliptic curve (a curve of genus one) in a parametrization by transcendental functions: write an equation for the curve in a special form, and show that
\[ x = \wp(z), \quad y = \left( \partial/\partial z \right) (\wp(z)) \]
satisfies this differential equation.

(d) Such a uniformization for elliptic curves was generalized by Koebe and Poincaré (1907) to curves of arbitrary genus at least two (mapping the Poincaré upper half plane onto the set of complex point of the curve considered, again by transcendental functions).
In the 20-th century we observe the problem that arithmetic properties are difficult to read off from transcendental functions. How do we determine rational points (say over a number field) on a given algebraic curve (e.g. as we like to do in FLT)?

(e) The great breakthrough was in the Shimura-Taniyama-Weil conjecture: elliptic curves over $\mathbb{Q}$ are rationally parametrized (not by $\mathbb{P}^1$, but) by modular curves (and these curves we know very well).

We see the genesis of mathematics, the developments from the roots to final theory (uniformizations of algebraic curves).
We come to a proof of the expectation by Gauss in his “Last Entry”:

\[
\text{Let } a + bi \text{ be a prime element in } \mathbb{Z}[i] \text{ such that } a - 1 + bi \text{ is divisible by } 2 + 2i; \text{ then }
\#(E(\mathbb{F}_p)) = (a - 1)^2 + b^2.
\]

We conclude that \(a^2 + b^2 = p\), a (rational) prime number with \(p \equiv 1 \pmod{4}\).
A version of the claim by Gauss:

\[ N = N(p) := \#\{(x, y) \in \mathbb{F}_p \mid y^2 = X(X + 1)(X - 1)\} + 1; \]

\[ p = a^2 + b^2, \quad a, b \in \mathbb{Z}; \]

we see that \( b \) is even, and \( a \) is odd and \( a - 1 \equiv b \pmod{4} \). Claim by Gauss:

\[ N = (a - 1)^2 + b^2. \]

We write \( \pi = a + bi \), and see \( N = \text{Norm}(\pi - 1) \).

Reminder, examples:

\( p = 5 \), and \( \pi = -1 + 2i; \quad N(5) = 8 \);
\( p = 13 \), and \( \pi = 3 + 2i; \quad N(13) = 8 \);
\( p = 17 \), and \( \pi = +1 + 4i; \quad N(17) = 16 \);
\( p = 29 \), and \( \pi = -5 + 2i; \quad N(29) = 32 \); etc.

We discuss a proof given in 20-th century language and notation.
(10) An aside: a historical line, from Gauss into the 20-th century. We work over a finite field.

(a) The Last Entry by Gauss, giving a (possible) interpretation of the number of rational points as the norm of an endomorphism.

(b) The Riemann Hypothesis for the zeros of the zeta function.

(c) The PhD thesis (1921, 1924) by Emil Artin, where a characteristic $p$ analog of RH is launched, with many proofs, and checking the pRH in 40 different cases.
(d) Proofs by Hasse of pRH for elliptic curves (1933, 1934, 1936).

(e) Proofs by F. K. Schmidt, Deuring and many others trying to understand the zeta function of an algebraic curve over a finite field.

(f) Proofs by Weil of the pRH for curves of arbitrary genus and for abelian varieties (1940, 1941, 1948). Algebraic geometry over an arbitrary field had to be described in great generality.
(g) Weil gives an interpretation of the pRH as an analogue of the Fixed Point Theorem of Lefschetz, formulation of the Weil conjectures.


(j) “Standard conjectures” (Grothendieck), still wide open.
A comment. Gauss considered solutions of this equation modulo $p$. Only much later $E_{\mathbb{F}_p}$ was considered as an independent mathematical object, not necessarily a set of modulo $p$ solutions of a characteristic zero polynomial. In the beginning as the set of valuations of a function field, as in the PhD-thesis by Emil Artin, 1921/1924. For elliptic curves this was an accessible concept, but for curves of higher genus (leave alone for varieties of higher dimensions) this was cumbersome. A next step was to consider instead a geometric object over a finite field; a whole new aspect of (arithmetic) algebraic geometry had to be developed before we could proceed. Each of these new insights was not easily derived; however, as a reward we now have a rich theory, and a thorough understanding of the impact of ideas as in the Last Entry of Gauss.
(11) Frobenius maps and formulas. We recall some theory developed by Emil Artin, F.K. Schmidt, Hasse, Deuring, Weil and many others, now well-known, and later incorporated in the general theory concerning “the Riemann Hypothesis in positive characteristic”. For simplicity we only discuss the case of the ground field $\mathbb{F}_p$. 
Easy, but crucial observation.
Let $\mathbb{F}_p \subset k$. The map $x \mapsto x^p$ is a field homomorphism $k \to k$. Moreover the set of fixed points is:

$$\mathbb{F}_p = \{x \in k \mid x^p = x\}.$$

This simple-minded fact, via a geometric interpretation of higher dimensional analogues, is the basis of our proof (and of the proof of the Riemann Hypothesis in positive characteristic for curves and for abelian varieties).
The Frobenius morphism. For a variety $V$ over the field $\mathbb{F}_p$ we construct a morphism

$$\text{Frob}_V = F : V \to V,$$

defined by “raising all coordinates to the power $p$”. Note that if $(x)$ is a zero of $f = \sum a_\alpha X^\alpha$, then indeed $(x^p)$ is a zero of $\sum a_\alpha^p X^\alpha$, because $a_\alpha^p = a_\alpha$,

$$f(x)^p = (\sum a_\alpha x^{\alpha})^p = \sum a_\alpha^p (x^{\alpha})^p = f(x^p).$$

This morphism was considered by Hasse in 1930.

(Side remark. An analogous morphism can be defined for any variety over $\mathbb{F}_q$ for any $q = p^n$.)
Easy fact: the set invariants of $\text{Frob}_V : V \to V$ is exactly $V(\mathbb{F}_p)$, the set of $\mathbb{F}_p$-rational points on $V$.

A little warning. The morphism $\pi : V \to V$ induces a bijection $\pi(k) : V(k) \to V(k)$ for every algebraically closed field $k \supset \mathbb{F}_p$; however (in case the dimension of $V$ is at least one) $\pi : V \to V$ is not an isomorphism.

We illustrate this in one example. Let $V = \mathbb{A}^1_{\mathbb{F}_p}$, the affine line. The morphism $\text{Frob}_V : V \to V$ on points is given by $x \mapsto x^p$; on coordinate rings it is given by $K[T] \leftarrow K[S]$ with $T^p \leftarrow S$. We see this is not an isomorphism. However note that $V(x) = x$ if and only if $x \in \mathbb{A}^1_{\mathbb{F}_p}(\mathbb{F}_p) = \mathbb{F}_p$. Moreover note that for any perfect field $\Omega \supset \mathbb{F}_p$ (e.g. a finite field, or an algebraically closed field) we obtain a bijection

$$\text{Frob}(\Omega) : \Omega = V(\Omega) \to \Omega = V(\Omega),$$

injective because $y^p = z^p$ implies $(y - z)^p = 0$ hence $y = z$; surjective because $\sqrt[p]{z} \mapsto z$. 

We are going to apply this to our elliptic curve \( E \). We write \( \pi = \text{Frob}_E \). Using addition on \( E \) we can write, and see

\[
\text{Ker}(\pi - 1 : E \to E) = E(\mathbb{F}_p).
\]

It turns to that we can consider \( \pi \in \text{End}(E) \) as a complex number. (For specialists: \( \text{End}(E) \) is a characteristic zero domain.) A small argument shows that

\[
\text{Norm}(\pi - 1) = \#(E(\mathbb{F}_p)) =: N.
\]

(For specialists: \( \pi - 1 \) is separable, hence all fixed points have multiplicity one.)
Moreover for the complex conjugate $\pi$ we have $\pi \cdot \bar{\pi} = p$. Write $\beta := \pi + \bar{\pi}$, the trace of $\pi$. We see that $\pi$ is a zero of

$$T^2 - \beta \cdot T + p; \quad |\pi| = \sqrt{p};$$

$$N = \text{Norm}(\pi - 1) = (\pi - 1)(\bar{\pi} - 1) = 1 - \beta + p.$$

**Crucial observation** by André Weil. Stare at the equality

$$\#(E(\mathbb{F}_p)) = 1 - \text{Trace}(\pi) + p,$$

where $\#(E(\mathbb{F}_p))$ is the number of fixed points of $\pi : E \to E$. What is the interpretation (and possible generalization) of this?
The statements usually indicated by “the Riemann hypothesis in positive characteristic” I tend to indicate by pRH, in order to distinguish this from the classical Riemann hypothesis RH. For any elliptic curve $C$ over a finite field $\mathbb{F}_q$ one can define its zeta function (as can be done for more general curves, and more general varieties over a finite field). As E. Artin and F. K. Schmidt showed, for an elliptic curve over $\mathbb{F}_q$ we have

$$Z(E, T) = \frac{(1 - \rho T)(1 - \bar{\rho} T)}{(1 - T)(1 - qT)}.$$

As is usual, the (complex) variable $s$ is defined by $T = q^{-s}$. The theorem proved by Hasse is

$$|\rho| = \sqrt{q} = |\bar{\rho}|; \quad \text{this translates into} \quad s = \frac{1}{2} \quad (\text{pRH}),$$

and we see the analogy with the classical RH, which explains the terminology pRH.
A comment. Gauss considered solutions of this equation modulo $p$. Only much later $E_{\mathbb{F}_p}$ was considered as an independent mathematical object, not necessarily a set of modulo $p$ solutions of a characteristic zero polynomial. In the beginning as the set of valuations of a function field, as in the PhD-thesis by Emil Artin, 1921/1924. For elliptic curves this was an accessible concept, but for curves of higher genus (leave alone for varieties of higher dimensions) this was cumbersome. A next step was to consider instead a geometric object over a finite field; a whole now aspect of (arithmetic) algebraic geometry had to be developed before we could proceed. Each of these new insights was not easily derived; however, as a reward we now have a rich theory, and a thorough understanding of the impact of ideas as in the Last Entry of Gauss.
(12) A proof for the statement by Gauss in his Last Entry. We have $p = a^2 + b^2$, where $2 + 2i$ divides $a - 1 + bi$. Hence 8 divides $(a - 1)^2 + b^2$. We conclude:

$p \equiv 1 \pmod{4}$, with $a$ odd and $b$ even; the sign of $a$ is determined by

$$a - 1 \equiv b \pmod{4},$$

explicitly,

$p \equiv 1 \pmod{8}$ then $a \equiv 1 \pmod{4}$

$p \equiv 5 \pmod{8}$ then $a \equiv 3 \pmod{4}$;

proof: 8 divides

$$(a - 1)^2 + b^2 = a^2 - 2a + 1 + b^2 = p + 1 - 2a;$$

either $p \equiv 1 \pmod{8}$ and $2a \equiv 2 \pmod{8}$
or $p \equiv 5 \pmod{8}$ and $2a \equiv 6 \pmod{8}$. 
**Expectation / Theorem** (Gauss 1841, Herglotz 1921).

\[ \#(E(\mathbb{F}_p)) = \text{Norm}(\pi - 1) = (a - 1)^2 + b^2. \]

**Proof** (assuming our knowledge for \( \pi \in \text{End}(E) \)).

We know \( \pi \cdot \pi = p \) and

\[ \text{Norm}(\pi - 1) = \#(E(\mathbb{F}_p)) =: N \]

and

\[ N = 1 - \text{Trace}(\pi) + p. \]

Hence

\[ N = 1 - 2a + a^2 + b^2 = (a - 1)^2 + b^2. \]
Thank you for this opportunity to talk to you,

and thank you for your attention.