

ANNALES DE LA FACULTÉ DES SCIENCES DE TOULOUSE Mathématiques

GERHARD FREY

The Way to the Proof of Fermat's Last Theorem

Tome XVIII, n° S2 (2009), p. 5-23.

http://afst.cedram.org/item?id=AFST_2009_6_18_S2_5_0

© Université Paul Sabatier, Toulouse, 2009, tous droits réservés.

L'accès aux articles de la revue « Annales de la faculté des sciences de Toulouse Mathématiques » (<http://afst.cedram.org/>), implique l'accord avec les conditions générales d'utilisation (<http://afst.cedram.org/legal/>). Toute reproduction en tout ou partie cet article sous quelque forme que ce soit pour tout usage autre que l'utilisation à fin strictement personnelle du copiste est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

cedram

Article mis en ligne dans le cadre du
Centre de diffusion des revues académiques de mathématiques
<http://www.cedram.org/>

The Way to the Proof of Fermat's Last Theorem

GERHARD FREY^(*)

1. Fermat's Claim and Wiles' Theorem

More than 360 years ago one of the most exciting stories in the history of Mathematics began when *Pierre de Fermat* stated on the margin of a copy of Diophant's work the

CONJECTURE 1. — *Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos, et generaliter nullam in infinitum ultra quadratum potestatem in duas ejusdem nominis fas est dividere...*

or in modern language

There are no natural numbers $n \geq 3, x, y, z$ such that

$$x^n + y^n = z^n \quad \text{(FLT)}.$$

In fact Fermat did not state a conjecture but he **claimed** to have a proof:

... cujus rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.

(*) This article is based on a lecture delivered at the conference «*Fermat, Quatre cents ans après*» at the University of Toulouse in 2001. Though it was written immediately after the conference and hence a long time ago the author decided not to change it thoroughly during proof reading. He only wants to remark that very important new results were obtained in the areas discussed in the paper. So Serre's conjecture was proved in full generality by work of Khare, Kisin, Wintenberger, et al. The author would like to thank the organizers of the conference very much for the warm hospitality and the interesting lectures he could enjoy in Toulouse, and for their persisting efforts which finally led to the publication of the lectures.

The content of Fermat's statement is easy to understand and the methods to prove it seem to lie at hand: Use the symmetry of the equation and then try to find conditions for solutions by factorization of both sides of the equation. For small exponent n this elementary approach is successful. Fermat himself gave a proof for $n = 4$ by using the beautiful method of **descente infinie**. But already for $n = 3$ Euler had to go to new grounds of number theory and compute with third roots of unity (and he made some minor mistakes at the first attempt), and though the exponents 5 and 7 still can be treated "by elementary number theory" (Lamé) it becomes clear that one has to replace the field of rational by algebraic number fields, especially cyclotomic fields obtained by adjoining n -th roots of unity ζ_n to \mathbb{Q} .

It is easily seen that it will be enough to prove **FLT** for exponents which are odd prime numbers and so we shall replace the exponent n by the exponent p with p a prime different from 2 from now on.

Here **E. E. Kummer** did his ground breaking and celebrated work. One high point is his theorem that Fermat's claim is true if the exponent is a regular prime p . It may be worth while to state this result in the frame of Galois theory:

THEOREM 1.1 (Kummer). — *If $\mathbb{Q}(\zeta_p)$ has no unramified cyclic extension of degree p then Fermat's claim is true for the exponent p .*

But even if p does not have this property one can find conditions for the truth of FLT by using **Kummer congruences**. Philosophically they should establish a local-global principle (which is not true for the Fermat equation itself):

By studying local conditions at the archimedean and non archimedean places of $\mathbb{Q}(\zeta_p)$ imposed on assumed solutions one hopes to find a contradiction.

These congruence conditions have been refined in a very remarkable way during the development of algebraic number theory (for details cf. [Ri]) and it cannot be excluded that eventually they will establish another proof of FLT but the present situation led H.M.Edwards 1978 in a survey article to the conclusion that though Fermat's claim is one of the most famous mathematical problems it is not subject to recent mathematical research since no one knows how to attack it.

In fact the methods for the solution of the problems were available at that time already but they do not come from "classical" number theory. The aim of this article is to explain how **Galois representations** yield local-global principles which finally are strong enough to prove Fermat's

claim.

In fact much more is proved!

The following result was announced by **Andrew Wiles** 1993 and proved 1994 ([W]) relying on a joint work with Richard Taylor ([T-W]):

THEOREM 1.2. — *Semi stable elliptic curves over \mathbb{Q} are modular.*¹

In the following we try to explain the meaning of Wiles' theorem and why it settles Fermat's claim and we would like to convince the reader that Fermat's Last Theorem is not true because of an accident but because of a reason derived from general principles concerning the Galois group of the rational numbers and its geometric and automorphic representations.

2. Arithmetic of torsion points of elliptic curves

In this section we shall explain how Fermat's claim is related to torsion points of very special elliptic curves.

2.1. Definitions

An **elliptic curve** E over a field K is (the K -isomorphism class of) a plane irreducible projective cubic curve without singularities with a K -rational point.

If we fix such a rational point as "point at infinity" we can describe E by an affine plane cubic. If $\text{char}(K) \neq 2$ we find for E an equation

$$Y^2 = X^3 + AX^2 + BX + C =: f_3(X)$$

with $A, B, C \in K$.

The non singularity of E is equivalent to the fact that the discriminant Δ_E of $f_3(X)$ is not equal to 0.

Δ_E is determined by E up to 12-th powers.

The absolute invariant j_E is uniquely determined by E and determines E over \bar{K} .

We give its definition only if $\text{char}(K) \neq 2, 3$. In this case we can transform E to an isomorphic curve E' given by

$$E' : Y^2 = X^3 - g_2X - g_3.$$

⁽¹⁾ In fact Wiles used semi stability only at the primes 3 and 5; the general result that all elliptic curves are modular was proved by Breuil, Conrad, Diamond and Taylor in 1999.

Then

$$j_E = 12^3 \cdot \frac{4 \cdot g_2^3}{\Delta_{E'}}$$

with $\Delta_{E'} = 4g_2^3 - 27g_3^2$.

2.2. Reduction of Elliptic Curves

Take $K = \mathbb{Q}$. It is easy to see that we can choose $A, B, C \in \mathbb{Z}$ **and** such that Δ_E has minimal absolute value.

From now on we shall always assume that this is the case.

Let p be a prime. For simplicity we shall assume in the following definitions that p is odd. But analogues definitions can be done for $p = 2$, too (cf. [Si]).

The reduction of E at p is the cubic $E^{(p)}$ over $\mathbb{Z}/p =: \mathbb{F}_p$ one gets by taking the residues of A, B, C modulo p .

- E has good reduction at p if $f_3(X) \bmod p$ has three different zeroes in $\overline{\mathbb{F}_p}$.
Otherwise E has bad reduction.
- E has semi stable reduction at p if $f_3(X) \bmod p$ has at least 2 different zeroes in $\overline{\mathbb{F}_p}$.

The conductor N_E is a number which is divisible exactly by the primes at which E has bad reduction, hence $N_E \mid \Delta_E$.

For all primes p including $p = 2$ we have:
 E is semi stable at p iff $v_p(N_E) \leq 1$.

For $p \neq 2, 3$ the exponent $v_p(N_E)$ is bounded by 2, for $p = 3$ it is bounded by 5 and for $p = 2$ by 8. The exact definition and an algorithm for the computation of N_E are given in [Ta].

2.3. Torsion points

The set of algebraic points of E is

$$E(\overline{\mathbb{Q}}) := \{(x, y) \in \overline{\mathbb{Q}} \times \overline{\mathbb{Q}}; y^2 = f_3(x)\} \cup \{\infty\}.$$

This set is an abelian group in which the addition law is given by rational functions with coefficients in \mathbb{Q} and so it is compatible with the action of $G_{\mathbb{Q}}$ [Si].

For $n \in \mathbb{N}$ define

$$E_n := \{P \in E(\bar{\mathbb{Q}}); n \cdot P = \infty\}.$$

This is a $G_{\mathbb{Q}}$ -module which is isomorphic to $\mathbb{Z}/n \times \mathbb{Z}/n$ as abelian group.

The coordinates of points in E_n are zeroes of polynomials with coefficients in \mathbb{Q} and hence are algebraic numbers.

Let K_n be the field obtained by adjoining the coordinates of all points of order n of E . It is a Galois extension with Galois group embeddable into $Gl(2, \mathbb{Z}/n)$.

There is a non degenerate $G_{\mathbb{Q}}$ -compatible symplectic form on E_n with values in the group μ_n of roots of unity of order n called Weil pairing.

Consequence: K_n contains $\mathbb{Q}(\mu_n)$, and the Galois group of K_n over $\mathbb{Q}(\mu_n)$ is contained in $Sl(2, \mathbb{Z}/n)$. In general the image is as large as possible:

If E has no complex multiplication (e.g. if j_E is not an integer) for almost all primes p the Galois group of K_p/\mathbb{Q} is equal to $Gl(2, \mathbb{Z}/p)$ ([Se1]).

So it is a highly interesting diophantine question to find curves E for which the field K_p is "small" for some p .

For instance one can look for elliptic curves for which a point of order p has coordinates in \mathbb{Q} or another given fixed number field K or, weaker, that a cyclic subgroups of E_p is invariant under the action of the Galois group of \mathbb{Q} .

This question was already studied by B. Levi in the early years of the last century. He gave a list of torsion points of elliptic curves over \mathbb{Q} and conjectured that it was complete. Much later (in the sixties of the last century) this problem was brought to the attention of the mathematicians again by the work of A. Ogg. But now the situation was better than at Levi's time. Due to the work of Néron and Kodaira the reduction theory sketched above was well understood and an explicit classification of the special fiber of the minimal model was available. Moreover the Tate curve (cf. [Ro] or [Si]) described the rational points of semi stable elliptic curves with bad reduction over p -adic fields by a p -adic analytic parametrisation compatible with the Galois action of the local fields.

So the problem of torsion points could be studied **locally** to find necessary conditions for the existence of torsion points. The hope was that they were so sharp that contradiction to global properties would result (cf. [He] and [F1]).

To give a flavor of the results obtained we cite a result which can be easily deduced from [F1]:

PROPOSITION 2.1. — *Let $p \geq 5$ be a prime.*

Then E is semi stable at all primes of K_p and for primes q not dividing 2 the q -adic value of the minimal discriminant of E over K_p is divisible by p .

These arithmetical properties of E can be expressed nicely by divisorial properties of the X -coordinates of points of order 2:

COROLLARY 2.2. — *We assume that the points of order 2 of E are \mathbb{Q} -rational. We can find an equation for E which is of the form*

$$E : Y^2 = X(X - x_1)(X - x_2)$$

where $\{0, x_1, x_2\} \subset \mathbb{Z}$ are the X -coordinates of points of order 2.

Then the principal divisors (x_i) are equal to $p \cdot d_i$ with d_i a divisor of K_p , and for $(x_1 - x_2)$ we get:

$(x_1 - x_2) = d_0 + p \cdot d_3$ with d_0 an effective divisor which is only divisible by divisors of 2.

We remark that up to factors related to 2 we get a "divisorial solution" of Fermat's equation with exponent p **in the field K_p** just by using the coordinates of points of order. If K_p would be equal to $\mathbb{Q}(\zeta_p)$ we would get "nearly" a solution of Fermat's equation!

2.4. The Turning Point

But this last condition will never be satisfied for in 1976 **Barry Mazur** published his celebrated paper [Ma] and amongst other results he could list **all** elliptic curves defined over \mathbb{Q} with isogenies. As a consequence Mazur got a bound for the order of cyclic isogenies rational over \mathbb{Q} and of the primes dividing the order of rational torsion points (≤ 7).

The tools Mazur used to get these results were global:

Mazur used the arithmetic of **modular curves** and their differentials (**modular forms**).

So much of the research work devoted to the study of torsion points of elliptic curves became obsolete. But now it was exciting to reverse the argumentation.

Thanks to Mazur we know now that for $p > 7$ and curves without complex multiplication the field K_p is a large Galois extension. By the local computations described above we can determine its ramification behavior.

So **assume** that

$$(a, b, c) \text{ is a solution of } X^p + Y^p = Z^p$$

for $abc \neq 0$ and $p > 7$.

Without loss of generality we can assume that a and b are odd and that $b \equiv 3 \pmod{4}$.

Then define

$$E_{a,b} : Y^2 = X(X - a^p)(X - b^p).$$

Reduction theory applied to this curve yields

THEOREM 2.3. —

1. $E_{a,b}$ is semi stable at all primes l .
2. K_p/\mathbb{Q} is unramified outside of primes dividing $2p$.
3. K_p is little ramified² at places \mathfrak{p} dividing p , i.e. the completion $K_{\mathfrak{p}}$ is obtained by tamely ramified extensions of $\mathbb{Q}_{\mathfrak{p}}$ followed by Kummer extensions of degree p where the radicals are \mathfrak{p} -units.
4. The Galois group $G_p = G(K_p/\mathbb{Q})$ is isomorphic to $Gl(2, p)$.

For details of the proof cf. [F2].

Hence a solution of Fermat's equation implies the existence of a nearly unramified extension of \mathbb{Q} with Galois group $Gl(2, p)$ containing the p -th roots of unity, and in order to prove FLT one has to prove the non-existence of such extensions.

Here the reader should remember Kummer's criterion. But this criterion was about abelian extensions over $\mathbb{Q}(\mu_p)$ and so we can (at least nowadays) use class field theory to discuss it.

In our case we have a Galois extension with a large non solvable group, and so class field theory cannot help. But we have translated the Fermat problem into the question whether a certain representations of $G_{\mathbb{Q}}$ of **dimension 2** can exist, and as it is well known from work of Serre, Deligne,

⁽²⁾ cf. [Se2]

Ribet, Shimura and many others such representations are closely related to modular forms which were so useful in Mazur's paper.

3. Galois representations of \mathbb{Q}

At the end of the last section we have seen that a non trivial solution of Fermat's equation with exponent $p > 7$ would imply that a nearly unramified Galois extension of \mathbb{Q} with Galois group isomorphic to $Gl(2, p)$ would exist. To deal with such extensions we have to study Galois representations.

3.1. Definitions and notation

A Galois representation of $G_{\mathbb{Q}}$ is a continuous homomorphisms (with respect to the Krull topology on $G_{\mathbb{Q}}$)

$$\rho : G_{\mathbb{Q}} \rightarrow GL_n(R)$$

with $GL_n(R)$ the set of invertible $n \times n$ -matrices over a ring R with a given topology.

For us $n = 2$ is the most important case. As coefficients R we mostly use finite fields, finite quotients of \mathbb{Z} or finite extensions of \mathbb{Z}_l for some prime number l . But one important ingredient in Wiles' proof uses $R = \mathbb{C}$.

3.1.1. Conductor

Let ρ be a representation of $G_{\mathbb{Q}}$ with R a finite field k of characteristic p .

Let K_{ρ} be the fixed field of the kernel of ρ and let V be a representation space of ρ .

Attached to ρ is the **Artin conductor** N'_{ρ} defined in the following way:

Let q be a prime and \mathfrak{q} a divisor of q in K_{ρ} . For $i \geq 0$ let $G_i(q)$ be the i -th ramification group of $G(K_{\rho}/\mathbb{Q})$ with respect to \mathfrak{q} . Then

$$n_q := \sum_{i \geq 0} [G_0 : G_i]^{-1} \text{codim}_k(V^{G_i})$$

and

$$N'_{\rho} = \prod_{q \text{ prime}} q^{n_q}.$$

Especially we get: A prime q divides N'_{ρ} if and only if it is ramified in K_{ρ}/\mathbb{Q} .

DEFINITION 3.1. — *The Serre conductor N_ρ is equal to the prime-to- p -part of the Artin conductor N'_ρ .*

3.1.2. Example: the cyclotomic character

We describe a most important representation of dimension 1.

Take $n \in \mathbb{N}$ and $\zeta_n = e^{2\pi i/n}$. For all $\sigma \in G_{\mathbb{Q}}$ we get:

$$\sigma(\zeta_n) = \zeta_n^{k_\sigma}$$

where k_σ is a number prime to n and uniquely determined modulo n .

$$\chi_n : G_{\mathbb{Q}} \rightarrow (\mathbb{Z}/n)^*$$

with

$$\sigma \mapsto \chi_n(\sigma) = k_\sigma \bmod n$$

is a one-dimensional representation with representation space \mathbb{Z}/n . It is called the **cyclotomic character**.

3.1.3. Semisimple representations

Let ρ be as above.

Let σ be an element of $G_{\mathbb{Q}}$.

By

$$\chi_{\rho(\sigma)}(T)$$

we denote the characteristic polynomial of $\rho(\sigma)$.

For example take $n = 2$. Then

$$\chi_{\rho(\sigma)}(T) = T^2 - \text{Tr}(\rho(\sigma))T + \det(\rho(\sigma)).$$

DEFINITION 3.2. — *ρ is semisimple if ρ is determined (up to equivalence) by $\{\chi_{\rho(\sigma)}(T); \sigma \in G_{\mathbb{Q}}\}$.*

3.1.4. The Frobenius automorphisms

The key ingredient to relate arithmetic with group theory are the *Frobenius automorphisms*:

Let l be a prime. We recall the simple polynomial identity

$$(X + Y)^l \equiv X^l + Y^l \pmod{l}.$$

By evaluating this with x and y in $\bar{\mathbb{Z}}$, the ring of algebraic integers, we get that exponentiation with l is compatible with addition (and of course with multiplication) in $\bar{\mathbb{Z}}$ modulo every prime ideal \mathfrak{l} containing l . (The error term is in $\bar{\mathbb{Z}}$ divisible by l .)

DEFINITION 3.3. — *Let l be a prime number. $\sigma \in G_{\mathbb{Q}}$ is a Frobenius automorphism to l if there is a prime ideal \mathfrak{l} of $\bar{\mathbb{Z}}$ containing l such that for all $x \in \bar{\mathbb{Z}}$ holds: $\sigma(x) - x^l \in \mathfrak{l}$.*

For fixed l there are (infinitely many) different Frobenius automorphisms but they are closely related: they are conjugates in $G_{\mathbb{Q}}$. So we choose one Frobenius automorphism σ_l for each prime l but we have to make sure that our assertions and definitions do not depend on this choice.

3.1.5. Chebotarev's density theorem

If we look at **all** primes l and at corresponding Frobenius automorphisms we have a very important and powerful **Local-Global**-principle.

Chebotarev's Density Theorem. — *Let ρ be a semisimple Galois representation of $G_{\mathbb{Q}}$. Let S be a finite set of prime numbers containing the primes which ramify in K_{ρ} . Then ρ is determined by*

$$\{\chi_{\rho(\sigma_l)}(T); l \text{ runs over primes of } \mathbb{Z} \setminus S\}.$$

3.2. Galois representations attached to elliptic curves

Let E be an elliptic curve defined over \mathbb{Q} .

Recall that for $n \in \mathbb{N}$ the group $E_n(\bar{\mathbb{Q}})$ is isomorphic to $\mathbb{Z}/n \times \mathbb{Z}/n$.

Choose a base (P_1, P_2) of E_n .

Take $\sigma \in G_{\mathbb{Q}}$ and write

$$\begin{aligned}\sigma(P_1) &= a_{\sigma}P_1 \oplus c_{\sigma}P_2; \\ \sigma(P_2) &= b_{\sigma}P_1 \oplus d_{\sigma}P_2.\end{aligned}$$

The map

$$\sigma \mapsto \begin{pmatrix} a_{\sigma} & b_{\sigma} \\ c_{\sigma} & d_{\sigma} \end{pmatrix}$$

defines a two dimensional representation

$$\rho_{E,n} : G_{\mathbb{Q}} \rightarrow Gl(2, \mathbb{Z}/n).$$

A generalization:

Take a prime p and define the p -adic Tate module $T_p(E)$ as projective limit of E_{p^k} .

It is not difficult to see that $G_{\mathbb{Q}}$ acts continuously (w.r.t. the p -adic topology) on $T_p(E)$.

The corresponding Galois representation is denoted by

$$\tilde{\rho}_{E,p}.$$

Since it is equal to $\rho_{E,p}$ modulo $p \cdot T_p(E)$ it is called a **lifting** of $\rho_{E,p}$.

Remark. — $\rho_{E,p}$ has many liftings to p -adic representations. They are described by **deformation spaces** which will play a most important role later on.

We have the following deep results:

THEOREM 3.4. —

1. $\tilde{\rho}_{E,p}$ is semisimple (**Faltings**).
2. $\rho_{E,p}$ is semisimple for almost all primes p (**Serre**).

So we can apply Chebotarev's density theorem and it becomes important to compute the characteristic polynomials of the images of Frobenius automorphisms σ_l corresponding to prime numbers l .

This leads us to ...

4. L-series

4.1. Local L-series

We take a prime l different from p which does not divide the conductor of E and hence not the conductor of $\tilde{\rho}_{E,p}$.

σ_l can be interpreted as generator of the Galois group of the field \mathbb{F}_l in a natural way, and since E has good reduction $E^{(l)}$ modulo l it acts (by Hensel's Lemma) on the points of order prime to l of $E^{(l)}$ in "the same" way as on the corresponding lifted points.

Moreover the fixed points of $\sigma_l - id_{E^{(l)}}$ are exactly the \mathbb{F}_l -rational points of $E^{(l)}$.

Using this and the properties of the Weil pairing we mentioned above one shows the following

FACTS:

Let n be a natural number prime to l .

Let a_l be the number of points of $E \bmod l$. Then

$$\text{Tr}(\rho_{E,n}(\sigma_l)) \equiv l + 1 - a_l \bmod n$$

$$\text{and } \det(\rho_{E,n}(\sigma_l)) \equiv \chi_n(\sigma_l) \equiv l \bmod n.$$

PROPOSITION 4.1. — *The polynomials*

$$\{\chi_l(T) = T^2 + (a_l - l - 1)T + l; \ l \text{ prime numbers not dividing } N_E\}$$

determine almost all representations $\rho_{E,p}$ and all $\tilde{\rho}_{E,p}$.

DEFINITION 4.2. — $\chi_l(T)$ is the **local** L -series of E at l .

To define the local L -function for primes dividing the conductor of E one uses an explicit recipe depending on the special fiber of the minimal model of E ([Si]).

4.2. Global L-series

We assemble the local informations by forming the infinite product

$$L_E(s) := f^*(s) \cdot \prod_{l \text{ prime to } N_E} (1 - (l + 1 - a_l)l^{-s} + l^{1-s})^{-1}$$

with a rational function $f^*(s)$ coming from the local factors belonging to the divisors of N_E .

This product has to be seen as an analogue of the Riemann Zeta-function and it is called the **L-series of E** .

This series converges for complex numbers s with real part $> 3/2$ and can be written in this half plane as Dirichlet series

$$L_E(s) = \sum_{n=1}^{\infty} b_n n^{-s}.$$

Hence $L_E(s)$ is an analytic function in a complex half plane.

CONJECTURE 2 (*Hasse-Weil*). — $L_E(s)$ has an analytic continuation to \mathbb{C} satisfying the following functional equation:

Put

$$\Lambda_E(s) := N_E^{s/2} (2\pi)^{-2s} \Gamma(s) L_E(s).$$

Then

$$\Lambda_E(s) = W(E) \Lambda_E(2 - s)$$

with $W(E) \in \{1, -1\}$.

This means: the local data used to define L_E are tied together in such a way that a very special analytic function is created.

5. Modular Elliptic Curves

5.1. Modular curves and cusp forms

Finally we can explain the **conjecture of Taniyama** stated 1955. In the last section we formulated the Hasse-Weil conjecture and said rather vaguely that the L-series of E is expected to be a very special function. To make this more precise we need the following notions:

Define

$$\mathbb{H} := \{z \in \mathbb{C} \mid \text{Im}(z) > 0\}$$

and

$$\mathbb{H}^* := \mathbb{H} \cup \mathbb{Q} \cup \{\infty\}.$$

The group

$$\Gamma_0(N) = \left\{ \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}_2(\mathbb{Z}) \mid c \equiv 0 \pmod{N} \right\}$$

operates on \mathbb{H}^* by sending z to $(az + b)/(cz + d)$. Define

$$\Gamma_0(N) \backslash \mathbb{H}^* =: X_0(N).$$

$X_0(N)$ is a compact Riemann surface and so a projective algebraic curve defined over \mathbb{C} .

A modular form of level N , weight k and nebentypus χ (which is a Dirichlet character) is a function $f(z)$ on \mathbb{H}^* such that:

1. $f(z)$ is holomorphic in \mathbb{H} ,

2. $f(\alpha)(z) = \chi(d)(cz + d)^k f(z) \quad \forall z \in \mathbb{H}, \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$,
3. $f(z)$ is holomorphic at the cusps.

If in addition f vanishes at the cusps, then f is called a cusp form. Because of the transformation rules a cusp form $f(z)$ has a Fourier expansion

$$f(z) = \sum_{n=1}^{\infty} a_n q^n \quad \text{with } q := e^{2\pi iz}, a_n \in \mathbb{C}$$

called **q -expansion** which determines f .

The space of cusps forms of weight k and trivial nebentypus is denoted by $S_k(N)$.

It is not difficult to see that the map

$$f(z) \mapsto 2\pi i f(z) dz$$

is an isomorphism between $S_2(N)$ and the space of holomorphic differentials $\Omega^1(X_0(N))$ and hence $S_2(N)$ has dimension equal to the genus of $X_0(N)$.

By using the Hurwitz genus formula one can compute the genus of $X_0(N)$ as a function of N .

For instance we get:
The genus of $X_0(2) = 0$.

This has the

Consequence:

There is no non trivial cusp form of weight 2 and level 2.

This observation will become important later on.

Till now we have discussed modular curves and modular forms over the complex numbers. But since the points on $X_0(N)$ have a modular interpretation (points which are not cusps correspond to pairs (E, η) where E is an elliptic curve and η an isogeny of E with cyclic kernel of order N) it follows that $X_0(N)$ can be defined over \mathbb{Z} .

Hence it makes sense to speak about cusp forms defined over commutative unitary rings R . In fact these forms have a q -expansion with coefficients in R and are determined by this expansion.

The cusp forms of weight 2 and level N over R are denoted by $S_2(N)(R)$.

5.2. Modular elliptic curves

Taniyama stated the following conjecture:

CONJECTURE 3. — *Assume that the Hasse-Weil conjecture is true for the L -series*

$$L_E(s) = \sum_{n=1}^{\infty} b_n n^{-s}.$$

Then

$$f_E(z) := \sum_{n=1}^{\infty} b_n e^{2\pi i n z}$$

is a cusp form.

This conjecture has been made more precise by A.Weil and G. Shimura:

THEOREM 5.1. — *Taniyama's and the Hasse-Weil conjecture is equivalent with the existence of a non trivial map*

$$\phi : X_0(N_E) \rightarrow E$$

defined over \mathbb{Q} .

We call an elliptic curve E over \mathbb{Q} **modular** if a map ϕ like in the theorem exists. With this notation we can reformulate Taniyama's conjecture:

CONJECTURE 4 (**Taniyama-Shimura-Weil**). — *Every elliptic curve defined over \mathbb{Q} is modular.*

We see that Theorem 1.1 of A.Wiles proves part of this conjecture.

5.3. Modular Representations

In this section we shall give a Galois theoretic criterion for the modularity of an elliptic curve.

For this we need the notion of a modular representation in our special situation.

DEFINITION 5.2. — *A representation*

$$\rho : G_{\mathbb{Q}} \rightarrow Gl(2, \mathbb{Z}/p^k)$$

is modular of level N and weight 2 if there is a ring of integers O in a number field K and a cusp form $f \in S_2(N)(O)$ given by

$$f(z) = \sum_{n=1}^{\infty} b_n e^{2\pi i n z} \quad \text{with } b_n \in O, b_1 = 1$$

such that for all prime numbers l outside of a finite set we have:

$$\text{Tr}(\rho(\sigma_l)) \equiv b_l \pmod{\mathfrak{p}^k}$$

where \mathfrak{p} is a split prime ideal of O containing p and σ_l is a Frobenius automorphism to l .

Example. — Let E be a modular elliptic curve. Then ρ_{E,p^k} is modular of level N_E and weight 2 for all primes p and all natural numbers k . As modular form we can take $f_E(z)$.

Now we can characterize modular elliptic curves:

THEOREM 5.3. — *Let E/\mathbb{Q} be an elliptic curve with $L_E(s) = \sum_{n=1}^{\infty} b_n n^s$. The following properties are equivalent:*

1. E is modular.
2. The Hasse-Weil conjecture holds for $L_E(s)$.
3. $d \sum_{n=1}^{\infty} b_n q^n \in S_2(N_E)(\mathbb{Z})$
4. There is a number N such that for all primes l and all $k \in \mathbb{N}$ the representations ρ_{E,l^k} are modular of level N and weight 2.
5. There is a number N such that for one prime l and all $k \in \mathbb{N}$ the representations ρ_{E,l^k} are modular of level N and weight 2.

We have explained already that 1),2) and 3) are equivalent. To get the connection to representation theory one uses the Eichler-Shimura-congruence which relates Hecke operators to Frobenius automorphisms to construct 2-dimensional l -adic representations attached to modular forms (cf. [Rib]).

6. A conditional proof of Fermat's claim

6.1. Lowering the level

We recall that a solution (a, b, c) of Fermat's equation with exponent p gives rise to a two dimensional Galois representation ρ_p attached to the

p -torsion points of an elliptic curve $E_{a,b}$ with the additional property that the conductor is very small.

Now assume that E were modular. Then ρ_p would be modular of level N_E . But since its conductor is much smaller than N_E we can hope to find a better modular description.

And indeed this is so. The key ingredient is the phenomenon that cusp forms to different levels can become congruent modulo special primes q called **congruence primes** and so they induce the same representations modulo q .

Hence for a given ρ one can look for forms of minimal level related to ρ . A recipe for this minimal level was given by J.P. Serre in principle already in the seventies and precisely formulated 1986 (cf. [Se2]).

In the relevant example this recipe was proved by Ribet (cf.[Rib]) in the same year.

We state his result for the example we need:

THEOREM 6.1. — *Assume that E is a modular elliptic curve which is everywhere semi stable, that $p > 7$ and that the fixed field of the kernel of $\rho_{E,p}$ is little ramified at p .*

Then $\rho_{E,p}$ is modular of level $N_{\rho_{E,p}}$ and weight 2.

Now we can use our local results:

COROLLARY 6.2. — *Take*

$$E : Y^2 = X(X - A)(X - B)$$

with $A, B \in \mathbb{Z}$ relatively prime. Assume that E is modular and that p divides $AB(A - B)$ with a power divisible by p .

Then $\rho_{E,p}$ is modular of level

$$N_p = 2 \prod_{p \nmid v_l(AB(A-B))} l.$$

6.2. Application to FLT

Take $A = a^p, B = b^p$ and assume that $A - B = c^p$.

Take the corresponding elliptic curve $E_{a,b}$ and assume that $E_{a,b}$ is modular.

Then $\rho_{E,p}$ is modular of level 2. But we know already that there are no non trivial cusp forms of this level and so we get a **contradiction**.

Conclusion: *The conjecture of Taniyama-Shimura for semistable elliptic curves implies Fermat's Last Theorem.*

This was the state of the art 1986 before Wiles' work.

7. The Theorem of Wiles

In the last sections we have explained why the theorem 1.2 proves that Fermat' claim is true.

It is not possible even to sketch the proof of this result. So we have to restrict ourselves to a short description of the **strategy of the proof**.

Wiles proves criterion 5 of theorem 5.3 for $l = 3$. His starting point is $\rho_{E,3}$ so all his input information is the action of $G_{\mathbb{Q}}$ on 8 points!

The reason for this choice is that Langlands and Tunnell proved Artin's conjecture for the **complex** representation attached to $\rho_{E,3}$. By Deligne-Serre this implies that $\rho_{E,3} =: \rho$ is modular and so the first step is done.

Next he has to study the deformation problem for Galois representations which can be described by a "tangent space": This space is computable if appropriate local conditions \mathcal{D} (i.e. conditions for the restrictions of ρ' to the Galois group of l -adic fields) are imposed such that there exists a universal deformation represented by a ring $R_{\mathcal{D}}$. Hence the deformations of ρ of type \mathcal{D} correspond one-one to homomorphisms of $R_{\mathcal{D}}$.

Modular deformations are described by a ring $H_{\mathcal{D}}$ (related with the algebra of Hecke operators on modular forms) and the first step implies that there is a homomorphism

$$\eta_{\mathcal{D}} : R_{\mathcal{D}} \rightarrow H_{\mathcal{D}}$$

which is surjective because of Chebotarev's density theorem.

The main step is to show that $\eta_{\mathcal{D}}$ is an isomorphism if \mathcal{D} is chosen carefully.

By using algebraic number theory Wiles can describe the algebraic properties of $R_{\mathcal{D}}$ and he can control how this ring changes if one replaces the type \mathcal{D} by another (less or more restrictive) type. By using variants of Ribet's

theorem (used in the last section already) a similar computation can be done for $H_{\mathcal{D}}$. This establishes the first important fact (based on the “numerical criterion” of Wiles) of the proof: It is sufficient to show the injectivity of $\eta_{\mathcal{D}}$ for minimal types where minimality is determined by ρ (and not by $\rho_{E,3^k}$). For the proof in the minimal case Wiles uses another criterion which has a more geometrical flavour. It is written up in the paper of Taylor and Wiles and has been simplified by Faltings, Schoof, Diamond and others. To apply it Wiles adds carefully chosen auxiliary primes to \mathcal{D} which make the structure of $R_{\mathcal{D}}$ and $H_{\mathcal{D}}$ so “easy” (Gorenstein property, complete intersection) such that commutative algebra finally gives the result.

Bibliography

- [MF] Modular forms and Fermat's Last Theorem; ed. G. Cornell, J.H. Silverman, G. Stevens, New York (1997).
- [F1] FREY (G.). — Some remarks concerning points of finite order on elliptic curves over global fields; *Arkiv för Mat.* 15, p. 1-19 (1977).
- [F2] FREY (G.). — Links between stable elliptic curves and certain Diophantine equations; *Ann. Univ. Saraviensis*, 1, p. 1-40 (1986).
- [F3] FREY (G.). — On ternary equations of Fermat type and relations with elliptic curves; in [MF], 527-548.
- [He] HELLEGOUARCH (Y.). — Points d'ordre 2^{p^h} sur les courbes elliptiques; *Acta Arith.* 26, p. 253-263 (1975).
- [Ma] MAZUR (B.). — Modular curves and the Eisenstein ideal; *Publ. math. IHES* 47, p. 33-186 (1977).
- [Ri] RIBENBOIM (P.). — 13 Lectures on Fermat's Last Theorem; New York (1982).
- [Rib] RIBET (K.). — On modular representations of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms; *Inv. Math.* 100, p. 431-476 (1990).
- [Ro] ROQUETTE (P.). — Analytic theory of elliptic functions over local fields; *Hamb. Math. Einzelschriften, Neue Folge-Heft 1*, Vandenhoeck und Ruprecht, Göttingen (1969).
- [Se1] SERRE (J.-P.). — Propriétés galoisiennes des points d'ordre finis des courbes elliptiques; *Inv. Math.* 15, p. 259-331 (1972).
- [Se2] SERRE (J.-P.). — Sur les représentations modulaires de degré 2 de $G(\overline{\mathbb{Q}}/\mathbb{Q})$; *Duke Math. J.* 54, p. 179-230 (1987).
- [Si] SILVERMAN (J.H.). — The Arithmetic of Elliptic Curves; GTM 106, Berlin and New York (1986).
- [Ta] TATE (J.). — The arithmetic of elliptic curves; *Inv. Math.* 23, p. 179-206 (1974).
- [T-W] TAYLOR (R.), WILES (A.). — Ring theoretic properties of certain Hecke algebras; *Annals of Math.* 141, p. 553-572 (1995).
- [W] WILES (A.). — Modular elliptic curves and Fermat's Last Theorem; *Annals of Math.* 142, p. 443-551 (1995).