Preliminary version
February 2002

II Università degli
Studi di Roma

Wiles's proof of the Taniyama-Weil conjecture for semi-stable elliptic curves over Q.

René Schoof

Dipartimento di Matematica
2ª Università di Roma "Tor Vergata"
I-00133 Roma ITALY
Email: schoof@science.uva.nl

1. The Taniyama-Weil conjecture.

In this paper we sketch the proof of the Taniyama-Weil conjecture for semi-stable elliptic curves over $\mathbf{Q}$ obtained by Andrew Wiles (completed by Richard Taylor and Andrew Wiles). We often omit certain details hoping to present the essential lines of the proof in a clear way. We refer the interested reader to the original papers [8, 9] and to [1] for the details. See also [3, 5, 7]. I thank Bas Edixhoven for his help with the preparation of this paper. In this first section we explain what the Taniyama-Weil conjecture says about elliptic curves that are defined over the rational number field $\mathbf{Q}$.

Let E be an elliptic curve over $\mathbf{Q}$. Such a curve can be embedded as a smooth cubic curve in the projective plane $\mathbf{P}^2$ by means of a so-called Weierstrass equation with coefficients $a_i \in \mathbf{Q}$:

$$Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6.$$

The point at infinity $0 = (0 : 1 : 0)$ is the neutral element of the group law on E. One usually defines

$$\begin{aligned} b_2 &= a_1^2 + 4a_2, \\ b_4 &= a_1a_3 + 2a_4, \\ b_6 &= a_3^2 + 4a_6, \\ b_8 &= a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2, \\ \Delta &= 9b_2b_4b_6 - b_2^2b_8 - 8b_4^3 - 27b_6^2. \end{aligned}$$

The fact that E is smooth is reflected by the fact that the discriminant Δ does not vanish. One can always find a "minimal" Weierstrass equation, i.e., an equation with $a_i \in \mathbf{Z}$ for which $|\Delta|$ is minimal. This minimal discriminant is called the discriminant $\Delta(E)$ of E. A prime number p is called "a prime of good reduction" if the minimal equation modulo p describes again a smooth cubic

curve. A prime is called "a prime of bad reduction" if the minimal equation modulo p describes a singular curve. The bad primes are precisely the prime divisors of $\Delta(E)$.

The curve E modulo a bad prime p possesses a unique singular point with coordinates in $\mathbf{F}_p$. One can always make a linear substitution so that the singular point becomes $(0,0)$ modulo all bad primes p. This does not change the discriminant of the equation and implies that $a_3, a_4, a_6 \equiv 0 \pmod{p}$ for all bad p. The slopes of the tangent lines to the curve at the singular point $(0,0)$ are then the zeroes of the polynomal $X^2 + a_1X - a_2$ modulo p.

We define the coefficients $a(p) \in \mathbf{Z}$ associated to E as follows:

$$\begin{aligned} p + 1 - a(p) &= \#E(\mathbf{F}_p), \qquad \text{when } p \text{ is good;} \\ 1 + a(p) &= \#\{\text{solutions in } \mathbf{F}_p \text{ of } X^2 + a_1X - a_2 = 0\}, \qquad \text{when } p \text{ is bad;} \end{aligned}$$

Then the L-series of E is defined by

$$L(E,s) = \prod_{p \text{ good}} \frac{1}{1 - a(p)p^{-s} + p \cdot p^{-2s}} \prod_{p \text{ bad}} \frac{1}{1 - a(p)p^{-s}} \qquad s \in \mathbf{C},\ \mathrm{Re}\, s > 3/2.$$

We define coefficients $a(n)$ for every $n \geq 1$ by expanding $L(E,s)$ as a Dirichlet series

$$L(E,s) = \sum_{n=1}^{\infty} \frac{a(n)}{n^s} \qquad s \in \mathbf{C},\ \mathrm{Re}\, s > 3/2.$$

The elliptic curve E is called *semi-stable* at a prime p if either p is a prime of good reduction, or if p is bad and the two tangent lines at the singular point $(0,0)$ are distinct. This means that the discriminant b_2 of the polynomial $X^2 + a_1X - a_2$ is not zero mod p and that $a(p) = \pm 1$. The curve E is called semi-stable, if it is semi-stable at all primes p.

The *conductor* N of an elliptic curve is a positive integer that captures the reduction behaviour of E. We have

$$N = \prod_p p^{\delta_p}.$$

The exponent δ_p is zero for all good primes p. For bad primes of semi-stable reduction $\delta_p = 1$. For the other bad primes $\delta_p \geq 2$. For semi-stable curves E the conductor N of E is simply the product of the bad primes.

Conjecture *(Taniyama-Weil conjecture) The Fourier series*

$$f(\tau) = \sum_{n=1}^{\infty} a(n) e^{2\pi i n \tau}, \qquad \mathrm{Im}\, \tau > 0,$$

satisfies

$$f\left(\frac{a\tau + b}{c\tau + d}\right) = (c\tau + d)^2 f(\tau) \qquad \textit{for every } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N).$$

Here the subgroup $\Gamma_0(N)$ *of* $\mathrm{SL}_2(\mathbf{Z})$ *is defined by*

$$\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) : N \textit{ divides } c\}.$$

Moreover, for every $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbf{Z})$, *the function* $\tau \mapsto (c\tau + d)^{-2} f\left(\frac{a\tau+b}{c\tau+d}\right)$ *admits a Fourier expansion in positive powers of* $e^{2\pi i \tau/N}$.

Another way to phrase this, is to say that $f(\tau) = \sum_{n=1}^{\infty} a(n) e^{2\pi n i \tau}$ is a modular form of weight 2 for the group $\Gamma_0(N)$. It is then automatically a cusp form and a normalized eigenform for the Hecke algebra acting on the vector space of cusp forms of weight 2 for $\Gamma_0(N)$.

One of the motivations for believing the Taniyama-Weil conjecture is the following. Let N be the smallest integer N such that f is a modular form of weight 2 for the group $\Gamma_0(N)$. Then the involution

$$w_N = \begin{pmatrix} 0 & 1 \\ -N & 0 \end{pmatrix} \in \mathrm{SL}_2(\mathbf{R})$$

normalizes $\Gamma_0(N)$. It acts on the space of cusp forms and commutes with the action of the Hecke algebra. Therefore $w_N(f) = \pm f$. This implies that the L-function can be extended holomorphically to all of $\mathbf{C}$ and satisfies the functional equation

$$\Lambda(E, s) = \mp N^{1-s} \Lambda(E, 2 - s).$$

Here $\Lambda(E, s) = (2\pi)^{-s}\Gamma(s)L(E, s)$ is the usual modification of the L-function.

Conversely, André Weil showed in the 1967 that, if for sufficiently many Dirichlet characters χ the "twisted" L-series $\sum_{n=1}^{\infty} \chi(n)a(n)/n^s$ admit functional equations of this type, then the Taniyama-Weil conjecture must be true. It is widely believed that such L-functions admit holomorphic extensions to $\mathbf{C}$. Indeed, the celebrated Birch-Swinnerton-Dyer conjecture predicts what the value of $L(E, s)$ in $s = 1$ should be. Since $s = 1$ is not in the domain of convergence of the series $\sum_{n=1}^{\infty} a(n)/n^s$, the Birch-Swinnerton-Dyer conjecture presupposes that the L-series can be continued analytically to $s = 1$.

For future purposes we reformulate the Taniyama-Weil conjecture somewhat.

The coefficients $a(p)$ can also be recovered as follows. Let l be a prime. For every n let

$$\rho_{E[l^n]} : G_{\mathbf{Q}} \longrightarrow \mathrm{Aut}(E[l^n])) \cong \mathrm{GL}_2(\mathbf{Z}/l^n\mathbf{Z})$$

be the representation given by the natural action of $G_{\mathbf{Q}}$ on the group $E[l^n]$ of l^n-torsion points. Taking the projective limit we obtain the l-adic representation

$$\rho_{E,l} : G_{\mathbf{Q}} \longrightarrow \mathrm{Aut}(\mathrm{Ta}_l(E)) \cong \mathrm{GL}_2(\mathbf{Z}_l)$$

on the Tate module $\mathrm{Ta}_l(E) = \varprojlim E[l^n]$, which is a module of rank 2 over the ring $\mathbf{Z}_l$ of l-adic integers. For every prime of good reduction $p \neq l$, any Frobenius element $\varphi_p \in G_{\mathbf{Q}}$, acts on the Tate module $\mathrm{Ta}_l(E)$ via a 2×2 l-adic matrix with characteristic polynomial equal to

$$X^2 - a(p)X + p.$$

For every positive integer N, the Hecke algebra $\mathbf{T}(N)$ is the $\mathbf{Z}$-algebra generated by the Hecke operators acting on the space of cusp forms of weight 2 for the modular group $\Gamma_0(N)$. More precisely, it is the $\mathbf{Z}$-algebra generated by the Hecke operators T_p for all p not dividing N. The ring $\mathbf{T}(N)$ is finite, free and reduced over $\mathbf{Z}$. Another way to say that the Fourier series $f(\tau)$ is a normalized eigenform for the Hecke algebra is to say that the map

$$\mathbf{T}(N) \longrightarrow \mathbf{Z}$$

induced by $T_p \mapsto a(p)$ is a ring homomorphism.

Let A be a $\mathbf{Z}_l$-algebra endowed with the l-adic toplogy. We say that a continuous representation

$$\rho : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(A)$$

is *modular of level* N if there is a ring homomorphism

$$g : \mathbf{T}(N) \longrightarrow A$$

such that for almost all primes p and any Frobenius element $\varphi_p \in G_{\mathbf{Q}}$, the characteristic polynomial of $\rho(\varphi_p)$ acting on $\mathrm{Ta}_l(E)$ is given by

$$X^2 - f(T_p)X + p.$$

To prove the Taniyama-Weil conjecture for semi-stable curves, it suffices to show the following:

Theorem 1.1. *Let E be a semi-stable elliptic curve over $\mathbf{Q}$ of conductor N, then for some prime l for which the l-torsion points $E[l]$ form an irreducible Galois module, the representation on the Tate module $\mathrm{Ta}_l(E)$*

$$\rho_{E,l} : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(\mathbf{Z}_l)$$

is modular of level N.

Indeed, by Cebotarev's density theorem and the irreducibility condition the representation associated to the modular form f is isomorphic to $\rho_{E,l}$. See [2]. Since $f(T_p) \in \mathbf{Z}$ for almost all p, the abelian variety A_f associated to f is an elliptic curve over $\mathbf{Q}$. By Faltings's isogeny theorem A_f is isogenous to E. In this particular case this already follows from an older result of Serre's. This implies that the coefficients $a(p)$ agree for all primes p of good reduction. See [2] for the fact that the same is true for the bad primes.

2. Lifing modular representations.

In this section we explain how the Taniyama-Weil conjecture follows from a certain "lifting" result proved by Wiles. Let l be an odd prime. By $\varepsilon : G_{\mathbf{Q}} \longrightarrow \mathbf{Z}_l^*$ we denote the *cyclotomic* character: $\sigma(\zeta) = \zeta^{\varepsilon(\sigma)}$ for all $\sigma \in G_{\mathbf{Q}}$ and for every l^n-th root of unity ζ.

Definition. Let A be a complete local Noetherian $\mathbf{Z}_l$-algebra. A continuous representation

$$\rho : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(A)$$

is called *semi-stable* if its restriction to a decomposition group G_l at l is either ordinary or flat. Here ρ is called *ordinary* if G_l acts via a subgroup of $\mathrm{GL}_2(A)$ conjugate to $\left(\begin{smallmatrix} \varepsilon & * \\ 0 & 1 \end{smallmatrix}\right)$ where $\varepsilon : G_{\mathbf{Q}} \to \mathbf{Z}_l^* \to A^*$ is the cyclotomic character. The representation $\rho : G_p \to A^*$ is called *flat* if for every finite quotient A' of A, there exists a finite flat group scheme M over $\mathbf{Z}_l$ such that $M(\overline{\mathbf{Q}}_l)$ is a free A'-module of rank 2 with the action of G_l induced by ρ.

Note that the conditions only depend on the *decomposition group* G_l, where l is the characteristic of the residue field of A. The fundamental examples of semi-stable representations are provided by the representations

$$\rho_{E[l^n]} : G_{\mathbf{Q}} \longrightarrow \mathrm{Aut}(E[l^n]) \cong \mathrm{GL}_2(\mathbf{Z}/l^n\mathbf{Z})$$

and

$$\rho_{E,l} : G_{\mathbf{Q}} \longrightarrow \mathrm{Aut}(\mathrm{Ta}_l(E)) \cong \mathrm{GL}_2(\mathbf{Z}_l)$$

associated to a semi-stable elliptic curve E over $\mathbf{Q}$. This follows from the properties of the curve E over the l-adic numbers, in particular from Tate's theory of l-adic uniformization. Wiles proves a "lifting" result. In the next sections we will sketch his proof. In this section we explain how the Taniyama-Weil conjecture follows from it.

Theorem 2.1. *(A .Wiles) Let l be an odd prime and let*

$$\overline{\rho} : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(\mathbf{F}_l)$$

be a continuous irreducible representation which satisfies:

- $\det(\overline{\rho}) = \varepsilon \pmod{l}$,
- *$\overline{\rho}$ is semi-stable,*
- *For all primes $p \neq l$, the image $\overline{\rho}(I_l)$ of the inertia group I_l is contained in a subgroup conjugate to $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$,*
- *$\overline{\rho}$ is modular.*

Then for every complete local Noetherian ring A with residue field $\mathbf{F}_l$, any continuous representation ρ for which the triangle

$$\begin{array}{ccc} G_{\mathbf{Q}} & \xrightarrow{\rho} & \mathrm{GL}_2(A) \\ & \overline{\rho}\searrow & \downarrow \\ & & \mathrm{GL}_2(\mathbf{F}_l) \end{array}$$

is commutative and satisfies

- $\det(\rho) = \varepsilon$,
- *ρ is semi-stable,*
- *There is a finite set S, such that for all primes $p \neq l$ and not in S, we have: $\rho(I_p)$ is contained in a subgroup conjugate to $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$ and if $\overline{\rho}$ is unramified at p, so is ρ.*

is modular.

To deduce the Taniyama-Weil conjecture for semi-stable curves from Theorem 2.1, Wiles proceeds as follows. Let E be a semi-stable elliptic curve. For any prime l, the representation on the l-torsion points $E[l]$ is semi-stable. One rarely knows whether this representation is modular. In two instances one knows this: if $l = 2$ this follows from the work of Hecke. In this case one observes that $\mathrm{Aut}E[2] \cong S_3 \hookrightarrow \mathrm{GL}_2(\mathbf{C})$ and one knows that the L-series associated to the resulting representation

$$\overline{\rho} : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(\mathbf{C})$$

is the L-series of a modular form of weight 1. This representation was, in fact, Wiles's starting point in 1986. However, he encountered several technical difficulties in his attempts to apply Iwasawa theory to the lifting problem. Indeed, many of the arguments in Wiles's eventual proof do not apply when $l = 2$.

The other situation where one knows that the representation on $E[l]$ is modular is the case $l = 3$. In this case it so happens that the reduction homomorphism

$$\mathrm{GL}_2(\mathbf{Z}[\sqrt{-2}]) \longrightarrow \mathrm{GL}_2(\mathbf{F}_3)$$

determined by $\sqrt{-2} \mapsto 1$, admits a section s and in this way one obtains a representation

$$\overline{\rho} : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(\mathbf{F}_3) \xrightarrow{s} \mathrm{GL}_2(\mathbf{Z}[\sqrt{-2}]) \subset \mathrm{GL}_2(\mathbf{C}).$$

Moreover —and this is essential—, in this case the group $\mathrm{GL}_2(\mathbf{F}_3)$ has order 48 and is *solvable.* This allowed R. Langlands and J. Tunnell in the 1970's to show that the representation $\overline{\rho}$ is modular. Their methods are based on analytic techniques. The proof makes use of the Selberg Trace Formula.

The result is that there exists a modular form g of weight 1 whose L-series is the Artin L-series associated to the representation $s \circ \overline{\rho}$.

Multiplying g by the ϑ-series

$$\sum_{n,m \in \mathbf{Z}} e^{2\pi i(n^2+nm+m^2)} = 1 + 6 \sum_{k=1}^{\infty} b(k) e^{2\pi i k \tau}$$

we obtain a cusp form g' of weight 2 whose coefficients are congruent modulo 3 to the coefficients of g. The form g may have a high level; that is, it may only be a modular form with respect to a relatively small modular group.

For the action of $G_{\mathbf{Q}}$ on $E[3]$ there are two possibilities: either the image of $G_{\mathbf{Q}}$ is all of $\mathrm{GL}_2(\mathbf{F}_3)$, or it is contained in a subgroup conjugate to $\left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right)$. In the first case Theorem 2.1 implies that E is modular. There exists a modular form f so that the representation $\rho_{f,3}$ associated to f is isomorphic to $\rho_{E,3}$. To ensure that f is actually a modular form with respect to the group $\Gamma_0(N)$, one invokes the "lowering the level" results of Carayol that imply that f actually has level equal to the conductor N of E. See [2] for more details. This completes the proof when $E[3]$ is an irreducible $G_{\mathbf{Q}}$-module.

In the second case we consider $E[5]$. Again the image of $G_{\mathbf{Q}}$ in $\mathrm{GL}_2(\mathbf{F}_5)$ is either all of $\mathrm{GL}_2(\mathbf{F}_5)$ or is contained in a subgroup conjugate to $\left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right)$. In the latter case, E admits a rational subgroup of order 15. This gives rise to a rational point on the modular curve $X_0(15)$. This curve has genus 1 and only four non-cuspidal rational points. They all correspond to elliptic curves over $\mathbf{Q}$ that have additive reduction at 5. Although not semi-stable, it is well known that all these curves are all modular.

Therefore we may assume that the action of $G_{\mathbf{Q}}$ on $E[5]$ is irreducible. The curve $X'(5)$ that parametrizes elliptic curves E' over $\mathbf{Q}$ for which $E'[5]$ is isomorphic to $E[5]$ as a Galois module, is a twisted form of the modular cure $X(5)$. This curve has four components, all curves of genus 0. The component that contains the rational point corresponding to $(E, E[5])$ is isomorphic to $\mathbf{P}^1$ over $\mathbf{Q}$. Therefore it contains infinitely many rational points. Consider the curve $X'(5;3)$ that parametrizes elliptic curves E' with $E'[5] \cong E[5]$ and with $E[3] \cong \mathbf{Z}/3\mathbf{Z} \times \mathbf{Z}/3\mathbf{Z}$. An application of Hilbert's irreducibility theorem to the covering $X'(5;3) \longrightarrow X(5)$ shows that there are infinitely many elliptic curves E' over $\mathbf{Q}$ that have $E'[5] \cong E[5]$ and $E[3]$ irreducible. By choosing E' 5-adically close to E, we can make sure that E' is semi-stable.

Now we apply Wiles's theorem to $E'[3]$. We conclude that $\rho_{E',3}$ is modular. Therefore so is $\rho_{E',5}$ and $\rho_{E'[5]} \cong \rho_{E[5]}$. Another application of Wiles' theorem, this time with $l = 5$, gives then that $\rho_{E,5}$ is also modular of some level. Then $\rho_{E,5}$ is actually modular of level equal to the conductor N of E, as required.

It is worth noting on how many "coincidences" the proof appears to depend: the prime $l = 2$ could not be used because of several technical reasons. For any prime $l \geq 5$, the group $GL_2(\mathbf{F}_l)$ is not solvable and results like the theorem of Langlands-Tunnell are lacking. Only $l = 3$ could work in this way. To cirumvent the problems that arise when one tries to prove directly an analogue of Wiles's theorem for *reducible* $E[3]$, Wiles used a second prime l. To make the argument above work, it was essential that the genus of $X(l)$ is zero, while the genus of the cure $X_0(3l)$ is not zero. This implies that $l \leq 5$ and $l \geq 5$ respectively. So $l = 5$ was the only choice Wiles had ...

3. Deformation rings and Hecke rings.

In this section we study the following situation. We fix a prime $l \neq 2$ and a continuous irreducible representation

$$\overline{\rho} : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(\mathbf{F}_l).$$

We assume in addition that the determinant of the representation is equal to the cyclotomic character $\varepsilon \pmod{l} : G_{\mathbf{Q}} \longrightarrow \mathbf{F}_l^*$. As a consequence the representation $\overline{\rho}$ is absolutely irreducible. We assume that

- the representation $\overline{\rho}$ is *semi-stable.*
- for every prime $p \neq l$, the image $\overline{\rho}(I_p)$ is contained in a subgroup of $\mathrm{GL}_2(\mathbf{F}_l)$ conjugate to

$$\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}.$$

If E is a semi-stable elliptic curve over $\mathbf{Q}$ for which $E[l]$ is an irreducible Galois module, the representation

$$\overline{\rho}_{E,l} : G_{\mathbf{Q}} \longrightarrow \mathrm{Aut}(E[l])$$

of $G_{\mathbf{Q}}$ enjoys these properties. Wiles views the representation $\rho_{E,l} \longrightarrow \mathrm{Aut}(\mathrm{Ta}_l(E))$ as a *deformation* of $\overline{\rho}$. Here a deformation of $\overline{\rho}$ is a continuous representation ρ

$$\begin{array}{ccc} G_{\mathbf{Q}} & \xrightarrow{\rho} & \mathrm{GL}_2(A) \\ & \overline{\rho} \searrow & \downarrow \\ & & \mathrm{GL}_2(\mathbf{F}_l) \end{array}$$

where A is a complete local Noetherian $\mathbf{Z}_l$-algebra with residue field $\mathbf{F}_l$ which is endowed with the l-adic topology. Moreover, modulo the maximal ideal of A, the representation ρ is conjugate to $\overline{\rho}$.

If the representation $\overline{\rho}$ is the representation of $G_{\mathbf{Q}}$ acting on the l-torsion points $E[l]$ of a semi-stable elliptic curve E over $\mathbf{Q}$, then the basic examples of such deformations are the representations

$$\rho_{E[l^n]} : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(\mathbf{Z}/l^n\mathbf{Z})$$

on the l^n-torsion points of E. The representation on the l-adic Tate-module

$$\rho_{E,l} : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(\mathbf{Z}_l)$$

is another example of a deformation of $\overline{\rho}$. If $\overline{\rho}$ is *modular,* i.e., if the traces of the $\overline{\rho}(\varphi)$ are the coefficients of some eigenform f, then the l-adic respresentation associated to the modular form $\rho_{f,l}$ is yet another example of a deformation of $\overline{\rho}$. All these are deformations of a rather special kind: the representations are at most ramified at a finite set of primes of bad reduction, the local Galois groups G_l act in a rather restricted way, the determinant is given by the cyclotomic character ε ... etc.

The strategy of the proof is to study *all* deformations that satisfy these kind of properties and show that they are *all* modular. This implies then in particular that the deformation $\rho_{E,l}$ is modular. In some sense, this follows from a counting argument: one exhibits many modular deformations of $\overline{\rho}$ and then shows that their number is equal to the number of all deformations.

We need to be more precise about the kind of deformations we will be studying. Roughly speaking, the more restrictive the type of deformations we consider, the more control we have

on them and the easier it becomes to study the collection of deformations. On the other hand, the restrictions should not be so severe that the deformation $\rho_{E,l}$ does not satisfy them and since eventually the representations are supposed to be modular, it seems reasonable to insist that at least the respresentations associated to the candidate modular forms also satisfy the restrictions.

Rather surprisingly, Wiles drops the first requirement : he starts off by studying deformations of such a severely restricted type, that the representation on the l-adic Tate module of the elliptic curve E may not even meet all the requirements. Once these restricted, so-called *minimal*, deformations have been understood, Wiles relaxes the conditions of his deformation problems. This leads to the following notion.

Definition. *Let S be a finite set of primes. An S-deformation of $\overline{\rho}$ is a continuous representation*

$$\rho : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(A)$$

where A is a complete local Noetherian $\mathbf{Z}_l$-algebra with residue field $\mathbf{F}_l$. It satisfies the following conditions.

- *The reduction of ρ modulo the maximal ideal of A is $\overline{\rho}$.*
- *The determinant of ρ is the cyclotomic character $\varepsilon : G_{\mathbf{Q}} \to \mathbf{Z}_l^* \to A^*$.*
- *The restriction of ρ to G_l is semi-stable. In other words G_l acts via a subgroup conjugate to*

$$\begin{pmatrix} \varepsilon & * \\ 0 & 1 \end{pmatrix} \subset \mathrm{GL}_2(A)$$

  or $\rho|_{G_l}$ is flat, which means that for every finite quotient A' of A there is a finite flat group scheme M over $\mathbf{Z}_l$ such that $M(\overline{\mathbf{Q}}_l)$ is a free A'-module of rank 2 with G_l-action via ρ.

At the primes $p \in S$ that are different from l we do not put any restrictions on ρ at all. At the primes $p \notin S$ we insist that the ramification is "as limited as is reasonable":

- *If $p = l$ and $\overline{\rho}$ is flat, then ρ is flat.*
- *If $p \neq l$ and $\overline{\rho}$ is unramified at p , then ρ is unramified at p.*
- *If $p \neq l$ and $\overline{\rho}$ is ramified at p , then $\rho(I_p)$ is contained in a subgroup conjugate to $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$.*

For example, let E be a semi-stable elliptic curve over $\mathbf{Q}$ and let S be the set of primes of bad reduction of E. Then the representation

$$\rho_{E,l} : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{Z}_l)$$

is an S-deformation of the action of $G_{\mathbf{Q}}$ on $E[l]$.

For every set S there exists a *universal* S-representation in the following sense.

Theorem 3.1. *(B. Mazur, R. Ramakrishna) There exists a complete local Noetherian $\mathbf{Z}_l$-algebra R_S with residue field $\mathbf{F}_l$ together with an S-deformation*

$$\rho_S^{\mathrm{univ}} : G_{\mathbf{Q}} \to \mathrm{GL}_2(R_S)$$

which is universal in the sense that for every S-deformation

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(A)$$

with A a complete local Noetherian $\mathbf{Z}_l$-algebra with residue field $\mathbf{F}_l$, there is a unique $\mathbf{Z}_l$-algebra homomorphism $g : R_S \to A$ inducing a representation conjugate to ρ, i.e. $\rho \sim g \cdot \rho_S^{\mathrm{univ}}$.

These universal rings are, in general, poorly understood. One can compute their "cotangent spaces" $\mathfrak{m}_S/(l,\mathfrak{m}_S^2)$ and hence estimate their Krull dimensions. Here $\mathfrak{m}_S$ denotes the maximal ideal of the universal deformation ring R_S. We have

$$\mathrm{Hom}(\mathfrak{m}_S/(l,\mathfrak{m}_S^2),\mathbf{F}_l) \cong H^1_S(G_{\mathbf{Q}},\mathfrak{sl}_2(\overline{\rho})).$$

Here $\mathfrak{sl}_2(\overline{\rho})$ denotes the 3-dimensional $\mathbf{F}_l$-vector space of 2×2-matrices of trace 0. The Galois group $G_{\mathbf{Q}}$ acts on $\mathfrak{sl}_2(\overline{\rho})$ via $\overline{\rho}$ by conjugation. The Galois cohomology group $H^1_S(G_{\mathbf{Q}},\mathfrak{sl}_2(\overline{\rho}))$ is a finite dimensional $\mathbf{F}_l$-vector space and consists of 1-cocycles with restricted local behaviour, reflecting the conditions that the S-deformations of $\overline{\rho}$ must satisfy, modulo certain coboundaries. Given the set S, the cohomology group can be computed explicitly.

From now on we assume in addition that our fixed semi-stable representation

$$\overline{\rho} : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(\mathbf{F}_l)$$

is *modular*. This means that there exists a normalized eigenform $f(\tau) = \sum_{n=1}^{\infty} a(n)e^{2\pi i n\tau}$ such that the characteristic polynomial of the image $\overline{\rho}(\varphi_p)$ of the Frobenius element $\varphi_p \in G_{\mathbf{Q}}$ is, for almost all primes p equal to

$$X^2 - a(p)X + p.$$

The representation $\overline{\rho}$ is ramified at only finitely many primes. Let M denote the product of those primes except for l when $\overline{\rho}$ is flat at l. Difficult results of Carayol, Coleman, Diamond, Edixhoven, Gross, Ribet, Taylor et. al. imply that one can "lower the level" of the form f àla Ribet [2]. More precisely, there exists a modular form f of minimal level M whose Fourier coefficients are congruent to the ones of f modulo l. From now on we will assume that the modular form is of this minimal level M. For simplicity we also assume that that the coefficients of f are in $\mathbf{Z}$. This assumption does not change the discussion in an essential way. Let

$$\rho_{f,l} : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(\mathbf{Z}_l)$$

denote the l-adic representation associated to f.

The representation $\rho_{f,l}$ is a deformation of $\overline{\rho}$. It is a deformation of the most restrictive kind: it is an S-deformation with $S = \emptyset$. This follows from the study of these representations by mathematicians like Shimura, Carayol, Langlands, Deligne-Rapoport, ... By the universal property of the deformation rings there exists therefore a homomorphism of rings

$$h : R_{\emptyset} \longrightarrow \mathbf{Z}_l$$

so that $\rho_{f,l} = h \cdot \rho_{\emptyset}^{\mathrm{univ}}$. The modular form f is, in general, not unique. Even if we restrict our attention to modular forms of minimal level M, there may be several modular forms of level M whose Fourier coefficients are congruent modulo l to the traces of the matrices $\overline{\rho}(\varphi_p)$. For every such modular form g there is a homomorphism of rings $\mathbf{T}(M) \longrightarrow O_g$. Here O_g denotes the $\mathbf{Z}_l$-algebra generated by the coefficients of g. The kernels of all these homomorphisms are contained in the same maximal ideal $\mathfrak{m}$ which is the kernel of the homomorphism

$$\mathbf{T}(M) \xrightarrow{f} \mathbf{Z} \longrightarrow \mathbf{F}_l.$$

We let the *Hecke ring* $\mathbf{T}_{\emptyset}$ denote the completion of $\mathbf{T}(M)$ with respect to $\mathfrak{m}$. It is a local complete $\mathbf{Z}_l$-algebra contained in $\prod_g O_g$. The irreducible components of the spectrum of $\mathbf{T}_{\emptyset}$ are of the form $\mathrm{Spec}(O)$ where O is a subring of the ring of integers in a finite extension of $\mathbf{Q}_l$. The components

are glued at their closed points. Each component corresponds to a modular form of level M whose Fourier coefficients are congruent to those of f modulo l.

Taking all l-adic representations ρ_g together we obtain a representation

$$\rho : G_{\mathbf{Q}} \longrightarrow \mathbf{T}_{\emptyset}$$

which, by a result of Carayol, is again an S-deformation with $S = \emptyset$. It has the property that the characteristic polynomial of $\rho(\varphi)$ is equal to $X^2 - T_p X + p$ for almost all primes p. Here we have written T_p for the image of the Hecke operator T_p in $\mathbf{T}_{\emptyset}$.

By the universal property of $R_{\emptyset}$ we obtain therefore a commutative diagram

$$\begin{array}{ccc} R_{\emptyset} & \longrightarrow & \mathbf{T}_{\emptyset} \\ & \searrow \quad \swarrow & \\ & \mathbf{Z}_l & \end{array}$$

If $S \neq \emptyset$, we can construct Hecke rings $\mathbf{T}_S$ in a similar way: the level N is the lcm of the minimal level M and the squares of the primes in S except for the l-part. The prime l divides N at most once, namely when $l \in S$ or when $\overline{\rho}$ is not flat at l. The modular form f of minimal level gives rise to a homomorphism of rings which we denote by f as well

$$f : \mathbf{T}(N) \longrightarrow \mathbf{T}(M) \longrightarrow \mathbf{Z}$$

Let $\mathfrak{m}$ denote the kernel of the homomorphism

$$\mathbf{T}(N) \xrightarrow{f} \mathbf{Z} \longrightarrow \mathbf{F}_l$$

and let $\mathbf{T}_S$ denote the completion of $\mathbf{T}(N)$ with respect to the maximal ideal $\mathfrak{m}$. There is a representation

$$\rho : G_{\mathbf{Q}} \longrightarrow \mathbf{T}_S$$

which is an S-deformation of $\overline{\rho}$. It has the property that the characteristic polynomial of $\rho(\varphi_p)$ is equal to $X^2 - T_p X + p$ for almost all primes p.

The proofs of these facts are difficult and generalize certain results of B. Mazur. The rings $\mathbf{T}_S$ are finite free $\mathbf{Z}_l$-algebras. To prove the existence of the representations ρ, it is important that they are also *Gorenstein rings.*

For every set S we have a commutative diagram of ring homomorphisms

$$\begin{array}{ccc} R_S & \longrightarrow & \mathbf{T}_S \\ & \searrow \quad \swarrow & \\ & \mathbf{Z}_l & \end{array}$$

Theorem 3.2. *(A. Wiles) For every S, the homomorphism*

$$R_S \longrightarrow \mathbf{T}_S$$

is an isomorphism.

The proof of this result will be discussed in the next sections. It implies that $\mathbf{T}_S$ plays the role of the universal deformation ring. In particular, this implies that for every S-deformation $\rho : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(A)$ there is a unique homomorphism $g : \mathbf{T}_S \longrightarrow A$ inducing ρ. In other words the characteristic polynomials of $\rho(\varphi_p)$ are, for almost all primes p equal to $X^2 - g(T_p)X + p$.

Therefore Theorem 1.1 and the Taniyama-Weil conjecture for semi-stable elliptic curves follow.

4. The minimal deformation problem.

In this section we show how to prove Theorem 3.2 for $S = \emptyset$. The proof of this result is given in the paper by Taylor and Wiles [8] and in the appendix to that paper, which contains a simplification due to Faltings. We remark that for the proof one should actually consider a somewhat more general situation. Wiles only assumes that $\overline{\rho}$ is a $\mathrm{GL}_2(k)$ representation for some finite extension k of $\mathbf{F}_l$ and considers $\mathrm{GL}_2(A)$ represention where A is a complete local Noetherian O-algebra with residue field k. Here O is the ring of integers of a finite extension of $\mathbf{Q}_l$. We ignore this and simply work with $\mathbf{F}_l$ and with $\mathbf{Z}_l$-algebras.

If $\overline{\rho}$ happens to be the representation given by the action of $G_{\mathbf{Q}}$ on the l-torsion points of a semi-stable elliptic curve E, then the level M of the modular form f of section 3 divides the conductor N of E and may actually be a *proper* divisor of N. Indeed, if p is a prime dividor of the conductor N which has the property that l divides the p-adic valuation of the discriminant $\Delta(E)$, then $E[l]$ extends to a finite flat group scheme over $\mathbf{Z}_p$. When $p \neq l$, this means that $\overline{\rho}$ is unramified at p. In any case, such primes p do *not* divide the level M. Therefore the representation $\rho_{E,l}$ is not an S-deformation with $S = \emptyset$ in this case.

It is not very difficult to see that the homomorphism of section 3

$$R_\emptyset \longrightarrow \mathbf{T}_\emptyset$$

is a surjection. It follows from the fact that the Hecke operators that generate $\mathbf{T}_\emptyset$ occur as traces. The problem is to show that the map is an isomorphism. We know that the Hecke ring $\mathbf{T}_\emptyset$ is "small": it is a finite free $\mathbf{Z}_l$-algebra. On the other hand, we know very little about the deformation ring $R_\emptyset$. We tried our best to make it small by putting constraints on the type of deformations we were considering, but we only know it is a Noetherian $\mathbf{Z}_l$-algebra whose Krull dimension we can estimate by means of a Galois cohomology group:

$$\dim_{\mathrm{Krull}} R_\emptyset \leq \dim_{\mathbf{F}_l} H^1_\emptyset(G_{\mathbf{Q}}, \mathfrak{sl}_2(\overline{\rho})) + 1.$$

The important idea is to consider suitable S-deformations with $S \neq \emptyset$ as well. Let S be a finite set of primes $p \equiv 1 \pmod{l}$ at which the representation $\overline{\rho}$ is unramified. Moreover, assume that the Frobenius automorphism φ_p acts via a 2×2-matrix $\overline{\rho}(\varphi_p)$ with *distinct* eigenvalues. Note that these kind of S-deformations of $\overline{\rho}$ do *not* occur as the deformations $\rho_{E,l} : G_{\mathbf{Q}} \longrightarrow \mathrm{Ta}_l(E)$ of some semi-stable elliptic curve E.

Because of these conditions on the primes p, the decomposition groups D_p act via matrices in a subgroup of $\mathrm{GL}_2(\mathbf{F}_l)$ conjugate to $\left(\begin{smallmatrix} * & 0 \\ 0 & * \end{smallmatrix}\right)$. This implies that ρ_S^{univ} restricted to the inertia group I_p is of the form

$$\begin{pmatrix} \chi & 0 \\ 0 & \chi^{-1} \end{pmatrix}$$

where χ is a tame character of the inertia group I_p of l-power order. The universal deformation $\rho_S^{\mathrm{univ}}(I_p)$ induces therefore a homomorphism $\prod_p I_p \longrightarrow R_S^* \hookrightarrow \mathrm{GL}_2(R_S)$ which factors over the maximal quotient of l-power order Δ_S of $\prod_p I_p$. This gives rise to a $\mathbf{Z}_l$-algebra homomorphism $\mathbf{Z}_l[\Delta_S] \longrightarrow R_S$.

There is a commutative diagram of surjective homomorphisms

$$\begin{array}{ccc} R_S & \longrightarrow & \mathbf{T}_S \\ \downarrow & & \downarrow \\ R_\emptyset & \longrightarrow & \mathbf{T}_\emptyset \end{array}$$

The homomorphism $\mathbf{Z}_l[\Delta_S] \longrightarrow R_S$ turns, via this homomorphism, the ring $\mathbf{T}_S$ into an algebra over the complete local ring $\mathbf{Z}_l[\Delta_S]$. The $\mathbf{Z}_l[\Delta_S]$-algebra structure of the Hecke ring $\mathbf{T}_S$ is related to the action of the so-called "diamond" operators. Let I denote the augmentation ideal of $\mathbf{Z}_l[\Delta_S]$. The vertical maps induce isomorphisms $R_S/IR_S \cong R_\emptyset$ and $\mathbf{T}_S/I\mathbf{T}_S \cong \mathbf{T}_\emptyset$.

Because of the *minimality* of the deformation problem associated to $R_\emptyset$, one can do the following:

Let $r = \dim H^1_\emptyset(G_{\mathbf{Q}}, \mathrm{Hom}(\mathfrak{sl}_2(\overline{\rho}), \mu_l))$. For every power l^n there exist a set S of r primes p congruent to 1 (mod l^n) so that both

- $\dim H^1_S(G_{\mathbf{Q}}, \mathfrak{sl}_2(\overline{\rho})) \leq r$; this implies that there exists a surjective $\mathbf{Z}_l$-algebra homomorphism
$$\mathbf{Z}_l[[X_1, \ldots, X_r]] \twoheadrightarrow R_S.$$
- the ring $\mathbf{T}_S$ is a *free* module over the ring of diamond operators $\mathbf{Z}_l[\Delta_S]$.

In other words, the ring R_S has bounded Krull dimension while the ring $\mathbf{T}_S$, being free over $\mathbf{Z}_l[\Delta_S]$, contains a large algebra. Note that, since Δ_S admits $(\mathbf{Z}/l^n\mathbf{Z})^r$ as a quotient, there is a surjective map

$$\mathbf{Z}_l[\Delta_S] \twoheadrightarrow \mathbf{Z}_l[[S_1, \ldots, S_r]]/((S_1+1)^{l^n} - 1, \ldots, (S_r+1)^{l^n} - 1).$$

In this notation, the augmentation ideal I of $\mathbf{Z}_l[\Delta_S]$ is just the image of $(S_1, S_2, \ldots, S_r)$.

The proof depends on some Galois cohomological computations and a clever application of Cebotarev's density theorem. It also depends on an argument of E. de Shalit. See [8] and [1] for the details.

We reduce everything modulo l: let $R = R_S/lR_S$ and $T = \mathbf{T}_S/l\mathbf{T}_S$. The following lemma is due to Karl Rubin [6] and replaces the construction in [8].

Lemma. *Let $\mathbf{F}_l[[X_1, \ldots, X_r]] \twoheadrightarrow R \twoheadrightarrow T$ be surjective morphisms of $\mathbf{F}_l[[S_1, \ldots, S_r]]$-algebras. Suppose that $d = \dim_{\mathbf{F}_l} T/(S_1, \ldots, S_r)T$ is finite. If, for some $k > r^{r-1}d^r$, there is an injective $\mathbf{F}_l[[S_1, \ldots, S_r]]$-algebra morphism*

$$\mathbf{F}_l[[S_1, \ldots, S_r]]/(S_1^k, \ldots, S_r^k) \hookrightarrow T,$$

then the morphism $R \twoheadrightarrow T$ induces an isomorphism

$$R/(S_1, \ldots, S_r)R \cong T/(S_1, \ldots, S_r)T$$

of complete intersection algebras.

We can apply the lemma by taking $p \equiv 1 \pmod{l^n}$ for sufficiently large n. It tells us that $R_\emptyset/lR_\emptyset$ and $\mathbf{T}_\emptyset/l\mathbf{T}_\emptyset$ are isomorphic complete intersection rings. An easy application of Nakayama's Lemma implies then that the homomorphism

$$R_\emptyset \twoheadrightarrow \mathbf{T}_\emptyset$$

is also an isomorphism of complete intersection algebras.

5. From the minimal deformation back to the elliptic curve.

In the previous section we have shown that when $S = \emptyset$, all S-deformations of the given modular representation

$$\overline{\rho} : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(\mathbf{F}_l)$$

are modular a well. This result suffices to prove the Taniyama-Weil conjecture for semi-stable elliptic curves that have square free discriminant and irreducible $E[3]$. In that case, all primes that divide the conductor N of E are also ramified in the representation on the 3-torsion points. In general, however, the representation

$$\rho_{E,l} : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(\mathbf{Z}_l)$$

on the Tate module may be ramified at primes p for which the representation $\overline{\rho}$ is *not* ramified. Therefore it is necessary to prove Theorem 3.2 for all S and not only for the minimal S-deformations with $S = \emptyset$.

Incidentally, this phenomenon is essential in Frey's application of elliptic curves to the proof of Fermat's Last Theorem: consider the representation on the l-torsion points of the Frey curve E associated to a hypothetical solution of the Fermat equation $a^l + b^l = c^l$. In this case the conductor N of E is equal to the product of the prime divisors of abc, but the representation $\overline{\rho}$ is only ramified at 2 and the minimum level of a deformation of $\overline{\rho}$ is 2.

For every set S there is, by Mazur's result, a universal deformation ring R_S which is a complete Noetherian $\mathbf{Z}_l$-algebra with residue field $\mathbf{F}_l$. Since the modular representation

$$\rho : G_{\mathbf{Q}} \longrightarrow \mathrm{GL}_2(\mathbf{T}_S)$$

is an S-deformation, there is a unique homomorphism

$$R_S \longrightarrow \mathbf{T}_S.$$

This homomorphism is surjective and by the universal properties there are commutative diagrams with surjective ring homomorphisms

$$\begin{array}{ccc} R_S & \twoheadrightarrow & \mathbf{T}_S \\ \downarrow & & \downarrow \\ R_\emptyset & \twoheadrightarrow & \mathbf{T}_\emptyset \\ & \searrow \quad \swarrow & \\ & \mathbf{Z}_l & \end{array}$$

To prove Theorem 1.2, we must show that the homomorphism $R_S \longrightarrow \mathbf{T}_S$ is, in fact an isomorphism for every S. The proof is based on a counting argument. We know that $R_\emptyset \longrightarrow \mathbf{T}_\emptyset$ is an isomorphism. Starting from this isomorphism, we "measure" the difference in size between R_S and $R_\emptyset$ and between $\mathbf{T}_S$ and $\mathbf{T}_\emptyset$ respectively. In order to make this precise, we introduce two invariants of local "pointed" $\mathbf{Z}_l$-algebras.

Definition. Let A be a finite free complete local Noetherian $\mathbf{Z}_l$-algebra provided with a $\mathbf{Z}_l$-point, i.e. a section π of the canonical morphism $\mathbf{Z}_l \longrightarrow A$:

$$\pi : A \longrightarrow \mathbf{Z}_l.$$

Let $I = \ker(\pi)$. Then the *cotangent space* of A is defined to be

$$\Phi(A) = I/I^2.$$

The second invariant is the "congruence ideal" $\eta(A) \subset \mathbf{Z}_l$:

$$\eta(A) = \pi(\mathrm{Ann}_A(I)).$$

Example 1. Let $A = \{(x,y) \in \mathbf{Z}_l \times \mathbf{Z}_l : x \equiv y \pmod{l^N}\}$ with $\pi : A \longrightarrow \mathbf{Z}_l$ given by $\pi(x,y) = x$. So, the spectrum of A consists of two copies of $\mathbf{Z}_l$ that are glued to the order l^N at their closed points. We have $A \cong \mathbf{Z}_l[X]/(X(X-l^N))$ with $\pi(X) = 0$ and I is the ideal generated by X. Then I/I^2 is I modulo $(X^2, X(X-l^N)) = (X^2, l^N X)$. It follows that $I/I^2 \cong \mathbf{Z}/l^N\mathbf{Z}$. On the other hand, $\mathrm{Ann}(I)$ is equal to the ideal generated by $X - l^N$, so that $\eta(A) = \pi(\mathrm{Ann}(I)) = l^N\mathbf{Z}_l$.

Example 2. Let $A = \{(x,y,z) \in \mathbf{Z}_l \times \mathbf{Z}_l \times \mathbf{Z}_l : x \equiv y \equiv z \pmod{l}\}$ with $\pi : A \longrightarrow \mathbf{Z}_l$ given by $\pi(x,y,z) = x$. We have $A \cong \mathbf{Z}_l[X,Y]/(X^2 - lX, Y^2 - lY, XY)$ with $\pi(X) = 0$ and $I = (X,Y)$. Then $I/I^2 \cong \mathbf{Z}/l\mathbf{Z} \times \mathbf{Z}/l\mathbf{Z}$. On the other hand, $\mathrm{Ann}(I)$ is the ideal generated by $X + Y - l$, so that $\eta(A) = \pi(\mathrm{Ann}(I)) = l\mathbf{Z}_l$.

It is easy to see that the invariant $\Phi(A)$ is at least as large as $\mathbf{Z}_l/\eta(A)$. Wiles found the following surprising equivalent condition for the equality of these invariants. The present formulation is slightly stronger than Wiles's original statement and is due to H.W. Lenstra [4].

Theorem 5.1. *Let A be a complete local Noetherian pointed $\mathbf{Z}_l$-algebra. Suppose that $\eta(A) \neq 0$. Then*

$$\#\Phi(A) \geq \#\mathbf{Z}_l/\eta(A)$$

and we have equality if and only if A is a complete intersection, i.e. if and only if there is an isomorphism of pointed $\mathbf{Z}_l$-algebras

$$A \cong \mathbf{Z}_l[[X_1, X_2, \ldots, X_r]]/(f_1, f_2, \ldots, f_r)$$

for some power series $f_i \in \mathbf{Z}_l[[X_1, X_2, \ldots, X_r]]$ with $f_i(X) = 0$. Here the "point" is the ring homomorphism $\pi : \mathbf{Z}_l[[X_1, X_2, \ldots, X_r]] \longrightarrow \mathbf{Z}_l$ given by $\pi(X_i) = 0$ for each i.

It is not difficult to derive the following criterion from this.

Criterion. *Suppose that*

$$\varphi : R \longrightarrow T$$

is a surjection of pointed $\mathbf{Z}_l$-algebras and suppose that $\eta(T) \neq 0$ If

$$\#\Phi(R) \geq \#\mathbf{Z}_l/\eta(T)$$

then φ is an isomorphism and both R and T are complete intersections.

By the results of the previous section, we know that

$$\begin{array}{ccc} R_\emptyset & \xrightarrow{\cong} & \mathbf{T}_\emptyset \\ & \searrow \quad \swarrow & \\ & \mathbf{Z}_l & \end{array}$$

is an isomorphism of complete intersections. Moreover, this is an isomorphism of pointed $\mathbf{Z}_l$-algebras via the morphism provided by the modular form f of minimal level. Therefore we have that $\#\Phi(R_\emptyset) = \#\mathbf{Z}_l/\eta(\mathbf{T}_\emptyset)$.

In order to apply the criterion one computes the difference between $\Phi(R_S)$ and $\Phi(R_\emptyset)$ and between $\eta(\mathbf{T}_S)$ and $\eta(\mathbf{T}_\emptyset)$. The difference between the Φ-invariants can be computed by means of the "inflation-restriction" cohomology sequence relating H^1_S and $H^1_\emptyset$. One finds

$$\frac{\#\Phi(R_S)}{\#\Phi(R_\emptyset)} \leq \prod_{p\in S} c_p$$

where c_p is the order of a Galois cohomology group over the residue field $\mathbf{F}_p$. For example, for primes $p \neq l$ at which $\overline{\rho}$ is unramified, one finds

$$\begin{aligned} c_p &= \#H^0(G_{\mathbf{F}_p}, \mathrm{Hom}(\mathfrak{sl}_2(\rho_f \pmod{p^k}), \mu_l)), \\ &= (p\alpha_p/\beta_p - 1)(p-1)(p\beta_p/\alpha_p - 1), \\ &= (p-1)((p+1)^2 - a(p)^2). \end{aligned}$$

Here k is very large and α_p and β_p are the zeroes of the characteristic polynomial $X^2 - a(p)X + p$ of Frobenius φ_p acting on the l-adic representation of $\rho_{f,l}$ associated to the modular form f. Since φ_p acts with eigenvalues α_p/β_p, 1 and β_p/α_p on the $\mathbf{Z}_l$-module $\mathfrak{sl}_2(\rho_{f,l})$ of rank 3, the result follows.

The difference between the η-invariants can also be estimated. Roughly speaking, the quotient of $\#\mathbf{Z}_l/\eta(\mathbf{T}_S)$ by $\#\mathbf{Z}_l/\eta(\mathbf{T}_\emptyset)$ measures the order of glueing between the irreducible components of $\mathrm{Spec}(\mathbf{T}_\emptyset)$ and the additional irreducible components of $\mathrm{Spec}(\mathbf{T}_S)$. Since the components correspond to modular forms, this corresponds to congruences between the "old" modular forms that come with $\mathbf{T}_\emptyset$ and the "new" ones that come with $\mathbf{T}_S$. To estimate $\#\mathbf{Z}_l/\eta(\mathbf{T}_S)/\#\mathbf{Z}_l/\eta(\mathbf{T}_\emptyset)$ one must show that there are many congruences modulo high powers of l. This kind of estimates had already been obtained by Ihara and Ribet. To do the computations one uses the fact that the ring $\mathbf{T}_S$ is known to be a Gorenstein ring.

The result is ... the same:

$$\frac{\#\mathbf{Z}_l/\eta(\mathbf{T}_S)}{\#\mathbf{Z}_l/\eta(\mathbf{T}_\emptyset)} \geq \prod_{p\in S} c_p.$$

Therefore $\#\Phi(R_S) \leq \#\mathbf{Z}_l/\eta(\mathbf{T}_S)$ and by Criterion 4.3 we obtain an isomorphism

$$\begin{array}{ccc} R_S & \xrightarrow{\cong} & \mathbf{T}_S \\ & \searrow \quad \swarrow & \\ & \mathbf{Z}_l & \end{array}$$

as required.

Bibliography

[1] Darmon, H., Diamond, F. and Taylor, R.: Fermat's Last Theorem, in *Current developments in Math, 1995*, International Press, Cambridge MA 1995

[2] Edixhoven, B.: Le rôle de la conjecture de Serre dans la démonstration du théorème de Fermat, *Gazette de Mathématiques* ce volume.

[3] Faltings, G.: The Proof of Fermat's Last Theorem by R. Taylor and A. Wiles, *Notices of the AMS* **42** (1995). 743–746.

[4] Lenstra, H.W.: Complete intersections and Gorenstein rings, in *Conference on Elliptic Curves, Hong-Kong 1993*, International Press, Cambridge MA 1995.

[5] Oesterlé, J.: Travaux de Wiles (et Taylor ...), partie II, in *Sém. Bourbaki*, **804**, 1994–1995, Paris juin 1995.

[6] Rubin, K.: private communication.

[7] Serre, J.-P.: Travaux de Wiles (et Taylor ...), partie I, in *Sém. Bourbaki*, **803**, 1994–1995, Paris juin 1995.

[8] Taylor, R. and Wiles, A.: Ring theoretic properties of certain Hecke rings, *Annals of Math.* **141** (1995), 553–572.

[9] Wiles, A.: Modular elliptic curves and Fermat's Last Theorem, *Annals of Math.* **141** (1995), 443–551.