MODULARITY OF ELLIPTIC CURVES

EHUD DE SHALIT

Notes for a Kazhdan seminar, Spring 2023

Contents

Introduction 2
1. Background 3
1.1. Elliptic curves (week 1) 3
1.2. Modular forms (week 2) 7
1.3. An overview of the proof of Wiles' Modularity Theorem (week 3) 9
2. Deformation theory and Galois cohomology 12
2.1. Galois cohomology of number fields (week 3, continued) 12
2.2. Deformation theory (week 4) 16
2.3. Some examples 22
2.4. Deformation conditions (week 5) 23
3. The universal deformation ring R_Σ 26
3.1. The residual representation 26
3.2. Global deformations of type Σ (Week 6) 29
3.3. Tangent spaces of type Σ and the Greenberg-Wiles formula 30
3.4. Computation of local terms at $p \neq \ell$ 33
3.5. Computation of local terms at ℓ (Week 7) 34
3.6. Fontaine-Laffaille theory 39
3.7. A bound on the number of generators 41
3.8. Taylor-Wiles primes (week 8) 43
4. The Hecke algebra $\mathbb{T}_\Sigma$ and the proof of $R_\Sigma \simeq \mathbb{T}_\Sigma$ 48
4.1. Modularity of the residual representation (Langlands-Tunnell) (week 9) 48
4.2. Some results on the Hecke algebra 51
4.3. The two commutative algebra criteria (week 10) 54
4.4. The proof of the Main Theorem 56
4.5. The 3-5 trick 57
5. Complements on the Hecke algebra (weeks 11,12) 58
5.1. The geometry behind $\mathbb{T}_Q$ 58
5.2. Congruence ideals and Hecke algebras 63
6. Commutative Algebra (weeks 13,14) 65
6.1. The cotangent space and the congruence ideal 65
6.2. Complete intersections and the Gorenstein property 69
6.3. Proof of the first criterion 71
6.4. J-structures and the second criterion 73
References 75

Introduction

The goal of this course is to go over the proof of the following theorem, proved by Andrew Wiles[1] [W95] in 1995.

Theorem 1 (Modularity Theorem). *Let E be a semistable elliptic curve defined over $\mathbb{Q}$. Then E is modular.*

The terms "semistable" and "modular" will be defined in the first chapter, which sets the background for the rest of the course. In that chapter we shall review certain topics from the theory of elliptic curves, modular forms and Galois representations, and will then give a rough overview of the proof.

As is well known, the theorem implies Fermat's Last Theorem. The reduction of Fermat's Last Theorem to the Modularity Theorem is based on a construction of Gerhard Frey and subsequent work of Barry Mazur, Jean-Pierre Serre and Ken Ribet, that predated Wiles' theorem. For lack of time we shall not deal with this spectacular application, and refer the reader to Ribet's original paper [Ri90], and to the survey paper [St97].

The Modularity Theorem is known to hold today without the semistability assumption: *every* elliptic curve over $\mathbb{Q}$ is modular. In this form it apparently originated as a conjecture in 1955 and became known as the Shimura-Taniyama-Weil[2] conjecture. It later became clear that it is an instance of the much more general, still conjectural, Langlands Correspondence.

The generalization of the Modularity Theorem to arbitrary elliptic curves over $\mathbb{Q}$ resulted from a series of improvements on [W95, T-W95]. They started with [Di96], in which the semistability of E was only needed at 3 and 5, and culminated in the work of Christophe Breuil, Brian Conrad, Fred Diamond and Richard Taylor [B-C-D-T], which appeared in 2001, completing the proof of the Shimura-Taniyama-Weil conjecture.

Many more "modularity theorems", of elliptic curves over totally real or CM fields, of K3 surfaces, and of abelian varieties of higher dimension, are known today.

In addition, much progress on related topics followed in the footsteps of Wiles' work. Let us mention (i) Serre's Modularity Conjecture (proved by Khare and Wintenberger in 2008), (ii) the Fontaine-Mazur Conjecture (proved in many cases by Calegari, Dieulefait and Kisin), (iii) Sato-Tate's Conjecture (proved by Barnet-Lamb, Clozel, Geraghty, Harris, Shepherd-Barron and Taylor in two papers from 2008 and 2011), and (iv) Artin's Conjecture on L-functions (of which many new cases now follow from Serre's Conjecture or, independently, from Taylor's work).

I am not up-to-date on all these developments and do not feel qualified to survey them. There are plenty of good introductions and expositions on the web.

[1] Wiles worked on his theorem in isolation for seven years, and announced his result at the Newton Institute in Cambridge in 1993. A few months later, a gap was found in one of the steps of the proof. With the help of Richard Taylor, Wiles changed the strategy dealing with the problematic step, and closed the gap. Technically speaking, the use of "Flach Euler systems" was replaced by a method known today as "Taylor-Wiles patching". The Taylor-Wiles paper [T-W95] appeared as a companion to the main paper by Wiles, and both were published as a special issue of the Annals of Mathematics in 1995.

[2] There is some controversy about who should be credited with it. We included all three mathematicians, in alphabetical order, and refrain from delving into this question.

Our course will follow the survey paper [D-D-T], which in turn follows the original proof, with a few simplifications on the commutative algebra side. There have been several important new ideas introduced in subsequent work, by Fred Diamond, Mark Kisin, Frank Calegari, David Geraghty and others. While the new approaches are absolutely crucial for the generalizations mentioned above, I find the old paper by Darmon, Diamond and Taylor still the best complete introduction to this circle of ideas.

You may want to watch the talk "*Thirty years of modularity*" by Frank Calegari, delivered at the ICM [Cal], for an overview of the activity in this area.

1. Background

1.1. **Elliptic curves (week 1).**

1.1.1. *The Galois representations associated to elliptic curves.* Let F be a number field and E an elliptic curve over F. We denote by

$$\overline{\rho}_{E,\ell} : Gal(\overline{F}/F) \to GL_2(\mathbb{F}_\ell)$$

the representation on $E(\overline{F})[\ell] \simeq \mathbb{F}_\ell^2$ and by

$$\rho^0_{E,\ell} : Gal(\overline{F}/F) \to GL_2(\mathbb{Z}_\ell)$$

the representation on $T_\ell E = \lim_{\leftarrow} E(\overline{F})[\ell^r] \simeq \mathbb{Z}_\ell^2$. We let $\rho_{E,\ell}$ be the representation $\rho^0_{E,\ell}$, followed by the inclusion of $GL_2(\mathbb{Z}_\ell)$ in $GL_2(\mathbb{Q}_\ell)$.

Both $\overline{\rho}_{E,\ell}$ and $\rho_{E,\ell}$ are continuous and well-defined up to conjugation. Furthermore, they are unramified outside $S_{bad}(E) \cup \{v|\ell\}$, where $S_{bad}(E)$ is the finite set of primes of F where E has bad reduction. Let v be a good prime, $v \nmid \ell$, and denote by $\sigma_v \in Gal(\overline{F}/F)$ an (arithmetic) Frobenius at v. Then the characteristic polynomial of $\rho_{E,\ell}(\sigma_v)$ is

$$\det(XI - \rho_{E,\ell}(\sigma_v)) = X^2 - a_v X + q_v = (X - \alpha_v)(X - \alpha'_v)$$

where $q_v = \mathbb{N}v$, $a_v \in \mathbb{Z}$, and is independent of ℓ. That

$$\det(\rho_{E,\ell}(\sigma_v)) = q_v = \epsilon_\ell(\sigma_v),$$

where ϵ_ℓ is the ℓ-adic cyclotomic character, follows from the existence of the Weil pairing. Furthermore, α_v, α'_v are complex conjugate with $|\alpha_v| = |\alpha'_v| = \sqrt{q_v}$. This was conjectured by Emil Artin and proved by Helmut Hasse in 1933. It is often given in terms of the Hasse bound $|a_v| \leq 2\sqrt{q_v}$.

If $\kappa_v = \mathcal{O}_F/v$ is the residue field of the good prime v, and E_v is the reduction of E modulo v, then

$$\#E_v(\kappa_v) = 1 - a_v + q_v,$$

so the Hasse bound estimates the deviation of the number of κ_v-rational points on the reduction from $1 + q_v$.

The representation $\rho_{E,\ell} : Gal(\overline{F}/F) \to GL_2(\mathbb{Q}_\ell)$ is irreducible unless E has CM. However, if $F = \mathbb{Q}$ it is always irreducible.

The representation $\overline{\rho}_{E,\ell}$, in contrast, need not be irreducible or semisimple. We denote by $\overline{\rho}^{ss}_{E,\ell}$ its *semisimplification*. While $\rho^0_{E,\ell} \mod \ell = \overline{\rho}_{E,\ell}$, $\rho_{E,\ell}$ does not determine $\overline{\rho}_{E,\ell}$, only $\overline{\rho}^{ss}_{E,\ell}$ (by the Brauer-Nesbitt theorem).

1.1.2. *Semistable elliptic curves.* Let $v \in S_{bad}$. Let $\mathcal{O}_v$ be the ring of integers of F_v and κ_v its residue field. Let $\mathcal{E}$ be the *Néron model* of E over $\mathcal{O}_v$. Recall that this is a smooth group scheme, whose generic fiber is E, having the following universal property: *for every smooth $\mathcal{O}_v$-scheme S, any $S_\eta = S \times_{\mathcal{O}_v} F_v$-point of E extends uniquely to an S-point of $\mathcal{E}$.* Let $\mathcal{E}_v$ be the special fiber of $\mathcal{E}$ and $\mathcal{E}_v^0$ its connected component. Then there is a short exact sequence of κ_v-group schemes

$$0 \to \mathcal{E}_v^0 \to \mathcal{E}_v \to \Phi_v \to 0,$$

where Φ_v is finite étale over κ_v. The Néron model is unique up to isomorphism.

The curve E also has a unique *minimal regular model* $\overline{\mathcal{E}}$ over $\mathcal{O}_v$, and the Néron model can be identified with the smooth locus of $\overline{\mathcal{E}}$.

E is said to have *semistable (or multiplicative) reduction* at v if $\mathcal{E}_v^0$ is a twisted form of $\mathbb{G}_m$. In this case the special fiber of the minimal regular model $\overline{\mathcal{E}}_v$ becomes, over the quadratic extension κ_v' of κ_v, a polygon of rational curves, intersecting at κ_v'-rational nodes. The group Φ_v becomes, over κ_v', a constant cyclic group.

We say that E has *split multiplicative reduction at v* if $\mathcal{E}_v^0 \simeq \mathbb{G}_m$ already over κ_v. In this case Φ_v is constant cyclic already over κ_v. Otherwise, E is said to have *nonsplit multiplicative reduction.* In this case $\mathcal{E}_v^0$ is isomorphic (as an algebraic group over κ_v) to

$$\ker(Nr : Res_{\kappa_v'/\kappa_v}\mathbb{G}_m \to \mathbb{G}_m)$$

and the non-trivial element of $Gal(\kappa_v'/\kappa_v)$ acts on Φ_v via $x \mapsto x^{-1}$.

If E is not semistable at v then $\mathcal{E}_v^0 \simeq \mathbb{G}_a$ and E is said to have *additive reduction* at v. In this case the structure of the Néron model can be complicated, especially if the residue characteristic of v is 2 or 3.

E is called *semistable* if every $v \in S_{bad}$ is a prime of multiplicative reduction. The semistable reduction theorem implies that every elliptic curve E becomes semistable over a finite extension of F.

Example 2. Let $p > 2$. Then over $\mathbb{Z}_p$, the curve $y^2 = x^3 - x$ has good reduction, and $y^2 = x^3 + x^2 + p$ has (split) multiplicative reduction. The curves $y^2 = x^3 - px$ and $y^2 = x^3 + px^2 + p^4$ both have additive reduction (since their minimal regular model has a cusp in the special fiber) but the first has potentially good reduction (over $\mathbb{Q}_p(p^{1/4})$ it can be written as $(p^{-3/4}y)^2 = (p^{-1/2}x)^3 - (p^{-1/2}x)$, so becomes isomorphic to $y^2 = x^3 - x$), while the second has potentially multiplicative reduction (over $\mathbb{Q}_p(p^{1/2})$ it can be written as $(p^{-3/2}y)^2 = (p^{-1}x)^3 + (p^{-1}x)^2 + p$). Note that $y^2 = x^3 - px$ has complex multiplication by $\mathbb{Z}[i]$. CM elliptic curves always have potentially good reduction.

1.1.3. *Tate's uniformization and the local Galois representation at a place with multiplicative reduction.* Let v be a prime of multiplicative reduction for E. Let $X = Hom(\mathcal{E}^0_{v,\overline{\kappa}_v}, \mathbb{G}_{m,\overline{\kappa}_v})$. Then X is an infinite cyclic group (isomorphic to $\mathbb{Z}$). In the split case, $Gal(\overline{\kappa}_v/\kappa_v)$ acts on X trivially. In the non-split case, it acts via inversion.

Laying the foundations to rigid analytic geometry, Tate found, in the multiplicative case, a uniformization of E, as a rigid analytic space over F_v, by the torus $Hom(X, \mathbb{G}_m^{an})$. The kernel of the uniformization

$$Hom(X, \mathbb{G}_m^{an}) \twoheadrightarrow E^{an}$$

is the subgroup $Hom(X, q_E^{\mathbb{Z}})$, where $q_E \in F_v^\times$, $|q_E|_v < 1$ is the *Tate period.* It is uniquely determined by E and satisfies

$$ord_{F_v}(q_E) = -ord_{F_v}(j_E).$$

Moreover, the relation between the Tate period and the j-invariant j_E is given by a universal power series

$$j_E = q_E^{-1} + \sum_{n=0}^{\infty} c_n q_E^n$$

with $c_n \in \mathbb{Z}$. This power series is nothing but the classical, complex, q-expansion of $j(z)$, viewed p-adically. Thanks to $|q_E|_v < 1$, it converges p-adically. Note that $|j_E|_v > 1$.

The relation between Tate's uniformization and the Néron model can also be made explicit in the multiplicative case. It turns out that over κ'_v the group of connected components Φ_v is cyclic of order $ord_{F_v}(q_E)$. Note that this order is divisible by ℓ if and only if $ord_{F_v}(q_E) \equiv 0 \mod \ell$.

Tate's uniformization has the following consequence regarding the local ℓ-adic and mod-ℓ representations at v.

Proposition 3. *Let E have multiplicative reduction at v. Let η_v be the quadratic unramified character of the decomposition group $G_v = Gal(\overline{F}_v/F_v)$. Then:*

(i) If E has split multiplicative reduction

$$\rho_{E,\ell}|_{G_v} \simeq \begin{pmatrix} \epsilon_\ell & * \\ & 1 \end{pmatrix},$$

while if E has non-split multiplicative reduction

$$\rho_{E,\ell}|_{G_v} \simeq \begin{pmatrix} \epsilon_\ell & * \\ & 1 \end{pmatrix} \otimes \eta_v.$$

These representations are always ramified (i.e. their restriction to I_v is non-trivial).

(ii) If $\ell \neq p$ (the characteristic of v), $\overline{\rho}_{E,\ell}|_{G_v}$ (the representation of G_v on $E(\overline{F}_v)[\ell]$) always has an unramified rank-1 subspace, the quotient by which is also unramified, and is unramified altogether if and only if $ord_{F_v}(q_E) \equiv 0 \mod \ell$.

(iii) Similarly, if $\ell = p$, $\overline{\rho}_{E,p}|_{G_v}$ has a rank-1 subspace associated with a height 1 finite flat group scheme (explicitly, with μ_p or an unramified twist of μ_p), with a quotient of the same type (and even unramified). The representation $\overline{\rho}_{E,p}|_{G_v}$ is "flat" (see below) if and only if $ord_{F_v}(q_E) \equiv 0 \mod p$, if and only if the class in $Ext^1_{G_p}(1, \overline{\epsilon}_p)$ represented by the $*$ is "peu ramifié" in Serre's terminology.

Remark 4. Assume, for simplicity, that E has split multiplicative reduction at v. The splitting field of $\overline{\rho}_{E,\ell}|_{G_v}$ (i.e. the fixed field of $H = \ker(\overline{\rho}_{E,\ell}|_{G_v}) \subset G_v$) is $F_v(\mu_\ell, q_E^{1/\ell})$. It is obtained from F_v in two steps: First, adjoining ℓ-th roots of unity one gets $F_v(\mu_\ell)$ which is unramified if $\ell \neq p$ and tamely ramified if $\ell = p$. Then, adjoining the ℓ-th roots of q_E one gets a Kummer extension of $F_v(\mu_\ell)$. Unless q_E happens to be an ℓ-th power in $F_v(\mu_\ell)$, this is a cyclic extension of degree ℓ.

If $\ell \neq p$, $F_v(\mu_\ell, q_E^{1/\ell})$ is ramified over $F_v(\mu_\ell)$ if $ord_v(q_E)$ is not divisible by ℓ, and is unramified otherwise. When ramified, it is tamely ramified, because its degree $\ell \neq p$.

If $\ell = p$ and $ord_v(q_E)$ is not divisible by p, then $F_v(\mu_p, q_E^{1/p})/F_v(\mu_p)$ is evidently ramified, and "très ramifié" in Serre's terminology. If $\ell = p$ and $ord_v(q_E)$ is divisible by p then $F_v(\mu_p, q_E^{1/p})/F_v(\mu_p)$ is "peu ramifié" (obtained by extracting the pth root of a unit). It may even happen to be unramified. In both the peu/très ramifié cases, if $F_v(\mu_p, q_E^{1/p})/F_v(\mu_p)$ is ramified, it is now wildly ramified, simply because its degree is p.

We make another remark concerning elliptic curves with non-integral j-invariant.

Remark 5. Suppose $|j_E|_v > 1$. Then over a quadratic extension L/F_v the elliptic curve E acquires split multiplicative reduction. If L/F_v is unramified, then E already has multiplicative (possibly non-split) reduction over F_v. However, if L has to be taken ramified, E has additive, potentially multiplicative, reduction at v. On the other hand, when $|j_E| \leq 1$, E has either good or additive potentially good reduction at v.

1.1.4. *The L-function of E/F.* The v-th Euler factor of the L-function of E/F is the evaluation at $X = q_v^{-s}$ of $\det(1 - \sigma_v X|(V_\ell E)_{I_v})$ $(\ell \neq p = char(v))$. In other words, we consider the maximal unramified quotient of the rational ℓ-adic Tate module of E, the (arithmetic) Frobenius $\sigma_v \in G_v/I_v$, and the "characteristic polynomial" $\det(1 - \sigma_v X|(V_\ell E)_{I_v})$. This polynomial, of degree ≤ 2, is independent of ℓ and has $\mathbb{Z}$-coefficients, so we can view it over $\mathbb{C}$ and substitute $X = q_v^{-s}$. By the discussion in the previous sections, it comes out to be:

- $1 - a_v q_v^{-s} + q_v^{1-2s}$ if v is good
- $1 - q_v^{-s}$ if E has split multiplicative reduction at v
- $1 + q_v^{-s}$ if E has non-split multiplicative reduction at v
- 1 if E has additive reduction at v.

Denoting the *inverse of the* Euler factor at v by $L_v(E, s)$ we get, as a result of Hasse's bound, that

$$L(E, s) = \prod_v L_v(E, s)$$

converges absolutely in $Re(s) > 3/2$.

1.1.5. *The Hasse-Weil conjecture.* In general, it is expected, but not known, that $L(E, s)$ admits an analytic continuation to all s, and satisfies a functional equation w.r.t. $s \mapsto 2 - s$. This is called the *Hasse-Weil conjecture*, and without further resriction on F it is known only if E has complex multiplication.

As a result of Wiles' modularity theorem and its generalization, the Hasse-Weil conjecture is known for any E when $F = \mathbb{Q}$. It is also known today (2023) whenever F is real quadratic or totally real cubic, and in many more cases when F is totally real or CM.

Weil's Converse Theorem [We67] said that if $L(E, s)$ and sufficiently many quadratic twists of it satisfied the expected analytic continuation and functional equation, then E was in fact modular. In the early days, this was the strongest evidence in support of the Shimura-Taniyama-Weil conjecture, because the good analytic properties of L-series such as $L(E, s)$ or its quadratic twists, were widely believed to be true.

1.1.6. *The conductor of E.* Let $F = \mathbb{Q}$ for simplicity and consider a prime $p \in S_{bad}(E)$. The *exponent of the conductor* of E at p is an integer $f_p(E/\mathbb{Q}) \geq 1$ that measures how much the ℓ-adic representation $\rho_{E,\ell}$ (for $\ell \neq p$) is ramified at p. While the general definition of the (Artin) conductor of a representation is subtle, and involves the higher ramification groups at p, for elliptic curves we have:

- $f_p(E/\mathbb{Q}) = 1$ if and only if E has multiplicative reduction at p (in which case the ℓ-adic representation is tamely ramified)
- $f_p(E/\mathbb{Q}) \geq 2$ if and only if E has additive reduction at p, and in this case $f_p(E/\mathbb{Q}) = 2$ if $p \neq 2, 3$.

The integer

$$N_E = \prod_{p \in S_{bad}(E)} p^{f_p(E/\mathbb{Q})}$$

is called the conductor of $E/\mathbb{Q}$. It follows from our discussion that E *is semistable if and only if its conductor is square-free.*

1.2. **Modular forms (week 2).**

1.2.1. *Galois representations attached to Hecke eigenforms.* Let $f \in S_k(N, \chi)$. This means that $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ is a Dirichlet character, and $f : \mathfrak{H} \to \mathbb{C}$ is a cusp form of weight k, level $\Gamma_1(N)$ and nebentypus χ.

Assume that f is a *normalized eigenform* of all the Hecke operators. This means that for every $n \geq 1$, $T_n f = a_n \cdot f$ and if

$$f(z) = \sum_{n=1}^{\infty} a_n(f) q^n,$$

$q = e^{2\pi i z}$, then $a_1(f) = 1$. It is then known that $a_n = a_n(f)$. It is also known that $E = \mathbb{Q}(a_n(f))$ is a finite extension of $\mathbb{Q}$, and the $a_n(f) \in \mathcal{O}_E$. Let ℓ be a rational prime, and λ a prime of E lying above ℓ. Let E_λ be the completion of E at λ.

Theorem 6 (Eichler, Shimura, Deligne, Deligne-Serre)**.** *There exists a unique-up-to-conjugation Galois representation $\rho_{f,\lambda} : G_{\mathbb{Q}} \to GL_2(E_\lambda)$ which is unramified outside the primes dividing $N\ell$, such that for every prime $p \nmid N\ell$, if σ_p is a Frobenius automorphism at p, then*

$$\det(XI - \rho_{f,\lambda}(\sigma_p)) = X^2 - a_p(f)X + \chi(p)p^{k-1}.$$

Remark 7. (i) Note that while $\rho_{f,\lambda}$ depended on the choice of λ, the characteristic polynomial of $\rho_{f,\lambda}(\sigma_p)$ did not. This is similar to what we saw for the ℓ-adic Galois representation associated with an elliptic curve.

(ii) If $k = 2$, $\chi = 1$ and $E = \mathbb{Q}$, so the $a_n(f) \in \mathbb{Z}$ and $\lambda = \ell$, we recover characteristic polynomials of the very same shape as those associated with an elliptic curve. This is not a coincidence. In fact, it is not difficult to see from the construction of $\rho_{f,\lambda}$ that in this case $\rho_{f,\lambda}$ *is* a $\rho_{E,\ell}$ for some elliptic curve E defined over $\mathbb{Q}$, with good reduction at any prime $p \nmid N$. If f is a *newform of level* N (an assumption that we can always make since we are interested only in the Hecke eigenvalues away from N), then the conductor N_E of the elliptic curve E associated to f is equal to N, to which one refers sometimes as the *analytic conductor* of E. The equality $N = N_E$ between the analytic and the arithmetic conductors is due to Carayol.

The modularity theorem is a converse to this statement: If E is an elliptic curve over $\mathbb{Q}$, then $\rho_{E,\ell} = \rho_{f,\ell}$ for a rational Hecke-eigenform f of weight 2 and level $\Gamma_0(N)$.

(iii) The theorem follows from the work of Eichler and Shimura when $k = 2$ [Sh58], was extended by Deligne to all $k \geq 2$ [De71] and finally, by Deligne and Serre [De-Se74] to weight $k = 1$.

In weight 2, the construction of $\rho_{f,\lambda}$ can be summarized as follows. Without loss of generality, assume that f is a *newform* of weight N. A construction of Shimura associates to f an abelian variety A_f of dimension $[E:\mathbb{Q}]$, which is a quotient of the Jacobian $J_1(N)$ of the modular curve $X_1(N)$. Via the canonical isomorphisms

$$S_2(\Gamma_1(N)) \simeq H^0(X_1(N),\Omega^1) \simeq H^0(J_1(N),\Omega^1) \simeq T^*J_1(N)|_0,$$

the cotangent space to A_f at 0 is identified with the subspace of $S_2(\Gamma_1(N)) \simeq T^*J_1(N)|_0$ spanned by $\{f^\sigma |\, \sigma \in Emb(E,\mathbb{C})\}$. The abelian variety A_f has endomorphisms by the subring $\mathcal{O} = \mathbb{Z}[a_n(f)] \subset E$. Its rational Tate module

$$V_\ell(A_f) = E \otimes_{\mathcal{O}} \lim_{\leftarrow} A_f(\overline{\mathbb{Q}})[\ell^r]$$

is free of rank 2 over $E_\ell = \mathbb{Q}_\ell \otimes_{\mathbb{Q}} E$. Projecting to E_λ we get the desired representation $\rho_{f,\lambda}$. It is easily seen to be unramified outside $N\ell$. Let σ_p be a Frobenius automorphism at p. The key relation, that for $p \nmid N\ell$ we have

$$Tr(\rho_{f,\lambda}(\sigma_p)) = a_p(f),$$

results from the *Eichler-Shimura congruence relation*

$$T_p \equiv \Pi_p + S_p \circ \Pi_p^t$$

in the ring of correspondences on $X_1(N)_{/\mathbb{F}_p}$. Here Π_p is the relative Frobenius of the curve, T_p the "p-th Hecke operator", i.e. the Hecke operator associated with the matrix $\begin{pmatrix} p & \\ & 1 \end{pmatrix}$, and S_p the "p-th diamond operator", the Hecke operator associated with the matrix $\begin{pmatrix} p & \\ & p \end{pmatrix}$. It should be remarked that the origin of this fundamental relation can be traced back to *Kronecker's congruence relation*

$$\Phi_p(X,j) \equiv (X^p - j)(X - j^p) \mod p.$$

Here $\Phi_p(X,j)$, Kronecker's polynomial, is a primitive polynomial in $\mathbb{Z}[X,j]$, which, viewd as a polynomial in $\mathbb{C}(j)[X]$, gives the monic irreducible polynomial of the function $j(pz)$ over the field $\mathbb{C}(j(z))$.

The construction of the representations $\rho_{f,\lambda}$ for weight $k \geq 2$, which brought with it the proof of Ramanujan's conjecture, was one of the earliest successes of étale cohomology. The extension to $k=1$, by Deligne and Serre, was one of the earliest instances of the method of p-adic deformations of Galois representations.

1.2.2. *Modular elliptic curves.* Let E be an elliptic curve over $\mathbb{Q}$ and N_E its conductor.

Definition 8. E is said to be *modular* if there exists an integer $N \geq 1$, and a rational[3] normalized Hecke eigenform $f \in S_2(N,1)$ such that for some prime ℓ, $\rho_{E,\ell} \simeq \rho_{f,\ell}$.

Tate's isogeny conjecture, proved by Serre for elliptic curves with at least one prime of bad multiplicative reduction, and by Faltings in general, implies that two elliptic curves E and E' over $\mathbb{Q}$ are isogenous over $\mathbb{Q}$ if and only if $\rho_{E,\ell} \simeq \rho_{E',\ell}$. This has the following consequence.

Proposition 9. *For an elliptic curve E over $\mathbb{Q}$, the following are equivalent:*

(1) *E is modular.*

[3]Meaning $a_n(f) \in \mathbb{Q}$.

(2) *There exists an integer $N \geq 1$, and a rational normalized Hecke eigenform $f \in S_2(N,1)$ such that for any prime ℓ, $\rho_{E,\ell} \simeq \rho_{f,\ell}$.*
(3) *E is a quotient of $J_0(N)$.*
(4) *There is a non-zero homomorphism $E \to J_0(N)$.*
(5) *There is a non-constant morphism $X_0(N) \to E$.*

Since the conductor of an elliptic curve is an isogeny-invariant, if the f guaranteed by the definition is *new of level* N, then $N = N_E$.

Since a semisimple 2-dimensional continuous ℓ-adic representation is uniquely determined by the traces of its values on a dense set of Galois automorphisms, and since, by Čebotarev's Theorem, the Frobenii of unramified primes are dense, Condition (2) is equivalent to

$$Tr(\rho_{E,\ell}(\sigma_p)) = a_p(f)$$

for all $p \nmid N\ell$.

Condition (5) is sometimes replaced by the apparently weaker condition that there exists a non-constant holomorphic map $X_0(N)(\mathbb{C}) \to E(\mathbb{C})$. That this implies (5) follows from the fact that if E is modular, so is every quadratic twist of E. Indeed, let D be a fundamental discriminant and $\varepsilon = \left(\frac{D}{\cdot}\right)$ the Legendre symbol modulo D. Then, if $f = \sum_{n=1}^{\infty} a_n q^n$ is a weight 2, rational normalized eigenform of level N, so is $f^\varepsilon = \sum_{n=1}^{\infty} \varepsilon(n) a_n q^n$, of level ND^2. It should be rather surprising, to anybody encountering the conjecture for the first time, that a condition on the existence of a map between Riemann surfaces has implications to Galois representations.

Conjecture 10 (Shimura-Taniyama-Weil)**.** *Every elliptic curve over $\mathbb{Q}$ is modular.*

As was explained in the introduction, [T-W95, W95] proved it for semistable curves, and the proof was completed in [B-C-D-T].

1.3. **An overview of the proof of Wiles' Modularity Theorem (week 3).** Let E be a semistable elliptic curve defined over $\mathbb{Q}$. Wiles' starting point is that for $\ell = 3$ the residual representation $\overline{\rho}_{E,\ell}$ is modular, in the sense that there exists a normalized cuspidal eigenform f (of weight 2 and some level $\Gamma_1(N)$) and a prime $\lambda|\ell$ of $\mathbb{Q}(a_n(f))$, with $\overline{\rho}_{f,\lambda} \simeq \overline{\rho}_{E,\ell}$ over $\mathbb{F}_\ell^{alg}$. This step relies on the fact that $GL_2(\mathbb{F}_3)$ is solvable (indeed, $PGL_2(\mathbb{F}_3) \simeq \mathfrak{S}_4$). By base-change theorems of Langlands and Tunnell, confirming the Artin conjecture in this case, it follows that one can find f of *weight 1* with $\overline{\rho}_{f,\lambda} \simeq \overline{\rho}_{E,\ell}$. A lemma on congruences between modular forms (the *Deligne-Serre Lemma*) allows to shift to weight 2.

Assume now that for some prime $\ell \geq 3$ we know (a) that the global residual representation $\overline{\rho} = \overline{\rho}_{E,\ell}$ is irreducible (b) that $\overline{\rho}$ is modular in the above sense. Note that since $\overline{\rho}$ is *odd* (a) implies that it is in fact absolutely irreducible. Let κ be a finite field over which we realize $\overline{\rho}$, and $k = \kappa^{alg}$ its algebraic closure. Let $W = W(k)$ be the Witt vectors over k. Following Mazur, one would like to construct a *universal deformation ring* $R = R(\overline{\rho})$ for $\overline{\rho}$. This should be a complete local noetherian W-algebra with residue field $R/\mathfrak{m}_R = k$, equipped with a "universal" Galois representation

$$\rho^{univ} : G = Gal(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL_2(R)$$

such that for any lifting ("deformation") $\rho : G \to GL_2(A)$ of $\overline{\rho}$ to a complete local noetherian ring A with residue field k, there exists a unique homomorphism $R \to A$

"bringing ρ^{univ} to ρ", and commuting with the identifications of both reduction with $\overline{\rho}$. For example, there should be such a specialization yielding $\rho_{E,\ell}$.

Now, without further restrictions on the deformations, such an R need not be noetherian, or would be way too large. One would like to impose as many restrictions on the deformations as possible, cutting the size of R, but at the same time accomodating $\rho_{E,\ell}$ as a possible deformation. For example, the deformations should factor through G_S, the Galois group of the maximal extension of $\mathbb{Q}$ which is unramified outside S, where S is a finite set of primes containing the primes of bad reduction of E and the prime ℓ. Their determinant should be the (W-valued) cyclotomic character. And they should satisfy "local conditions" on their restrictions to the decomposition groups G_p at the primes $p \in S$ where they are allowed to ramify. At the primes in $S_{bad}(E)$ these local conditions should be tailored according to the type of bad reduction E has, and here the semistability assumption becomes essential. The most difficult analysis of the local conditions is at the prime $p = \ell$, where one has to analyze "flat" deformations and invoke some results from p-adic Hodge theory, such as Fontaine-Laffaille theory. We emphasize that it is important, for technical reasons explained later, to allow S to be *larger* than $S_{bad}(E) \cup \{\ell\}$.

The assumption that $\overline{\rho}$ was modular yields a certain (complete, local, noetherian) quotient $\mathbb{T} = \mathbb{T}(\overline{\rho})$ of R, which captures all the deformations of $\overline{\rho}$, subject to the set of local conditions, which are "modular". (We have suppressed in the notation, both of R and of $\mathbb{T}$, the set of local conditions, which we shall denote for brevity $\mathcal{L}$.) This $\mathbb{T}$ is a completed Hecke algebra localized at a maximal ideal $\mathfrak{m}$; it is generated over W by "Hecke operators" $\{T_p, S_p | (p, N) = 1\}$, for an integer N which is divisible only by the primes in S and can be calculated from the set of local conditions $\mathcal{L}$. It is obtained by gluing together Hecke algebras acting on weight 2, level N, cuspidal eigenforms g, for which $\overline{\rho}_{g,\lambda} \simeq \overline{\rho}$. Among these g lies our original eigenform f. Thanks to the irreducibility of $\overline{\rho}$, $\mathbb{T}$ is "non-Eisenstein", which implies that it is finite and flat over W. Moreover, it has been known for some time (by Mazur and Tilouine), that the singularities of $\mathbb{T}$ are mild: it tends to be a Gorenstein ring, and in good cases even a local complete intersection (which, for finite flat W-algebras, is stronger than Gorenstein). Furthermore, not only the Hecke algebras glue. By a lemma of Carayol, the representations $\overline{\rho}_{g,\lambda}$ also glue to give a big Galois representation

$$\rho_{\mathfrak{m}} : G_S \to GL_2(\mathbb{T})$$

lifting $\overline{\rho}$.

Since $\rho_{\mathfrak{m}}$ is of type $\mathcal{L}$, we get a surjective homomorphism

$$R \twoheadrightarrow \mathbb{T}$$

of finite flat W-algebras, bringing ρ^{univ} to $\rho_{\mathfrak{m}}$. Our original $\rho_{E,\ell}$ is obtained (when we extend scalars from $\mathbb{Z}_\ell$ to W) by specializing ρ^{univ} via a homomorphism $\pi : R \to W$, while the specializations that factor through the homomorphism to $\mathbb{T}$ are, by construction, the modular ones. We "only" need to show that $R = \mathbb{T}$.

Deformation rings are in general pretty elusive. Remember that both R and $\mathbb{T}$ depended on the set of primes S and the set of local conditions $\mathcal{L}$ at each $p \in S$. One of Wiles' key observations was that while R is difficult to control, when S is enriched by a carefully selected finite set of auxiliary primes q (the "Taylor-Wiles primes"), and the local conditions at these q are appropriately formulated, R becomes gradually "smoother" and more manageable. This is done in a way that does not increase the number of generators of R as a W-algebra, yet increases its "depth", giving more and more room for the diamond operators (the Hecke operators S_q) at the auxiliary prime q. Since these operators appear also in $\mathbb{T}$, one is eventually lead to a proof of a theorem of the type $R_\infty \simeq \mathbb{T}_\infty$, not for R and $\mathbb{T}$ themselves, but for suitable large limits of them (when we keep changing S and $\mathcal{L}$). Then one descends back to the desired equality $R = \mathbb{T}$.

This method, called the "Taylor-Wiles patching", requires (a) comparing the size of R and $\mathbb{T}$ (b) controlling, for either R or $\mathbb{T}$, the way they change when we change S and $\mathcal{L}$. Here enter into the picture tools from Galois cohomology, p-adic Hodge theory and commutative algebra.

One important invariant of a complete noetherian local W-algebra R is its reduced cotangent space

$$\mathfrak{m}_R/(\ell, \mathfrak{m}_R^2).$$

For the universal deformation ring $R = R^{univ}$ this is identified with the k-dual of the Galois cohomology group

$$H^1_{\mathcal{L}}(G_S, Ad(\overline{\rho}))$$

(take $Ad^0(\overline{\rho})$ if the determinant is fixed), where the subscript $\mathcal{L}$ refers to the fact that we only look at cohomology classes satisfying various local conditions. This is a generalized "Selmer group" and its study occupies a great deal of the proof.

Wiles attaches two invariants Φ_A and η_A to a complete local noetherian ring A like R or $\mathbb{T}$, which is equipped in addition with a homomorphism $\pi_A : A \to W$, like the homomorphism yielding $\rho_{E,\ell}$ (when $A = R$). Let $\mathfrak{p}_A = \ker(\pi_A)$, a prime ideal $\subset \mathfrak{m}_A$. The first invariant is the cotangent space "at the point $Spec(W) \to Spec(A)$",

$$\Phi_A = \mathfrak{p}_A/\mathfrak{p}_A^2,$$

which for $A = R^{univ}$ is again dual to some Selmer group. The second is

$$\eta_A = \pi_A(\mathrm{Ann}_A \mathfrak{p}_A).$$

(Following [D-D-T], we shall give some examples of complete local noetherian A's where these two invariants can be calculated easily, to get some feeling for them.) Certain inequalities between the lengths of these invariants and delicate commutative algebra relating them to the singularity of A, give a numerical criterion for $R_\infty \simeq \mathbb{T}_\infty$, which Taylor and Wiles are able to verify (*the first numerical criterion*). This commutative algebra has seen, since the publication of [T-W95], various improvements. In particular, Rubin has given a version that does not require a passge to the infinite limit $R_\infty \simeq \mathbb{T}_\infty$, but works "at a finite level" with a suitable large, but fixed, set of auxilary primes. We might follow Rubin's proof at this stage.

The proof of $R = \mathbb{T}$ breaks into two cases: at first, one starts (prior to introducing the auxiliary Taylor-Wiles primes q) with deformations ρ that are *minimally* ramified. Roughly speaking, ρ involves as little ramification as is forced on us by $\overline{\rho}$. The Taylor-Wiles patching method works best in this set-up, thanks to some numerical coincidences for which I have no a-priori explanation. They just come out of the Galois cohomology computations and may be regarded as a case of good luck (or Divine Providence, depending on one's belief). Getting around these numerical coincidences was, to my understanding, one of the major stumbling blocks in proving higher cases of modularity. The generalization to a *non-minimal* deformation problem (needed to treat cases where there is a prime p where $\rho_{E,\ell}$ is ramified although $\overline{\rho}_{E,\ell}$ is unramified), requires a separate set of tools, and a *second numerical criterion.*

Finally, the whole approach via deformation theory stipulates that $\overline{\rho}_{E,3}$, known to be modular thanks to Langlands-Tunnell, is irreducible. When this is not the case, an ingenious trick (the 3-5 trick) replaces $\overline{\rho}_{E,3}$ by $\overline{\rho}_{E,5}$. Fortunately, for semistable elliptic curves over $\mathbb{Q}$, either $\overline{\rho}_{E,3}$ or $\overline{\rho}_{E,5}$ must be irreducible.

2. Deformation theory and Galois cohomology

This section develops background in Galois representations needed in the proof of the Modularity Theorem. The ultimate goal is to understand the geometry of a certain universal deformation ring R of a residual representation $\overline{\rho} : G_S = Gal(\mathbb{Q}_S/\mathbb{Q}) \to GL_2(\kappa)$, where κ is a finite field. The prototypical example is, of course, $\overline{\rho} = \overline{\rho}_{E,\ell}$. Here S is a finite set of primes containing ∞ and $\ell = char(\kappa)$, and $\mathbb{Q}_S$ is the maximal extension of $\mathbb{Q}$ which is unramified outside S. Following Wiles, we shall study deformations subject to various local conditions at the primes in S, the most subtle ones at the prime $\ell = char(\kappa)$. In addition, it is convenient (although not really necessary) to fix the determinant of all the deformations to be equal to the cyclotomic character ϵ_ℓ, assuming of course that $\det(\overline{\rho}) = \overline{\epsilon}_\ell$. The exact local conditions and the corresponding universal deformation rings will be discussed later.

Besides Mazur's theory of deformations of Galois representations we shall need several deep results from Galois cohomology of number fields. Galois cohomology enters the picture when we try to quantify how R, or rather, its cotangent space, change when we change the local conditions, or enlarge the set S. We therefore start by assembling a quite impressive toolkit from Galois cohomology.

Deformation rings are difficult to analyze. One of Wiles' insights was that the set of local conditions $\mathcal{L}$ has a "dual" set of local conditions $\mathcal{L}^*$. While it is difficult to analyze the deformations subject to each of these sets of conditions separately, it is possible to say something about their "ratio", and this turns out to be enough, thanks to a careful choice of the set of auxiliary primes by which we enlarge S.[4]

Modular forms, Hecke algebras or geometry will not show up in this section or the next one. The discussion will be purely algebraic, relying on the arithmetic of number fields, and on Class Field Theory. General good references are the papers [Co, Maz, Wa].

2.1. Galois cohomology of number fields (week 3, continued).

2.1.1. *Generalities.* References are [Mi, N-S-W]. We consider a profinite group G and a discrete G-module M. Cohomology groups are based on continuous cocycles, so

$$H^i(G,M) = \lim_{\rightarrow} H^i(G/N, M^N)$$

where the limit is over open normal subgroup N in G.

[4]The reader may compare the use of $\mathcal{L}$ and $\mathcal{L}^*$ to the way the Riemann-Roch formula is applied in algebraic geometry. There, the dimension $\ell(D)$ of a linear system $|D|$ is difficult to determine, but the difference $\ell(D) - \ell(D^*)$ where $D^* = K - D$ is easy to compute by an Euler characteristic formula. When $\ell(D^*) = 0$, this yields a precise formula for $\ell(D)$. Although this analogy is only illustrative, here too, the relation between the $\mathcal{L}$-Selmer group and the $\mathcal{L}^*$-Selmer group results from an Euler characteristic formula in Galois cohomology.

For a finite group G we let $\widehat{H}^i(G,M)$ be Tate's cohomology group. For $i>0$ they agree with $H^i(G,M)$, but

$$\widehat{H}^0(G,M)=M^G/N_GM$$

where $N_G=\sum_{\sigma\in G}\sigma$. Note that this only makes sense if G is finite.

Besides the usual tools (long exact sequence, cup product etc.) we shall make use of the *Inflation-Restriction exact sequence.* It is the exact sequence of low terms in the Hochschild-Serre spectral sequence, and runs as follows. Let H be a closed normal subgroup of G, and M a G-module. Then there is a 5-term exact sequence

$$0\to H^1(G/H,M^H)\overset{Inf}{\to}H^1(G,M)\overset{Res}{\to}H^1(H,M)^{G/H}\to$$

$$\to H^2(G/H,M^H)\overset{Inf}{\to}H^2(G,M).$$

Example 11. If I_p is the inertia subgroup of a decomposition group G_p then

$$H^1(G_p/I_p,M^{I_p})=\ker(H^1(G_p,M)\to H^1(I_p,M)).$$

Since G_p/I_p is procyclic, generated by the Frobenius σ_p, for any module X we have $H^1(G_p/I_p,X)=X/(\sigma_p-1)X$. We call

$$H^1(G_p/I_p,M^{I_p})=M^{I_p}/(\sigma_p-1)M^{I_p} \tag{2.1}$$

the *unramified* classes in $H^1(G_p,M)$.

Corollary 12. *Let M be finite. Then*

$$\#H^1(G_p/I_p,M^{I_p})=\#H^0(G_p,M).$$

Proof. This follows from the exact sequence

$$0\to M^{G_p}\to M^{I_p}\overset{\sigma_p-1}{\to}M^{I_p}\to M^{I_p}/(\sigma_p-1)M^{I_p}\to 0$$

and (2.1). □

2.1.2. *Local Tate duality.* Let $G_p=Gal(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ and $\mu\subset\overline{\mathbb{Q}}_p^\times$ the group of roots of unity. If M is a finite G_p-module we let

$$M^*=Hom(M,\mu)$$

with the Galois action $\sigma(h)(x)=\sigma(h(\sigma^{-1}(x))$ $(h\in M^*,\ x\in M)$. For any finite abelian group A we let

$$A^\vee=Hom(A,\mathbb{Q}/\mathbb{Z})$$

be its Pontryagin dual.

Theorem 13. *Let M be a finite G_p-module.*

(i) The groups $H^i(G_p,M)$ are finite and vanish for $i\geq 3$ (the cohomological dimension of G_p is 2).

(ii) For $i=0,1,2$ cup product induces a non-degenerate pairing

$$H^i(G_p,M)\times H^{2-i}(G_p,M^*)\to H^2(G_p,\mu)=Br(\mathbb{Q}_p)=\mathbb{Q}/\mathbb{Z}.$$

Therefore

$$H^i(G_p,M)^\vee\simeq H^{2-i}(G_p,M^*).$$

(iii) If $(p, \#M) = 1$ then $H^1(G_p/I_p, M^{I_p})$ and $H^1(G_p/I_p, M^{*I_p})$ are exact annihilators of each other under the pairing $H^1(G_p, M) \times H^1(G_p, M^*) \to \mathbb{Q}/\mathbb{Z}$. In this case therefore

$$H^1(G_p/I_p, M^{I_p})^\vee \simeq \ker(H^1(I_p, M^*)^{G_p/I_p} \to H^2(G_p/I_p, M^{*I_p})).$$

The infinite prime deserves special attention, and calls for Tate's cohomology.

Proposition 14. *Let M be a $G_\mathbb{R} = \{1, c\}$ module of finite cardinality. Then $\#\widehat{H}^i(G_\mathbb{R}, M)$ are finite. For $i = 0, 1, 2$ cup product induces a non-degenerate pairing*

$$\widehat{H}^i(G_\mathbb{R}, M) \times \widehat{H}^{2-i}(G_\mathbb{R}, M^*) \to \widehat{H}^2(G_\mathbb{R}, \mu) = Br(\mathbb{R}) = \frac{1}{2}\mathbb{Z}/\mathbb{Z}.$$

Note that these cohomology groups all vanish if M has no 2-part.

2.1.3. *Local Euler characteristic.* Let M be a finite G_p-module. Then

$$\frac{\#H^1(G_p, M)}{\#H^0(G_p, M)\#H^2(G_p, M)} = p^{v_p(\#M)}.$$

Taken together, Tate's local duality and the local Euler characteristic reduce the computation of $\#H^i(G_p, M)$ for all i to the computation of $\#H^0(G_p, M)$ and $\#H^0(G_p, M^*)$, which are much easier to calculate in most cases.

2.1.4. *Global Poitou-Tate duality and the 9-term exact sequence.* Let M be a finite $G_\mathbb{Q}$-module. We turn to the cohomology of $G_S = Gal(\mathbb{Q}_S/\mathbb{Q})$ where S is a finite set of primes, containing ∞, the primes ramifying in M and the primes dividing $\#M$. We may therefore regard M and M^* as G_S-modules. Note that since S contains also the primes dividing $\#M = n$, $M^* = Hom(M, \mu_n)$ is also unramified outside S.

The classes of $H^1(G_\mathbb{Q}, M)$ which are *unramified outside* S are, by definition, the classes in

$$H^1(G_S, M) = \ker(H^1(G_\mathbb{Q}, M) \to \prod_{p \notin S} H^1(I_p, M)).$$

The equality follows from $H^1(Gal(\overline{\mathbb{Q}}/\mathbb{Q}_S), M) = Hom(Gal(\overline{\mathbb{Q}}/\mathbb{Q}_S), M)$ and similarly, for $p \notin S$, $H^1(I_p, M) = Hom(I_p, M)$. Since the group generated in $Gal(\overline{\mathbb{Q}}/\mathbb{Q}_S)$ by I_p for all $p \notin S$ is dense, a homomorphism from $Gal(\overline{\mathbb{Q}}/\mathbb{Q}_S)$ to M, all of whose restrtictions to I_p for $p \notin S$ vanish, is 0.

Lemma 15. *The group $H^1(G_S, M)$ is finite.*

Proof. Let K be a finite Galois extension of $\mathbb{Q}$ contained in $\mathbb{Q}_S$ such that G_K fixes M. Inflation-restriction gives an exact sequence

$$0 \to H^1(Gal(K/\mathbb{Q}), M) \to H^1(G_S, M) \to Hom(Gal(\mathbb{Q}_S/K), M).$$

By Class Field Theory, or by Hermite-Minkowski, the last group is finite. The first group is clearly finite. It follows that so is the group in the middle. □

Consider the localization map

$$\alpha_i : H^i(G_S, M) \to \widehat{H}^i(G_\mathbb{R}, M) \times \prod_{p \in S_f} H^i(G_p, M).$$

Using the same map for M^* in degree $2-i$ and then dualizing we get, by Tate's local duality, the map

$$\beta_i = \beta_{i,M} = \alpha^{\vee}_{2-i,M^*} : \widehat{H}^i(G_{\mathbb{R}}, M) \times \prod_{p \in S_f} H^i(G_p, M) \to H^{2-i}(G_S, M^*)^{\vee}.$$

Proposition 16. *α_0 is injective, β_2 is surjective, and for $i = 0, 1, 2$ we have* $\mathrm{Im}(\alpha_i) = \ker(\beta_i)$.

There is also a *global duality* resulting from Global Class Field Theory that we proceed to describe. For any $\mathbb{Q} \subset K \subset \mathbb{Q}_S$, $[K : \mathbb{Q}] < \infty$ let

$$I_{K,S} = \prod_{v \in S_K} K_v^{\times}, \;\; C_{K,S} = I_{K,S}/\mathcal{O}^{\times}_{K,S}$$

($\mathcal{O}_{K,S}$ is the ring of S-integers in K). Let $\mathcal{O}^{\times}_S, I_S$ and C_S denote the direct limits over $K \subset \mathbb{Q}_S$. We then have a short exact sequence of continuous G_S-modules

$$0 \to \mathcal{O}^{\times}_S \to I_S \to C_S \to 0. \tag{2.2}$$

We remark that $C_{K,S}$ is *not* the $Gal(\mathbb{Q}_S/K)$-invariants of C_S, as is the case (by Hilbert's theorem 90) for the classical, unrestricted, sequence obtained by taking the direct limit of $0 \to K^{\times} \to I_K \to C_K \to 0$. Galois cohomology of S-units can be tricky.

For any finite G_S-module M such that S contains the primes p dividing $\#M$, $M^* = Hom(M, \mathcal{O}^{\times}_S)$. Consider the short exact sequence gotten from (2.2) by $Hom(M, -)$ and apply cohomology. We get an exact sequence

$$H^0(G_S, Hom(M, C_S)) \to H^1(G_S, M^*) \xrightarrow{\alpha_1} \prod_{v \in S} H^1(G_v, M^*)$$

(where we identified $H^1(G_S, Hom(M, I_S))$ with $\prod_{v \in S} H^1(G_v, M^*)$, in which there is only one decomposition group for each $v \in S$, by Shapiro's lemma).

For $i = 1, 2$ define $\text{Ш}^i(G_S, M) = \ker(\alpha_i)$. For example, for $i = 1$ these are the cohomology classes that are unramified outside S and trivial at S. From the long exact sequence associated with (2.2) we obtain a surjection

$$H^0(G_S, Hom(M, C_S)) \twoheadrightarrow \text{Ш}^1(G_S, M^*),$$

whose kernel is the image of

$$H^0(G_S, Hom(M, I_S)) \simeq \prod_{v \in S} H^0(G_v, M^*)$$

in $H^0(G_S, Hom(M, C_S))$, again by Shapiro's lemma.

By definition, we also have an injection

$$\text{Ш}^2(G_S, M) \hookrightarrow H^2(G_S, M)$$

and a pairing

$$\begin{array}{ccccccccc} H^0(G_S, Hom(M, C_S)) & \times & H^2(G_S, M) & \xrightarrow{\cup} & H^2(G_S, C_S) & \simeq & \frac{1}{\#G_S}\mathbb{Z}/\mathbb{Z} & \subset & \mathbb{Q}/\mathbb{Z}. \\ \downarrow & & \uparrow & \nearrow & & & & & \\ \text{Ш}^1(G_S, M^*) & \times & \text{Ш}^2(G_S, M) & & & & & & \end{array}$$

(Here $\#G_S$ is a profinite number and the last isomorphism is by the theory of class formations in CFT). By the compatibility between the local and global pairings, the image of $H^0(G_S, Hom(M, I_S)) \simeq \prod_{v\in S} H^0(G_v, M^*)$ in $H^0(G_S, Hom(M, C_S))$ annihilates $\text{Ш}^2(G_S, M)$ so the above pairing induces a pairing

$$\text{Ш}^1(G_S, M^*) \times \text{Ш}^2(G_S, M) \to \mathbb{Q}/\mathbb{Z}. \tag{2.3}$$

Theorem 17 (Poitou-Tate duality)**.** *The pairing* (2.3) *is a perfect pairing between finite abelian groups.*

This is the hardest part of all the duality theorems. Combining Proposition 16 and Theorem 17 we get the following.

Corollary 18 (9-term exact sequence)**.** *Let S be a finite set of primes, and M a finite G_S-module. Assume that S contains the primes dividing $\#M$ and ∞. Then there is an exact sequence*

$$0 \to H^0(G_S, M) \overset{\alpha_0}{\to} \prod_{v\in S} \widehat{H}^0(G_v, M) \overset{\beta_0}{\to} H^2(G_S, M^*)^\vee \to$$

$$\to H^1(G_S, M) \overset{\alpha_1}{\to} \prod_{v\in S} H^1(G_v, M) \overset{\beta_1}{\to} H^1(G_S, M^*)^\vee \to$$

$$\to H^2(G_S, M) \overset{\alpha_2}{\to} \prod_{v\in S} H^2(G_v, M) \overset{\beta_2}{\to} H^0(G_S, M^*)^\vee \to 0.$$

Here the unmarked arrows between the lines are obtained from the identifications

$$\operatorname{coker}(\beta_i) \simeq \ker(\alpha_{2-i})^\vee$$

and the perfect pairings of Theorem 17.

Proof. Exactness at the middle of each row of the diagram follows from Proposition 16. Exactness at the first and last terms is elementary. Exactness at the 3rd, 4th, 6th and 7th terms follows from Theorem 17. □

2.1.5. *The global Euler characteristic formula.*

Theorem 19. *Let M be a finite G_S-module and assume that S contains the infinite prime and all the primes dividing $\#M$. Then*

$$\chi(G_S, M) = \frac{\#H^0(G_S, M)\#H^2(G_S, M)}{\#H^1(G_S, M)} = \frac{\#H^0(G_\mathbb{R}, M)}{\#M}.$$

The proof of both the local and global Euler characteristic formulae go via reduction ot the case of $M = \mu_p$. For this one uses Artin's theorem on induced characters. The case of μ_p is treated by Kummer theory and Class Field Theory. For full details see the references cited above.

2.2. **Deformation theory (week 4).**

2.2.1. *Generalities on deformations.* Let G be a profinite group. We shall need some finiteness assumption on G. The simplest one is to assume that G is topologically finitely generated. Unfortunately, this is not known for $G_S = Gal(\mathbb{Q}_S/\mathbb{Q})$, our main example. Shafarevich conjectured this was the case many years ago, but not all the experts agree, so this shouldn't be stated even as a conjecture. Mazur uses the weaker assumption of *ℓ-finiteness*: For every open $H \subset G$

$$\dim Hom(H, \mathbb{F}_\ell) < \infty.$$

Equivalenly, the maximal pro-ℓ quotient of H is topologically finitely generated. This is known to hold for G_S by Class Field Theory. It is possible to develop deformation theory without this assumption on G, but the deformation rings won't be noetherian in general, and since in our applications the assumption holds, we impose it.

Let E be a finite extension of $\mathbb{Q}_\ell$, $\mathcal{O}$ its ring of integers, λ its maximal ideal and $k = \mathcal{O}/\lambda$.

Let $\mathcal{C}_\mathcal{O}$ be the category of *local complete noetherian* $\mathcal{O}$-algebras R with residue field

$$R/\mathfrak{m}_R = k.$$

Morphisms in $\mathcal{C}_\mathcal{O}$ are local homomorphisms of $\mathcal{O}$-algebras. Every member of $\mathcal{C}_\mathcal{O}$ is of the form

$$R = \mathcal{O}[[X_1, \dots, X_n]]/(f_1, \dots, f_m)$$

where $f_i \in (\lambda, X_1, \dots, X_n)$. If E' is a finite extension of E and $\mathcal{O}'$ is its ring of integers, then $R \mapsto \mathcal{O}' \otimes_\mathcal{O} R$ is a functor from $\mathcal{C}_\mathcal{O}$ to $\mathcal{C}_{\mathcal{O}'}$ (k may change to k'). If R happens to be finite and flat over $\mathcal{O}$, then *possibly after such a base change*, it will admit a section $\pi : R \twoheadrightarrow \mathcal{O}$. A ring $R \in \mathcal{C}_\mathcal{O}$ will be called a *coefficient ring*, and a pair (R, π) as above a *pointed coefficient ring.*

Definition 20. Let $\overline{\rho} : G \to GL_d(k)$ be a continuous representation. A *lifting* (or a *framed deformation*) of $\overline{\rho}$ to $R \in \mathcal{C}_\mathcal{O}$ is a continuous homomorphism

$$\rho : G \to GL_d(R)$$

whose reduction modulo $\mathfrak{m}_R$ is $\overline{\rho}$. Two liftings ρ_1 and ρ_2 are *strictly equivalent* if there exists a $T \in GL_d(R)$, $T \equiv I \mod \mathfrak{m}_R$, such that $\rho_2(\sigma) = T\rho_1(\sigma)T^{-1}$. A *deformation* is a strict equivalence class of framed deformations.

Example 21. Suppose $\overline{\rho} = \overline{\rho}_{f,\lambda}$ is the residual representation associated to some cuspidal Hecke newform $f \in S_k(\Gamma_1(N), \chi)$ and a prime λ of a finite extension $F \supset \mathbb{Q}(a_n(f))$. Here $k = \mathcal{O}_F/\lambda$, $E = F_\lambda$ and $\mathcal{O} = \mathcal{O}_{F,\lambda}$. Suppose $g \in S_k(\Gamma_1(N), \chi)$ is another cuspidal Hecke newform in the same space and $F \supset \mathbb{Q}(a_n(g))$ as well. Suppose the q-expansions of f and g are congruent modulo λ. Then, written in an appropriate basis, $\rho_{g,\lambda} : G_S \to GL_2(\mathcal{O})$ is a deformation of $\overline{\rho}$.

We define the framed and unframed deformation functors as follows.

Definition 22. The framed deformation functor

$$D^\square_{\overline{\rho}} : \mathcal{C}_\mathcal{O} \rightsquigarrow \text{Sets}$$

is the (covariant) functor associating to $R \in \mathcal{C}_\mathcal{O}$ the set of framed deformations of $\overline{\rho}$ to R. The deformation functor is the functor $D_{\overline{\rho}}$ associating to R the set of deformations of $\overline{\rho}$ to R.

We remark that $D^\square_{\overline{\rho}}$ is continuous, in the sense $D^\square_{\overline{\rho}}(R) = \lim_\leftarrow D^\square_{\overline{\rho}}(R/\mathfrak{m}_R^n)$, so $D^\square_{\overline{\rho}}$ is determined by its restriction to the full subcategory $\mathcal{A}r_\mathcal{O}$ of $\mathcal{C}_\mathcal{O}$ of Artinian objects. The same holds for $D_{\overline{\rho}}$.

Recall that a (covariant) functor $\mathcal{F}$ from $\mathcal{C}_\mathcal{O}$ to Sets is *representable* if there exists an object $R^{univ} \in \mathcal{C}_\mathcal{O}$ and a natural equivalence of functors

$$\mathcal{F}(-) \simeq Hom_{\mathcal{C}_\mathcal{O}}(R^{univ}, -).$$

The element $\rho^{univ} \in \mathcal{F}(R^{univ})$ corresponding to the identity morphism of R^{univ} is called then the universal object (in our case, universal framed deformation or universal deformation). It is characterized by the property that for every $A \in \mathcal{C}_\mathcal{O}$

and for any $\rho \in \mathcal{F}(A)$ there exists a unique homomorphism $R^{univ} \to A$ "bringing ρ^{univ} to ρ".

If $\mathcal{F}$ is representable, then the pair (R^{univ}, ρ^{univ}) representing it is unique up to a unique isomorphism.

2.2.2. *Representability of the deformation functors.*

Theorem 23. *Suppose that G satisfies the condition of ℓ-finiteness. Then $D_{\overline{\rho}}^{\square}$ is representable.*

Proof. (Easy) Let $H = \ker(\overline{\rho})$ and $N \triangleleft H$ the closed normal subgroup such that H/N is the maximal pro-ℓ quotient of H (N is the intersection of all the closed normal subgroups U such that H/U is pro-ℓ; since H is profinite, we may even let U run over *open* normal subgroups with this property). The closed group N is normal in G as well, and G/N is topologically finitely generated, because H/N is t.f.g., and G/H is finite. Let $\gamma_1, \ldots, \gamma_g \in G/N$ be topological generators. Let $W = W(k) \subset \mathcal{O}$ and let $[\overline{\rho}(\gamma_i)] \in GL_d(W)$ be Teichmüller lifts of $\overline{\rho}(\gamma_i)$ (lift every entry). Note that if $\rho \in D_{\overline{\rho}}^{\square}(R)$ then $N \subset \ker(\rho)$, because $\rho(H) \subset \ker(GL_d(R) \to GL_d(k))$, which is pro-$\ell$. Thus ρ factors through G/N. We may therefore define elements $x_{\alpha,\beta}^{(i)} \in \mathfrak{m}_R$ by

$$\rho(\gamma_i) = [\overline{\rho}(\gamma_i)](I_d + (x_{\alpha,\beta}^{(i)})).$$

Let $X_{\alpha,\beta}^{(i)}$ be commuting variables. Let F_g be the free profinite group on the symbols $\{\gamma_i\}_{i=1}^{g}$ and define

$$r : F_g \to GL_d(\mathcal{O}[[X_{\alpha,\beta}^{(i)}]])$$

by the above formula, with the gd^2 variables $X_{\alpha,\beta}^{(i)}$ replacing the $x_{\alpha,\beta}^{(i)}$. Let $\widetilde{N} \triangleleft F_g$ be the kernel of the canonical surjection $F_g \twoheadrightarrow G/N$, so that $F_g/\widetilde{N} = G/N$. Let I be the ideal of $\mathcal{O}[[X_{\alpha,\beta}^{(i)}]]$ generated by the entries of $r(\sigma) - I_d$ for all $\sigma \in \widetilde{N}$.

It is easy to see that $R^{univ} = \mathcal{O}[[X_{\alpha,\beta}^{(i)}]]/I$ and $\rho^{univ} =$ image of r, are the universal objects representing $D_{\overline{\rho}}^{\square}$. Indeed, as we have seen, any $\rho \in D_{\overline{\rho}}^{\square}(R)$ factors through $F_g/\widetilde{N} = G/N$, and there are unique $x_{\alpha,\beta}^{(i)} \in \mathfrak{m}_R$ such that

$$\rho(\gamma_i) = [\overline{\rho}(\gamma_i)](I_d + (x_{\alpha,\beta}^{(i)})).$$

It follows that ρ is obtained from ρ^{univ} by the unique specialization $R^{univ} \to R$ taking $X_{\alpha,\beta}^{(i)}$ to $x_{\alpha,\beta}^{(i)}$. $\square$

The representability of $D_{\overline{\rho}}$ is more challenging. There are several ways to prove the following theorem.

Theorem 24. *Assume, in addition, that $End_{k[G]}(\overline{\rho}) = k$ (e.g. that $\overline{\rho}$ is absolutely irreducible). Then $D_{\overline{\rho}}$ is representable.*

(i) Kisin proves the theorem by observing that under the given assumptions, the formal group $\widehat{PGL_d}$ acts *freely* on the functor $D_{\overline{\rho}}^{\square}$ and the quotient is $D_{\overline{\rho}}$. He then applies some results from SGA to deduce the theorem.

(ii) Faltings gave a direct proof in the spirit of the proof of the representability of the framed deformations functor. The problem now is the ambiguity in the model of ρ, since it is only defined up to strict equivalence. To overcome it, Faltings proved that any deformation $[\rho]$ of $\overline{\rho}$ has a *unique* representative $\rho : G \to GL_d(R)$ which is "well-placed" in the sense that the vector $v_\rho = (\rho(\gamma_i) - [\overline{\rho}(\gamma_i)])_{i=1}^g$ lies in $V(R)$ for a certain fixed $\mathcal{O}$-submodule $V \subset M_{d,\mathcal{O}}^g$. See Theorem 2.28 and Lemma 2.29 in [D-D-T] for details.

(iii) The original approach of Mazur and Ramakrishna [Ra, Maz] was to verify the four conditions given by Schlessinger for representability. We explain these conditions below. Let $k[\varepsilon]$ be the ring of dual numbers over k.

Suppose that $\mathcal{F} : \mathcal{C}_\mathcal{O} \to \text{Sets}$ is a covariant continuous functor. If $\mathcal{F}$ is representable by R then:

(1) $\mathcal{F}(k)$ is a singleton, since $Hom_{\mathcal{C}_\mathcal{O}}(R,k)$ is a singleton.
(2) If $A \to C$ and $B \to C$ are arrows in $\mathcal{A}r_\mathcal{O}$ then

$$\mathcal{F}(A \times_C B) = Hom_{\mathcal{C}_\mathcal{O}}(R, A \times_C B)$$
$$= Hom_{\mathcal{C}_\mathcal{O}}(R,A) \times_{Hom_{\mathcal{C}_\mathcal{O}}(R,C)} Hom_{\mathcal{C}_\mathcal{O}}(R,B) = \mathcal{F}(A) \times_{\mathcal{F}(C)} \mathcal{F}(B).$$

(3) $\dim \mathcal{F}(k[\varepsilon]) < \infty$. Here $\mathcal{F}(k[\varepsilon])$ is given the structure of a k-vector space by means of the maps $\mathcal{F}([+])$ and $\mathcal{F}([\alpha])$, $(\alpha \in k)$ obtained by functoriality from the ring homomorphisms

$$[+] : k[\varepsilon] \times_k k[\varepsilon] \to k[\varepsilon], \ \ (a+b\varepsilon, a+c\varepsilon) \mapsto a + (b+c)\varepsilon$$
$$[\alpha] : k[\varepsilon] \to k[\varepsilon], \ \ a + b\varepsilon \mapsto a + \alpha b\varepsilon.$$

The reason for $\dim_k \mathcal{F}(k[\varepsilon]) < \infty$ is that if $\mathcal{F}$ is representable by R then

$$\mathcal{F}(k[\varepsilon]) = Hom_{\mathcal{C}_\mathcal{O}}(R, k[\varepsilon]) = Hom_k(\mathfrak{m}_R/(\mathfrak{m}_R^2, \lambda), k)$$

and $\dim_k \mathfrak{m}_R/(\mathfrak{m}_R^2,\lambda) < \infty$ since R is noetherian.

Definition 25. The vector space $\mathcal{F}(k[\varepsilon])$ is called the *tangent space* of $\mathcal{F}$.

Theorem 26 (Grothendieck)**.** *Conversely, if $\mathcal{F} : \mathcal{C}_\mathcal{O} \to$ Sets is a continuous covariant functor satisfying* (1) – (3), *then $\mathcal{F}$ is representable.*

In his thesis, Schlessinger replaced (2) by three special cases that are easier to check. A homomorphism $A \to C$ in $\mathcal{A}r_\mathcal{O}$ is called *small* if it is surjective, and its kernel is principal and annihilated by $\mathfrak{m}_A$.

Theorem 27 (Schlessinger's criteria)**.** *Let $\mathcal{F} : \mathcal{C}_\mathcal{O} \to$ Sets be a continuous covariant functor satisfying:*

(1) *$\mathcal{F}(k)$ is a singleton,*
(2) *Consider $\alpha : A \to C$, $\beta : B \to C$ arrows of $\mathcal{A}r_\mathcal{O}$ and the set map*

$$\phi : \mathcal{F}(A \times_C B) \to \mathcal{F}(A) \times_{\mathcal{F}(C)} \mathcal{F}(B)$$

induced by functoriality. Suppose that
   (a) *If α is small, then ϕ is a surjection,*
   (b) *If $A = k[\varepsilon]$ and $C = k$ then ϕ is bijective,*
   (c) *If $A = B$ and $\alpha = \beta$ is small, then ϕ is bijective,*
(3) $\dim_k \mathcal{F}(k[\varepsilon]) < \infty$.

Then $\mathcal{F}$ is representable.

Mazur proved the representability of $D_{\overline{\rho}}$ by verifying Schlessinger's criteria for the deformation functor. See [Maz, Ra]. Perhaps the least trivial is point (3), which follows from the cohomological interpretation of the tangent space $D_{\overline{\rho}}(k[\varepsilon])$. We discuss it next.

2.2.3. *The tangent space of the deformation functor.* Let $Ad\overline{\rho} = M_d(k)$ with the adjoint action of G, i.e.

$$Ad\overline{\rho}(\sigma)X = \overline{\rho}(\sigma)X\overline{\rho}(\sigma)^{-1}.$$

Let $Ad^0\overline{\rho}$ be the subrepresentation of trace-0 matrices.

Suppose $\rho : G \to GL_d(k[\varepsilon])$ lifts $\overline{\rho}$ and write

$$\rho(\sigma) = (1 + \varepsilon c(\sigma)) \cdot \overline{\rho}(\sigma).$$

Then $c : G \to M_d(k)$ is continuous and satisfies

$$c(\sigma\tau) = Ad\overline{\rho}(\sigma)(c(\tau)) + c(\sigma) = \sigma c(\tau) + c(\sigma)$$

(the cocycle condition). Thus $c \in Z^1(G, Ad\overline{\rho})$. It can be checked that the k-vector space structures of $D^{\square}_{\overline{\rho}}(k[\varepsilon])$ and $Z^1(G, Ad\overline{\rho})$ agree. Furtheremore, changing ρ by strict equivalence gets translated to changing c by a coboundary. Conversly, given $c \in Z^1(G, Ad\overline{\rho})$, the above formula gives a lift to $k[\varepsilon]$. We get the following result, relating the tangent spaces of the deformation functors to Galois cohomology.

Proposition 28. *There is a canonical isomorphism of vector spaces*

$$D^{\square}_{\overline{\rho}}(k[\varepsilon]) \simeq Z^1(G, Ad\overline{\rho}), \quad D_{\overline{\rho}}(k[\varepsilon]) \simeq H^1(G, Ad\overline{\rho}).$$

If R represents $D^{\square}_{\overline{\rho}}$ then $D^{\square}_{\overline{\rho}}(k[\varepsilon]) = Hom_{\mathcal{C}_{\mathcal{O}}}(R, k[\varepsilon]) \simeq Hom_k(\mathfrak{m}_R/(\mathfrak{m}_R^2, \lambda), k)$. Similarly for $D_{\overline{\rho}}$, in case it is representable.

The second assertion follows from the fact that $R = \mathcal{O} + \mathfrak{m}_R$, $\mathcal{O} \cap \mathfrak{m}_R = \lambda$, and a local homomorphism $R \to k[\varepsilon]$ lifting the identity on k is determined by its restriction to $\mathfrak{m}_R$, which is a k-linear map of $\mathfrak{m}_R/(\mathfrak{m}_R^2, \lambda)$ to k. Conversely, any such k-linear map determines a local homomorphism $R \to k[\varepsilon]$ lifting the identity on k.

If G is ℓ-finite (e.g. if $G = G_S$ in the arithmetic application we have in mind) it can be shown easily, using the inflation-restriction exact sequence, that $H^1(G, Ad\overline{\rho})$ is finite dimensional. The cohomological interpretation of the tangent space $D_{\overline{\rho}}(k[\varepsilon])$ given by the proposition implies then condition (3) in Schlessinger's theorem. Condition (1) is automatic, and the conditions (2)(a-c) are not difficult to verify.

2.2.4. *Relation between the framed and non-framed deformation rings.* Suppose that $End_{k[G]}(\overline{\rho}) = k$. Let R be the universal deformation ring representing $D_{\overline{\rho}}$ and $R^{\square}$ the universal framed deformation ring representing $D^{\square}_{\overline{\rho}}$. Let ρ^{univ} and $\rho^{\square,univ}$ be the universal deformation / framed deformation. The strict equivalence class of $\rho^{\square,univ}$ is an element of $D_{\overline{\rho}}(R^{\square})$, so corresponds to a canonical homomorphism

$$\iota : R \to R^{\square}$$

in $\mathcal{C}_{\mathcal{O}}$. This homomorphism is *formally smooth* (e.g. $R^{\square}$ could be a power series ring in some number of variables over R, and this would indeed be the case if $R = k$ and in a few other cases). Recall that being formally smooth means, in our context, that for any $B \in \mathcal{C}_{\mathcal{O}}$ and I an ideal of B with $I^2 = 0$, a framed B/I-deformation whose strict equivalence class lifts to B, lifts to B. This is clear because a lifting to B of the strict equivalence class of the deformation is, by defintion, a strict equivalence class of liftings.

Moreover, let

$$T \in \ker(GL_d(R^{\square}) \to GL_d(k)).$$

Then $T\rho^{\square,univ}T^{-1}$ is another lifting of $\overline{\rho}$ to $R^{\square}$, so there should be a homomorphism, in fact an automorphism,

$$\theta_T \in Aut(R^{\square})$$

bringing $\rho^{\square,univ}$ to $T\rho^{\square,univ}T^{-1}$. As these two representations are, by definition, strictly equivalent, $\theta_T \circ \iota = \iota$. In fact R should be the subring of $R^{\square}$ invariant by all such θ_T [??].

On the other hand, any representative ρ of ρ^{univ} (a strict equivalence class of representations), is an R-valued lift of $\overline{\rho}$, so determines a homomorphism

$$\pi_\rho : R^{\square} \to R,$$

bringing $\rho^{\square,univ}$ to ρ, and it is easily checked that $\pi_\rho \circ \iota = id_R$. Thus the choice of ρ allows us to regard R as an $R^{\square}$-algebra.

Exercise. Show that $\pi_\rho \circ \theta_T = \pi_{\pi_\rho(T)\rho\pi_\rho(T)^{-1}}$.

2.2.5. *Generators and relations.* Suppose that $End_{k[G]}(\overline{\rho}) = k$ and R is the ring that represents $D_{\overline{\rho}}$. It can be shown then, with the aid of Nakayama's lemma, and the computation we did of the tangent space, that R has the following structure

$$R \simeq \mathcal{O}[[X_1, \dots, X_g]]/(f_1, \dots, f_r)$$

where $g = \dim_k H^1(G, Ad\overline{\rho})$ and $r = \dim_k H^2(G, Ad\overline{\rho})$. See the survey of *obstruction theory* in [Maz], 1.6 for the emergence of H^2, or consult, more generally, chapter 6, "Elementary Deformation Theory", in [FGA].

This implies the following inequality for the Krull-dimension

$$\dim(R) \geq 1 + g - r.$$

Mazur raised the question whether, in the number field case, an equality always holds here. Fernando Gouvêa stated it as a conjecture.

Conjecture 29. *(Mazur-Gouvêa) Assume $G = G_S$ is the Galois group of the maximal unramified-outside-S extension of $\mathbb{Q}$, where S is a finite set that contains ∞, ℓ. Assume that $\overline{\rho}$ is absolutely irreducible and let R be the universal deformation ring of $\overline{\rho}$. Then all the irreducible components of $Spec(R)$ have the same Krull dimension, and equality holds*

$$\dim(R) = 1 + h_1 - h_2$$

where $h_i = \dim_k H^i(G_S, Ad\overline{\rho})$.

Already the case $d = 1$ of this conjecture, for a general totally real field F replacing $\mathbb{Q}$, is equivalent to Leopoldt's conjecture. The conjecture must therefore be very hard. In fact, in 2013 Sprang found a counterexample to the conjecture, if G_S is replaced by an arbitrary profinite group satisfying the ℓ-finiteness condition. If Gouvêa's conjecture is true, it must be because of delicate arithmetic, and not a pure algebra result.

If Gouvêa's conjecture holds, the universal deformation ring R is a local complete intersection, because it is "cut" in the regular local ring $\mathcal{O}[[X_1, \dots, X_{h_1}]]$ by as many elements as its codimension. One of Wiles' achievements was to prove that a certain (restricted, see the next section) deformation ring is a local complete intersection. He did it, however, in a roundabout way, and only as a consequence of identifying R with a certain Hecke algebra.

The deformation problem is called *unobstructed,* when $h_2 = 0$. Gouvêa's conjecture is then clearly satisfied, and the universal deformation space is *formally smooth:* it is a power series in h_1 variables over $\mathcal{O}$.

2.3. **Some examples.**

2.3.1. $d = 1$. When $d = 1$ it is possible to obtain an explicit description of the universal deformation ring. It depends then only on G and k and not on the character $\overline{\rho}$. More generally, for any d Mazur shows that R^{univ} depends, up to a canonical isomorphism, only on the twisted-conjugacy class of $\overline{\rho}$.

Let Γ be the pro-ℓ completion of $G^{ab} = G/[G,G]$. By assumption, it is a finitely generated $\mathbb{Z}_\ell$-module. Let

$$\rho_0 : G \to W(k)^\times \subset \mathcal{O}^\times$$

be the Teichmüller lift of the character $\overline{\rho}$. Let $R = \mathcal{O}[[\Gamma]]$ (the Iwasawa algebra) and consider

$$\rho : G \to R^\times = GL_1(R),$$

defined by $\rho(g) = \rho_0(g)[g]$, where $[-] : G \twoheadrightarrow \Gamma \subset \mathcal{O}[[\Gamma]]^\times$ is the canonical homomorphism.

Proposition 30. *(R, ρ) is the universal deformation ring of $\overline{\rho}$.*

2.3.2. *Global representations.* We turn to $d = 2$. Assume that $\ell > 2$ and $G = G_S$ where $S = \{\infty, \ell\}$, the Galois group of the maximal extension of $\mathbb{Q}$ which is unramified outside ∞ and ℓ. This is (when S is more general), by far the most interesting case of the abstract theory. In this case, the global Euler characteristic formula allows us to obtain bounds on $\dim(R)$ (similar bounds can be obtained if $G = G_{F,S}$ for any number field F and a finite set of places S of F). Let

$$\overline{\rho} : G_S \to GL_2(\mathbb{F}_\ell)$$

be absolutely irreducible. The global Euler characteristic formula yields, quite easily, the following.

Proposition 31. *In this set-up, $h_1 - h_2 = 3$ if $\overline{\rho}$ is odd, and $h_1 - h_2 = 1$ if $\overline{\rho}$ is even.*

Nigel Boston and Mazur found examples where $\overline{\rho}$ is odd, $h_1 = 3$, $h_2 = 0$ and as a result the deformation problem is unobstructed, and the universal deformation ring is formally smooth.

Let ℓ be a prime of the form $\ell = 27 + 4a^3$, e.g. $\ell = 23, 31, 59, 283, 1399$. Let K be the cubic field $\mathbb{Q}(x)$ where x is a root of $x^3 + ax + 1 = 0$. Its discriminant is $-\ell$. Its Galois closure L is an $\mathfrak{S}_3$-extension of $\mathbb{Q}$. Let $S = \{\infty, \ell\}$. Let $\overline{\rho} : G_S \twoheadrightarrow Gal(L/\mathbb{Q}) \hookrightarrow GL_2(\mathbb{F}_\ell)$.

Proposition 32. *In this set-up, the universal deformation ring of $\overline{\rho}$ (with $G = G_S$) is isomorphic to $\mathbb{Z}_\ell[[T_1, T_2, T_3]]$.*

2.3.3. *Local representations.* Let p be a prime that may or may not be equal to ℓ and $G = G_p$, the absolute Galois group of $\mathbb{Q}_p$. Let

$$\overline{\rho} : G_p \to GL_2(\mathbb{F}_\ell)$$

be absolutely irreducible. Again, the local Euler characteristic formula gives an easy control of $h_1 - h_2$. Since $h_0 = \dim H^0(G_p, Ad(\overline{\rho})) = 1$ by Schur's lemma, we get

Lemma 33. *$h_1 - h_2 = 5$ if $\ell = p$ and $h_1 - h_2 = 1$ if $\ell \neq p$.*

In his thesis, Ramakrishna [Ra] showed that if $\ell = p$ the local deformation problem ($d = 2$, $k = \mathbb{F}_\ell$, $G = G_\ell$) is in fact unobstructed.

Theorem 34. *Let $G = Gal(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)$, $\ell > 2$. Let $\overline{\rho} : G \to GL_2(\mathbb{F}_\ell)$ be absolutely irreducible. Then the universal deformation problem of $\overline{\rho}$ is unobstructed and*

$$R^{univ} \simeq \mathbb{Z}_\ell[[T_1, \dots, T_5]].$$

To prove the theorem, Ramakrishna first finds an explicit model for $\overline{\rho}$ using Serre's second fundamental character on the tame inerta, and then shows that $H^2(G, Ad(\overline{\rho}))$, which by local Tate duality is dual to $H^0(G, Ad(\overline{\rho})(1))$, vanishes.

2.3.4. *Relation between the local and global deformation problems.* Let's put ourselves, to be explicit, in the situation where $G = G_S$, S contains ℓ, and $\overline{\rho} : G_S \to GL_2(\mathbb{F}_\ell)$ is (absolutely) irreducible and odd. Let (R_S, ρ_S) be the global universal deformation ring. By the above, its relative Krull dimension (over $\mathbb{Z}_\ell$) is bounded below by 3, and conjectured to be equal to 3. Assume that $\overline{\rho}|_{G_\ell}$ is still absolutely irreducible and let (R_ℓ, ρ_ℓ) be its universal deformation ring. By Ramakrishna's theorem, R_ℓ is a power series ring in 5 variables over $\mathbb{Z}_\ell$. Since $\rho_S|_{G_\ell}$ is a deformation of $\overline{\rho}|_{G_\ell}$, we get a homomorphism $R_\ell \to R_S$ "bringing ρ_ℓ to $\rho_S|_{G_\ell}$". This corresponds to a morphism

$$Spec(R_S) \to Spec(R_\ell).$$

Many questions arise: Is this morphism finite over its image? Assuming the relative dimensions are 3 and 5, what are the two conditions characterizing the image? How does it change when we increase S?

2.4. **Deformation conditions (week 5).**

2.4.1. *Abstract framed and non-framed deformation problems.* We shall need to study deformations restricted in certain ways (in the case $G = G_S$, by imposing local conditions on their restrictions to the decomposition groups, or on the determinant, conditions that must be met of course by $\overline{\rho}$). The abstract way to deal with it is this[5].

A class $\mathcal{D}^\square$ of lifts of $\overline{\rho}$ to pairs (A, ρ) where $A \in \mathcal{C}_\mathcal{O}$ is called a *deformation problem* if the following conditions hold:

- $(k, \overline{\rho}) \in \mathcal{D}^\square$.
- If $(A, \rho) \in \mathcal{D}^\square$ and $\phi : A \to B$ is a morphism in $\mathcal{C}_\mathcal{O}$ then $(B, \phi \circ \rho) \in \mathcal{D}^\square$.
- If $A \to C$ and $B \to C$ are morphisms in $\mathcal{C}_\mathcal{O}$ and (A, ρ_A), $(B, \rho_B) \in \mathcal{D}^\square$ map to the same ρ_C then $(A \times_C B, \rho_A \times_{\rho_C} \rho_B) \in \mathcal{D}^\square$.
- $\mathcal{D}^\square$ is closed under inverse limits.
- $\mathcal{D}^\square$ is closed under strict equivalence.
- If $A \hookrightarrow B$ is an injection in $\mathcal{C}_\mathcal{O}$ and (A, ρ) is such that $(B, \rho) \in \mathcal{D}^\square$, then $(A, \rho) \in \mathcal{D}^\square$.

[5]There are several ways to introduce an abstract notion of a "restricted deformation problem". They need not be equivalent, but the deformation problems with which we shall eventually be working comply with any of them. Instead of the approach of [D-D-T] we follow Patrick Allen's lecture notes [All].

In particular, the second axiom implies that $\mathcal{D}^\square \subset D^\square_{\overline{\rho}}$ is a sub-functor. The non-framed deformation problem associated with $\mathcal{D}^\square$ is the functor of strict equivalence classes in $\mathcal{D}^\square$, and yields a subfunctor $\mathcal{D} \subset D_{\overline{\rho}}$. It is well-defined since $\mathcal{D}^\square$ is closed under strict equivalence.

Proposition 35. *(i) Any representable sub-functor of $D^\square_{\overline{\rho}}$ closed under strict equivalences, is a framed deformation problem.*

(iii) Conversely, any framed deformation problem is representable by a quotient $R^\square_{\mathcal{D}}$ of $R^\square_{\overline{\rho}}$.

(ii) If $End_{k[G]}(\overline{\rho}) = k$, then the non-framed deformation problem associated with $\mathcal{D}$ is also representable, by a quotient $R_{\mathcal{D}}$ of the universal deformation ring $R_{\overline{\rho}}$ of $\overline{\rho}$.

We omit the easy proof. It follows from the axioms that $\mathcal{D}^\square(k[\varepsilon])$ is a sub vector space of $D^\square_{\overline{\rho}}(k[\varepsilon]) = Z^1(G, Ad\overline{\rho})$, which we denote by $Z^1_{\mathcal{D}}(G, Ad\overline{\rho})$. Since $\mathcal{D}^\square$ is closed under strict equivalence, it contains all the coboundaries, so

$$\mathcal{D}(k[\varepsilon]) = H^1_{\mathcal{D}}(G, Ad\overline{\rho})$$

is its image in $D_{\overline{\rho}}(k[\varepsilon]) = H^1(G, Ad\overline{\rho})$. We call it the *tangent space of the deformation problem $\mathcal{D}$*.

Example 36. *The fixed determinant condition.* Assume that $\ell = char(k)$ does not divide d. Fix a character $\epsilon : G \to \mathcal{O}^\times$ such that $\det(\overline{\rho}) = \overline{\epsilon}$. Let $\mathcal{D}^\square$ be the collection of liftings (A, ρ) with determinant $\epsilon : G \to \mathcal{O}^\times \to A^\times$. Then $\mathcal{D}^\square$ and $\mathcal{D}$ are deformation problems in the above sense and the tangent space to $\mathcal{D}$ is

$$H^1_{\mathcal{D}}(G, Ad\overline{\rho}) = H^1(G, Ad^0\overline{\rho})$$

(recall that $Ad^0 \subset Ad$ is the subrepresentation of trace-0 matrices). This is easily checked using the identity

$$\det(I + X\varepsilon) = 1 + tr(X)\varepsilon,$$

that holds in $GL_d(k[\varepsilon])$.

We now list certain types of deformations that show up in connection with modularity. In all of them $d = 2$, so

$$\overline{\rho} : G \to GL_2(k).$$

2.4.2. *Ordinary deformations.* In this example $G = G_\ell = Gal(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)$ and $I = I_\ell$ is its inertia subgroup. Suppose that

$$\overline{\rho} = \begin{pmatrix} \overline{\chi}_1 & * \\ & \overline{\chi}_2 \end{pmatrix}$$

with $\overline{\chi}_1|_I \neq 1$ and $\overline{\chi}_2|_I = 1$. Note that if $\overline{\epsilon} = \det(\overline{\rho})$ then $\overline{\epsilon}|_I = \overline{\chi}_1|_I$. Fix $\epsilon : G \to \mathcal{O}^\times$ lifting $\overline{\epsilon}$. For $A \in \mathcal{C}_{\mathcal{O}}$, let $\mathcal{D}^\square(A)$ be the collection of all the lifts $\rho : G \to A$ which are strictly equivalent, in $GL_2(A)$, to

$$\begin{pmatrix} \chi_1 & * \\ & \chi_2 \end{pmatrix}$$

with $\chi_1|_I = \epsilon|_I$ and $\chi_2|_I = 1$. (Note that $\det(\rho)$ is fixed only on I, but is allowed to deform on G.) Then $\mathcal{D}$ is a deformation problem called an *ordinary deformation problem* and is denoted by $\mathcal{D}_{ord}$. The role of the two characters along the diagonal may be switched (by dualizing, or by twisting by ϵ^{-1}). We shall denote the tangent space $H^1_{\mathcal{D}_{ord}}(G, Ad\overline{\rho})$ also by $H^1_{ord}(G, Ad\overline{\rho})$.

Showing that $\mathcal{D}_{ord}$ is a deformation problem reduces, by Proposition 35(i), to showing that $\mathcal{D}^{\square}_{ord}$ is representable. It is easy to check that the functor $\mathcal{D}_{Bor}$ of all lifts of $\overline{\rho}$ of the prescribed type which are upper-triangular is representable (but not closed under strict equivalence). So is the functor $L : \mathcal{C}_{\mathcal{O}} \rightsquigarrow$ Sets sending A to

$$L(A) = \left\{ \begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix} \mid z \in \mathfrak{m}_A \right\}.$$

In fact, L is representable by $\mathcal{O}[[Z]]$. Finally the map

$$L \times \mathcal{D}_{Bor} \to \mathcal{D}^{\square}_{ord}, \quad (u, \rho) \mapsto u\rho u^{-1}$$

is bijective when evaluated at any $A \in \mathcal{C}_{\mathcal{O}}$ (an isomorphism of functors).

2.4.3. *Flat deformations.* Again let $G = G_\ell$ be the decomposition group of ℓ, and M a finite G_ℓ-module. We say that M is *flat* if there exists a finite flat group scheme $\mathscr{G}$ over $\mathbb{Z}_\ell$ such that M is the Galois module associated to the generic fiber of M.

Theorem 37 (Raynaud)**.** [Ray] *(i) (relying on the absolute index of ramification being smaller than $\ell - 1$) The "generic fiber" functor*

$$\{\text{finite flat gp schemes}/\mathbb{Z}_\ell\} \rightsquigarrow \{G_\ell\text{−modules}\}$$

is fully faithful, and the flat modules are just those in its essential image. This is false without $e < \ell - 1$:μ_2 and $\mathbb{Z}/2\mathbb{Z}$ are non-isomorphic finite flat group schemes over $\mathbb{Z}_2$, but have the same generic fiber. Same for μ_ℓ and $\mathbb{Z}/\ell\mathbb{Z}$ over $\mathbb{Z}_\ell[\zeta_\ell]$.

(ii) The class of flat G_ℓ-modules is closed under taking sub-objects, quotients and finite direct sums. This has two consequences: (a) The category of finite flat gp schemes over $\mathbb{Z}_\ell$, or, equivalently, of flat G_ℓ-modules, is abelian (if $e \geq \ell - 1$ or over an arbitrary base, this is false; it is only an exact category, in general). (b)We may define, unambiguously, a profinite (continuous) G_ℓ-module to be flat if and only if every finite quotient of it is flat, equivalently if and only if it is an inverse limit of finite flat modules.

(iii) If M and M' are isomorphic as I_ℓ-modules, then M is flat if and only if M' is flat.

(iv) If M is a free *$\mathbb{Z}_\ell$-module of finite type, which is also a G_ℓ-module, then M is flat if and only if it is isomorphic to the Tate module of an ℓ-divisible group over $\mathbb{Z}_\ell$.*

Assume that $\overline{\rho} : G_\ell \to GL_2(k)$ is flat. For any $A \in \mathcal{C}_{\mathcal{O}}$ we let $\mathcal{D}^{\square}_{flat}(A) \subset D^{\square}_{\overline{\rho}}(A)$ be the liftings of $\overline{\rho}$ for which the profinite G_ℓ-module A^2 is flat. It turns out that this is a "deformation problem", and we denote as usual by $\mathcal{D}_{flat}$ the associated non-framed deformation functor, and by $H^1_{fl}(G_\ell, Ad\overline{\rho})$ its tangent space.

If $\overline{\rho}$ is flat, its shape can be made explicit (in the non-ordinary case, by means of Serre's fundamental character of level 2, see below). However, unlike the previous examples, it is hard to tell from the shape of a *deformation* ρ if it is flat or not. We shall have to study flat deformations using tools from integral p-adic Hodge theory, namely Fontaine-Laffaille modules.

2.4.4. *Minimally ramified deformations.* Take $G = G_p$ to be a decomposition group at a rational prime $p \neq \ell$, and $I = I_p$ its inertia subgroup. We give two examples of deformation problems that will be called *minimal.* In both $\overline{\rho}$ will be ramified, but the ramification in ρ will be as small as possible, given what is forced on it by $\overline{\rho}$.

(i) Type A: Suppose

$$1 \neq \overline{\rho}|_I \subset N(k) = \left\{ \begin{pmatrix} 1 & * \\ & 1 \end{pmatrix} \right\}.$$

Let $\mathcal{D}^{\square}_{min}$ be the class of liftings (A, ρ) which are strictly equivalent to a representation with $\rho|_I \subset N(A)$.

(ii) Type B: Suppose

$$\overline{\rho} = \begin{pmatrix} \overline{\chi}_1 & \\ & \overline{\chi}_2 \end{pmatrix}$$

with $\overline{\chi}_2|_I = 1$ and $\overline{\chi}_1|_I \neq 1$. Let χ_1 be the Teichmüller lift of $\overline{\chi}_1$. We let $\mathcal{D}^{\square}_{min}$ be the class of (A, ρ) which are strictly equivalent to representations of the same diagonal shape, with $\chi_1|_I$ and $\chi_2|_I = 1$ along the diagonal of $\rho|_I$.

More generally, if we assume that $\overline{\rho}(I)$ has order prime to ℓ, we may consider a deformation problem $\mathcal{D}_{min}$ by stipulating that $\rho(I) \to \overline{\rho}(I)$ is an isomorphism. [This is more general because it applies also to the case when $\overline{\rho}(I)$ is a non-split Cartan subgroup of $GL_2(k)$.]

In the two examples above, as well as in $\mathcal{D}_{ord}$ and $\mathcal{D}_{flat}$, we may impose also the condition that the determinant (on all of G, not only on I) is fixed.

2.4.5. *A variant: Λ-deformations (with or without conditions).* Let $\Lambda \in \mathcal{C}_{\mathcal{O}}$ and let $\mathcal{C}_\Lambda$ be the category $\mathcal{C}_{\mathcal{O}/\Lambda}$. We can define framed and non-framed deformation problems as we did when $\Lambda = \mathcal{O}$. One advantage is that now we may fix the determinant to be a character $\epsilon : G \to \Lambda^\times$. For example, we may take $\Gamma = G^{ab(\ell)}$ be the pro-ℓ completion of the abelianization of G, $\Lambda = \mathcal{O}[[\Gamma]]$ and

$$\epsilon(\sigma) = \epsilon_{cyc}(\sigma) \cdot [\sigma]$$

where ϵ_{cyc} is the cyclotomic character and $[\sigma]$ the projection of σ to $\Gamma \subset \Lambda^\times$. The universal deformation rings will now become Λ-algebras. The same can be done "with conditions" as above.

3. The universal deformation ring R_Σ

3.1. **The residual representation.**

3.1.1. *Running assumptions.* Let $\ell > 2$ be an odd prime, E a finite extension of $\mathbb{Q}_\ell$, $\mathcal{O}$ its ring of integers, λ its maximal ideal, and $k = \mathcal{O}/\lambda$.

Let $\overline{\rho} : G_{\mathbb{Q}} \to GL_2(k)$ be a continuous representation satisfying:

- $\overline{\rho}$ is odd an irreducible (Exercise: it is then absolutely irreducible).
- $\det \overline{\rho} = \overline{\epsilon}$ is the $\mod$ -ℓ cyclotomic character

$$\overline{\epsilon} : G_{\mathbb{Q}} \twoheadrightarrow Gal(\mathbb{Q}(\zeta_\ell)/\mathbb{Q}) \simeq \mathbb{F}_\ell^\times \subset k^\times.$$

- The restriction of $\overline{\rho}$ to a decomposition group G_ℓ is flat (2.4.3) or ordinary (2.4.2). (It could well be *both* flat *and* ordinary.)

- If $p \neq \ell$ and $\overline{\rho}$ is ramified at p, then it is of *type* A, i.e.

$$\{1\} \neq \overline{\rho}|_{I_p} \sim \begin{pmatrix} 1 & * \\ & 1 \end{pmatrix}$$

(Exercise: (i) $\overline{\rho}|_{G_p}$ is then also upper-triangular, with unramified characters along the diagonal, (ii) $\#\overline{\rho}(I_p) = \ell$.)

Example 38. (i) By Section 1.1, if A is a semistable elliptic curve over $\mathbb{Q}$, and $\overline{\rho}_{A,\ell}$ is irreducible, then $\overline{\rho}_{A,\ell}$ is such a $\overline{\rho}$. If A has good reduction at ℓ, then $\overline{\rho}_{A,\ell}|_{G_\ell}$ is flat (and also ordinary if and only if the reduction is ordinary). If A has multiplicative reduction at ℓ, then $\overline{\rho}_{A,\ell}|_{G_\ell}$ is ordinary (and also flat if and only if $ord_\ell(q_A) \equiv 0 \mod \ell$). By a theorem of Mazur, irreducibility holds if $\ell > 7$.

(ii) If $f \in S_2(\Gamma_0(N), \mathbb{C})$ is a *newform* of weight 2, *square-free* level $N = N_f$ and trivial nebentypus, and if λ is a prime of $\mathbb{Q}(a_n(f))$ above ℓ, then $\overline{\rho}_{f,\lambda}$ is such a $\overline{\rho}$, provided again it is irreducible. First, it is classical and easy that $\overline{\rho}_{f,\lambda}$ is odd, unramified outside the primes dividing N and ℓ, and that its determinant is $\overline{\epsilon}$. In fact, this holds already for $\rho_{f,\lambda}$ and follows from its construction via the abelian variety A_f associated to f by Shimura.

That the restriction of $\overline{\rho}_{f,\lambda}$ to the decomposition groups at the primes dividing N and ℓ is of the prescribed shape follows from the work of several people. In the results quoted below we do not have to assume that N is square-free.

(a) If $p \neq \ell$ is such that $p||N$, Carayol proved, building on work of Langlands, that

$$\rho_{f,\lambda}|_{G_p} \sim \begin{pmatrix} \eta^{-1}\epsilon & * \\ 0 & \eta \end{pmatrix}$$

where η is a quadratic unramified character, and $\eta(\sigma_p) = a_p(f) = \pm 1$. The point is that the local factor π_p of the automorphic representation π associated to f is "special". If N is square-free this holds for all $p|N$, and a-fortiori $\overline{\rho}_{f,\lambda}$ is "type A."

(b) If $\ell \nmid N$ then it is easy to see from the construction of $\overline{\rho}_{f,\lambda}$ that $\overline{\rho}_{f,\lambda}|_{G_\ell}$ is *flat*. If, moreover, $a_\ell(f)$ is a λ-adic unit, then $\rho_{f,\lambda}$ (and a-fortiori $\overline{\rho}_{f,\lambda}$) is *also ordinary* and

$$\rho_{f,\lambda}|_{G_\ell} \sim \begin{pmatrix} \chi^{-1}\epsilon & * \\ 0 & \chi \end{pmatrix}$$

where χ is unramified, and $\chi(\sigma_\ell)$ is the *unit root* (in $E = \mathbb{Q}(a_n(f))_\lambda$) of

$$X^2 - a_\ell(f)X + \ell = 0.$$

(c) Finally, if $\ell||N$ then $\rho_{f,\lambda}$ is ordinary and

$$\rho_{f,\lambda}|_{G_\ell} \sim \begin{pmatrix} \eta^{-1}\epsilon & * \\ 0 & \eta \end{pmatrix}$$

where η is a quadratic unramified character, and $\eta(\sigma_\ell) = a_\ell(f) = \pm 1$. This follows from work of Deligne and Rapoport. For a proof see [Gr], Proposition 12.1.

Definition. We say that $\overline{\rho}$ (or a deformation ρ) is *semistable* at ℓ if its restriction to G_ℓ is ordinary or flat.

3.1.2. *The restriction of* $\overline{\rho}$ *to* $\mathbb{Q}(\sqrt{(-1)^{(\ell-1)/2}\ell})$. We shall need another technical condition on $\overline{\rho}$. Let $L = \mathbb{Q}(\sqrt{(-1)^{(\ell-1)/2}\ell})$. This is the unique quadratic subfield of $\mathbb{Q}(\zeta_\ell)$. We impose the following condition:

- (L) The restriction $\overline{\rho}|_{G_L}$ is absolutely irreducible. (Note that since L is imaginary, oddness makes no sense over L, so irreducibility no longer implies absolute irreducibility.)

Fortunately for us, assumption (L) follows from the other assumptions made on $\overline{\rho}$, provided we know that $\overline{\rho}$ is modular.

Proposition 39. *Suppose $\overline{\rho}$ satisfies the running assumptions, and in addition is modular. Then (L) holds.*

Proof. Suppose (after possibly enlarging k) $\overline{\rho}|_{G_L}$ were reducible. If $\ell | \#\overline{\rho}(G_L)$ then $\overline{\rho}|_{G_L}$is not diagonalizable (even over the algebraic closure of k), so must have a unique invariant line, on which G_L acts via a character. Since $G_L \lhd G_{\mathbb{Q}}$, this line must be $G_{\mathbb{Q}}$-stable too, contradicting the irreducibility of $\overline{\rho}$. It follows that $\ell \nmid \#\overline{\rho}(G_L)$, and since $[L : \mathbb{Q}] = 2$, $\ell \nmid \#\overline{\rho}(G_{\mathbb{Q}})$. By our running assumptions on $\overline{\rho}|_{G_p}$, $p \neq \ell$, if $\overline{\rho}$ were ramified at p,we would have $\ell | \#\overline{\rho}(G_p)$. Thus, $\overline{\rho}$ is unramified outside ℓ. The prime-to-ℓ conductor $N(\overline{\rho})$ of $\overline{\rho}$ is therefore 1. Moreover, by the same argument

$$\overline{\rho}|_{I_\ell} \sim \begin{pmatrix} \overline{\epsilon} & \\ & 1 \end{pmatrix}$$

if $\overline{\rho}$ is ordinary at ℓ, so must be flat at ℓ, even if it is ordinary there. It now follows from Diamond's strengthening of Ribet's theorem on lowering the level ([Di93], Theorem 1.1) that $\overline{\rho}$ must be modular of weight 2 and level 1. But there are no weight 2 cusp forms of level 1, a contradiction. □

3.1.3. *Vanishing of $H^0(G_{\mathbb{Q}}, Ad^0\overline{\rho}^*)$.* Let $W = Ad^0\overline{\rho}$. The invariant pairing $Tr(XY)$ makes W a self-dual representation, so

$$W^* = Hom(W, \mu_\ell) \simeq W \otimes \mu_\ell = W(1).$$

Lemma 40. *We have $H^0(G_{\mathbb{Q}}, Ad^0\overline{\rho}^*) = 0$.*

Proof. Let V be the underlying space of $\overline{\rho}$. Since $\bigwedge V \simeq \mu_\ell$, $V^\vee \simeq V(1)$, and

$$Ad\overline{\rho}(1) = V \otimes V^\vee(1) \simeq V^\vee \otimes V^\vee.$$

Under this isomorphism $W(1) = Ad^0\overline{\rho}(1) \simeq Sym^2 V^\vee$. We therefore have to prove that there does not exist a non-zero symmetric $G_{\mathbb{Q}}$-invariant bilinear form on V. Suppose

$$0 \neq \beta(u, v) = {}^t uBv$$

is such a bilinear form. If it were degenerate, its kernel would be an invariant subspace of V, contradicting the irreducibility of $\overline{\rho}$. Thus $\det B \neq 0$. For $\sigma \in G_{\mathbb{Q}}$

$$\beta(u, v) = \beta(\sigma u, \sigma v) = {}^t u\, {}^t\overline{\rho}(\sigma) B \overline{\rho}(\sigma) v,$$

so ${}^t\overline{\rho}(\sigma)B\overline{\rho}(\sigma) = B$, and in particular $\overline{\epsilon}(\sigma)^2 = \det(\overline{\rho}(\sigma))^2 = 1$. If $\ell > 3$, this is a contradiction. If $\ell = 3$, we find that the image of $\overline{\rho}|_{G_L}$ lies in the group $SO(2)$ (we may assume that we are over the algebraic closure of k). But $SO(2)$ is diagonalizable (over the algebraic closure of k), contradicting the irreducibility of $\overline{\rho}|_{G_L}$. □

3.2. **Global deformations of type Σ (Week 6).** Let Σ be a finite set of *finite* primes, which may be empty. We define a global deformation problem $\mathcal{D}_\Sigma$ by stipulating that $\rho \in \mathcal{D}_\Sigma^\square(A)$ if and only if the following conditions hold:

- $\det(\rho) = \epsilon : G_\mathbb{Q} \twoheadrightarrow \mathbb{Z}_\ell^\times \subset \mathcal{O}^\times \to A^\times$.
- $\rho|_{G_\ell}$ is semistable: either *flat* (i.e. for any Artinian quotient of A the image of $\rho|_{G_\ell}$ is the Galois module associated with the generic fiber of a finite flat group scheme over $\mathbb{Z}_\ell$) or *ordinary*, i.e.
$$\rho|_{G_\ell} \sim \begin{pmatrix} \chi^{-1}\epsilon & * \\ 0 & \chi \end{pmatrix}$$
where $\chi : G_\ell \to A^\times$ is unramified.
- If $p \neq \ell$, $p \notin \Sigma$ and $\overline{\rho}|_{G_p}$ is unramified, then $\rho|_{G_p}$ is unramified as well.
- If $p \neq \ell$, $p \notin \Sigma$ and $\overline{\rho}|_{G_p}$ is ramified, then $\rho|_{G_p}$ is "type A", i.e.
$$\rho|_{G_p} \sim \begin{pmatrix} \eta^{-1}\epsilon & * \\ 0 & \eta \end{pmatrix}$$
with η unramified (and, necessarily, $*|_{I_p} \neq 0$).
- If $\ell \notin \Sigma$ and $\overline{\rho}|_{G_\ell}$ is flat, then $\rho|_{G_\ell}$ is flat. *We only include ℓ in Σ if $\overline{\rho}$ is ordinary and flat, but we want to consider deformations that might be ordinary and not flat.* In the other two cases, either $\overline{\rho}$ is non-ordinary, in which case it is flat and any deformation must be flat, by the second condition above, or it is non-flat, in which case it is ordinary and any deformation must be ordinary.

Thus, if p or ℓ are not in Σ, the local deformation is "minimally ramified" in the sense that it is of the same type as $\overline{\rho}$. At primes in Σ we do not impose any condition, except that at ℓ we retain the assumption that $\rho|_{G_\ell}$ is either flat or ordinary, and we always keep the condition on the determinant.

If $\Sigma \subset \Sigma'$, then clearly $\mathcal{D}_\Sigma \subset \mathcal{D}_{\Sigma'}$. If $\Sigma = \emptyset$, we say that $\mathcal{D}_\Sigma^\square$ is a *minimal global deformation problem.*

Example 41. (i) If $\overline{\rho} = \overline{\rho}_{A,\ell}$ for a semistable elliptic curve $A/\mathbb{Q}$, then $\rho_{A,\ell}$ is of type Σ if Σ contains all the places of bad reduction of A (but it can be of type Σ for a smaller set Σ).

(ii) If f is a weight 2, level N newform with trivial nebentypus, and λ a prime above ℓ in $\mathbb{Q}(a_n(f))$, and if N is square-free, then $\rho_{f,\lambda}$ is such a deformation of $\overline{\rho} = \overline{\rho}_{f,\lambda}$, with Σ the set of primes dividing N. Note that if $p \neq \ell$ (resp. ℓ) divides N then $\rho_{f,\lambda}$ would be ramified (resp. non-flat but ordinary) there, although $\overline{\rho}_{f,\lambda}$ might be non-ramified (resp. flat), so the deformation need not be minimal.

Proposition 42. *(i) $\mathcal{D}_\Sigma^\square$ is a framed "deformation problem".*

(ii) Let S be the set of prime consisting of ∞, ℓ, the primes where $\overline{\rho}$ is ramified, and the primes in Σ. Let $G = G_S$. The associated non-framed deformation problem is represented by a quotient ring R_Σ of $R_{\overline{\rho},S}^{univ}$, the universal deformation ring of $\overline{\rho} : G_S \to GL_2(k)$.

Proof. (i) Each of the local conditions is a "deformation problem". These are precisely the examples discussed before. The axioms defining a "deformation problem" are compatible with localization, so the global $\mathcal{D}_\Sigma^\square$ is, technically speaking, also a "deformation problem".

(ii) Let $R_{\overline{\rho}}$ be the (global) universal deformation ring of $\overline{\rho} : G_S \to GL_2(k)$ (with determinant ϵ) and $R^{\square}_{\overline{\rho}}$ the corresponding universal *framed* deformation ring (both without the local conditions). As explained before, a choice of ρ in the strict equivalence class $\rho^{univ} \in D_{\overline{\rho}}(R_{\overline{\rho}})$ determines a homomorphism $R^{\square}_{\overline{\rho}} \to R_{\overline{\rho}}$, bringing the universal framed deformation to ρ.

For $v \in S$, let $R^{\square}_{\overline{\rho},v}$ be the corresponding universal framed deformation rings for $\overline{\rho}|_{G_v}$. Since the restriction of the (global) universal framed deformation to the decomposition group is a "local framed deformation", the universal property of $R^{\square}_{\overline{\rho},v}$ yields a homomorphism $R^{\square}_{\overline{\rho},v} \to R^{\square}_{\overline{\rho}}$. Let

$$R^{\square}_{loc} = \prod_{v \in S,\, R^{\square}_{\overline{\rho}}} R^{\square}_{\overline{\rho},v}$$

(the fiber product of the local framed deformation rings over the global one). We obtain a homomorphism

$$R^{\square}_{loc} \to R^{\square}_{\overline{\rho}} \to R_{\overline{\rho}},$$

(the second arrow depending on the choice of ρ).

When we introduce the local conditions, we have a surjective homomorphism $R^{\square}_{\overline{\rho},v} \twoheadrightarrow R^{\square}_{\mathcal{D},v}$ for each $v \in S$, expressing the (local) universal framed deformation ring with condition $\mathcal{D}_v$ as a quotient of the corresponding ring without any conditions. Similarly, there is a surjective homomorphism $R^{\square}_{\overline{\rho}} \twoheadrightarrow R^{\square}_{\mathcal{D}}$ between the global framed deformation rings, with and without conditions. Put together, these give a homomorphism

$$R^{\square}_{loc} \twoheadrightarrow R^{\square}_{\mathcal{D},loc} = \prod_{v \in S,\, R^{\square}_{\mathcal{D}}} R^{\square}_{\mathcal{D},v}.$$

It is now straightforward to check that

$$R_{\Sigma} := R^{\square}_{\mathcal{D},loc} \otimes_{R^{\square}_{loc}} R_{\overline{\rho}}$$

is a universal (non-framed) deformation ring "with conditions" $\mathcal{D}_{\Sigma}$. Indeed, to give a homomorphism $R_{\Sigma} \to A$ is to give a homomorphism $R_{\overline{\rho}} \to A$, i.e. a specialization of the strict equivalence class ρ^{univ}, such that if we specialize the chosen representative ρ and get, say, a framed deformation ρ_A, its restriction to every G_v (determined by the map $R^{\square}_{\overline{\rho},v} \to R^{\square}_{\overline{\rho}} \to R_{\overline{\rho}} \to A$) satisfies condition $\mathcal{D}_v$ (i.e. factors through a map $R^{\square}_{\mathcal{D},v} \to A$).

We remark that the need to work both with framed and non-framed deformation rings resulted form the fact that locally, $\overline{\rho}|_{G_v}$ need not be irreducible, so need not have a universal non-framed deformation ring. When we quotient out the local deformation rings by the ideals defining the conditions in $\mathcal{D}_v$, we have to do it with *framed* deformation rings. Globally, however, we wanted to get the universal *non-framed* deformation ring R_{Σ}. □

3.3. Tangent spaces of type Σ and the Greenberg-Wiles formula.

3.3.1. *The global tangent space.* Let S be a set of primes *containing* ∞, ℓ, the primes where $\overline{\rho}$ is ramified, and the primes in Σ. Let

$$\boldsymbol{t}_{\overline{\rho}} := D_{\overline{\rho}}(k[\varepsilon]) \simeq H^1(G_S, Ad^0\overline{\rho})$$

be the tangent space of the deformation problem with the only conditions being (i) $\det = \epsilon$, (ii) unramified outside S (i.e. factoring through G_S).

At each *finite* $v \in S$ let

$$L_v \subset H^1(G_v, Ad^0\overline{\rho})$$

be the subspace $\mathcal{D}_v(k[\varepsilon])$ where $\mathcal{D}_v$ is the local condition, as defined above. Note that if v is a place different from ℓ, where $\overline{\rho}$ is unramified, and not in Σ, then $L_v = H^1(G_v/I_v, Ad^0\overline{\rho})$.

We let $\mathcal{L}_\Sigma = \{L_v | \, v \in S_f\}$. Although the notation stresses the role of Σ, this collection depends also on the choice of S. We shall make these subspaces explicit soon, but at the moment we treat them as a black box.

Proposition 43. *The tangent space* $\boldsymbol{t}_\Sigma = \mathcal{D}_\Sigma(k[\varepsilon]) \subset D_{\overline{\rho}}(k[\varepsilon]) = \boldsymbol{t}_{\overline{\rho}}$ *"of type* Σ*" is identified with the generalized Selmer group*

$$H^1_{\mathcal{L}_\Sigma}(G_S, Ad^0\overline{\rho}) := \ker\left(loc : H^1(G_S, Ad^0\overline{\rho}) \to \prod_{v\in S} H^1(G_v, Ad^0\overline{\rho})/L_v\right).$$

Proof. Clear. Note that in the product it is harmless to include $v = \infty$, as ℓ is odd, so $H^1(G_\mathbb{R}, Ad^0\overline{\rho}) = 0$. If $w \notin S$ we may replace S by $S \cup \{w\}$ without changing the Selmer group, because $L_w = H^1(G_w/I_w, Ad^0\overline{\rho})$ means that a cohomology class in $H^1_{\mathcal{L}_\Sigma}(G_{S\cup\{w\}}, -)$ is unramified at w, so belongs to $H^1_{\mathcal{L}_\Sigma}(G_S, -)$. However, if at the same time we replace Σ by $\Sigma \cup \{w\}$, the Selmer group grows, as the constraint of being unramified at w is dropped. □

3.3.2. *The dual Selmer group.* Let $L_v^\perp$ be the annihilator of L_v under the perfect pairing of abelian groups (local Tate duality)

$$H^1(G_v, Ad^0\overline{\rho}) \times H^1(G_v, Ad^0\overline{\rho}(1)) \to \mathbb{Q}/\mathbb{Z}.$$

Since the pairing

$$\langle , \rangle : Ad^0\overline{\rho} \times Ad^0\overline{\rho}(1) \simeq Ad^0\overline{\rho} \times Ad^0\overline{\rho}^* \to \mu$$

underlying it is a pairing of k-vector spaces (i.e. $\langle \alpha x, y\rangle = \langle x, \alpha y\rangle$ for $\alpha \in k$), and L_v is a k-vector space, so is $L_v^\perp$. We let $\mathcal{L}_\Sigma^* = \{L_v^\perp | \, v \in S_f\}$.

We emphasize that the collection $\mathcal{L}_\Sigma^*$ need *not* be associated with any deformation type. It is called the system of local conditions *dual* to $\mathcal{L}_\Sigma$. The *dual* Selmer group is

$$H^1_{\mathcal{L}_\Sigma^*}(G_S, Ad^0\overline{\rho}(1)) := \ker\left(loc : H^1(G_S, Ad^0\overline{\rho}(1)) \to \prod_{v\in S} H^1(G_v, Ad^0\overline{\rho}(1))/L_v^\perp\right).$$

3.3.3. *The Greenberg-Wiles formula.* For the moment, let us be more general. Let M be a finite $G_\mathbb{Q}$-module (such as $Ad^0\overline{\rho}$) and

$$\mathcal{L} = \{L_v\}$$

a family of subgroups $L_v \subset H^1(G_v, M)$ (v runs over all the places of $\mathbb{Q}$, including ∞) such that for all but finitely many v, $L_v = H^1(G_v/I_v, M^{I_v})$. The *generalized Selmer group* is

$$H^1_\mathcal{L}(G_\mathbb{Q}, M) = \left\{x \in H^1(G_\mathbb{Q}, M) | \, \forall v \, res_v(x) \in L_v\right\}.$$

The dual set of local conditions $\mathcal{L}^*$ is defined by letting $L_v^* = L_v^\perp \subset H^1(G_v, M^*)$ under the duality between $H^1(G_v, M)$ and $H^1(G_v, M^*)$ (local Tate duality). Note that if v is finite, $L_v = H^1(G_v/I_v, M^{I_v})$ and v does not divide $\#M$, also $L_v^* = H^1(G_v/I_v, M^{*I_v})$.

The main result of Wiles concerning these Selmer groups is the following theorem, inspired by earlier work of Ralph Greenberg.

Theorem 44 (Greenberg-Wiles formula). *Both $H^1_{\mathcal{L}}(G_{\mathbb{Q}}, M)$ and $H^1_{\mathcal{L}^*}(G_{\mathbb{Q}}, M^*)$ are finite and*

$$\frac{\#H^1_{\mathcal{L}}(G_{\mathbb{Q}}, M)}{\#H^1_{\mathcal{L}^*}(G_{\mathbb{Q}}, M^*)} = \frac{\#H^0(G_{\mathbb{Q}}, M)}{\#H^0(G_{\mathbb{Q}}, M^*)} \cdot \prod_v \frac{\#L_v}{\#H^0(G_v, M)}. \tag{3.1}$$

Since we have seen in Corollary 12 that for a finite place v

$$\#H^0(G_v, M) = \#H^1(G_v/I_v, M^{I_v}),$$

all but finitely many terms in the infinite product are 1. We emphasize that the product ranges over all the v, including $v = \infty$.

Given the set of local conditions $\mathcal{L}$ we take S to be any finite set of places containing ∞, the primes dividing $\#M$, and the places where $L_v \neq H^1(G_v/I_v, M^{I_v})$. The same S will work then for $\mathcal{L}^*$.

Anticipating the application to the proof of the Modularity Theorem we remark that, enlarging the set S by a carefully selected set of auxilary primes q, and taking the least restrictive $L_q = H^1(G_q, M)$ for the new q's, Wiles manages to guarantee that $H^1_{\mathcal{L}^*}(G_{\mathbb{Q}}, M^*) = 0$. His formula gives him then a precise control over $\#H^1_{\mathcal{L}}(G_{\mathbb{Q}}, M)$.

Corollary 45. *Suppose $\mathcal{L}'$ is obtained from $\mathcal{L}$ by replacing $H^1(G_q/I_q, M^{I_q})$ by $H^1(G_q, M)$ for some prime $q \nmid \#M$. Then*

$$\frac{\#H^1_{\mathcal{L}'}(G_{\mathbb{Q}}, M)}{\#H^1_{\mathcal{L}}(G_{\mathbb{Q}}, M)} = \frac{\#H^1_{\mathcal{L}'^*}(G_{\mathbb{Q}}, M^*)}{\#H^1_{\mathcal{L}^*}(G_{\mathbb{Q}}, M^*)} \cdot \#H^0(G_q, M^*) \leq \#H^0(G_q, M^*),$$

with equality if $H^1_{\mathcal{L}^*}(G_{\mathbb{Q}}, M^*) = 0$ already.

Proof. (of Corollary) When we change $\mathcal{L}$ to $\mathcal{L}'$, the RHS of the expression in the theorem changes by

$$\frac{\#H^1(G_q, M)}{\#H^1(G_q/I_q, M^{I_q})} = \frac{\#H^1(G_q, M)}{\#H^0(G_q, M)} = \#H^2(G_q, M) = \#H^0(G_q, M^*)$$

by the local Euler characteristic formula (and the fact that $q \nmid \#M$) and by Tate's local duality.

Proof (of Theorem): We shall show how to derive the theorem from the Poitou-Tate 9-term exact sequence. We have already noted the finiteness of $H^1(G_S, M)$ in Lemma 15. The group $H^1_{\mathcal{L}}(G_{\mathbb{Q}}, M)$ is a subgroup of it, hence clearly finite, and similarly $H^1_{\mathcal{L}^*}(G_{\mathbb{Q}}, M^*)$ is finite.

By definition, we have an exact sequence of finite abelian groups

$$0 \to H^1_{\mathcal{L}^*}(G_{\mathbb{Q}}, M^*) \to H^1(G_S, M^*) \to \prod_{v \in S} \frac{H^1(G_v, M^*)}{L_v^{\perp}}.$$

Dualizing, we get an exact sequence

$$0 \leftarrow H^1_{\mathcal{L}^*}(G_{\mathbb{Q}}, M^*)^{\vee} \leftarrow H^1(G_S, M^*)^{\vee} \leftarrow \prod_{v \in S} L_v.$$

Splicing it into the 9-term exact sequence we get

$$0 \to H^0(G_S, M) \xrightarrow{\alpha_0} \prod_{v \in S} \widehat{H}^0(G_v, M) \xrightarrow{\beta_0} H^2(G_S, M^*)^{\vee} \to$$

$$\to H^1_{\mathcal{L}}(G_{\mathbb{Q}}, M) \xrightarrow{\alpha_1} \prod_{v\in S} L_v \xrightarrow{\beta_1} H^1(G_S, M^*)^\vee \to H^1_{\mathcal{L}^*}(G_{\mathbb{Q}}, M^*)^\vee \to 0.$$

The theorem follows from this, from the global Euler characteristic formula, and from the fact that

$$\#M = \#M^* = \#(1+c)M \cdot \#H^0(G_{\mathbb{R}}, M^*)$$

which we leave as an easy exercise. (In the applications our M will have an odd order, in which case the last equality boils down to the obvious $\#M = \#(M^+) \cdot \#(M^-)$.) □

3.4. **Computation of local terms at $p \neq \ell$.** We want to calculate the local terms appearing on the RHS of (3.1). That the local terms are computable, is in principle not surprising. After all, local cohomologies are easier than the global ones, and Tate's duality and the local Euler characteristic often reduce their computation to that of H^0's. *That they come out to be what they are, and eventually lead to surprisingly pleasant results for the orders of the global Selmer groups, is a "numerical coincidence", or sheer good luck.* In fact, in the generalizations to modularity theorems for higher fields, in the work of Calegari and Geraghty, this is not the case any more, and the same Galois cohomology computations had to be radically upgraded.

Let $W = Ad^0\overline{\rho}$.

- If $v = \infty$ then $L_\infty = 0$ (because $\ell > 2$) and $\dim H^0(G_\infty, W) = 1$.
- **Claim**: If $p \neq \ell$ is finite and $p \notin \Sigma$ then $\#L_p = \#H^0(G_p, W)$ because $L_p = H^1(G_p/I_p, W^{I_p})$.

This is clear if $\overline{\rho}$ is unramified at p. Let us show that the same formula remains valid if $\overline{\rho}$ is ramified, in which case it is "type A". Informally, saying that "a deformation $\rho : G_p \to GL_2(k[\varepsilon])$ is as little ramified as is forced upon it by $\overline{\rho}$" means that while $\overline{\rho}$ is ramified, the *extension class* (of k by W) defining ρ is unramified (i.e. is a push-out of an unramified extension of k by W^{I_p}). Specifically, let V be the underlying space of $\overline{\rho}$ and $V_1 \subset V$ the I_p-invariant line. In a basis consisting of a vector from V_1 and a vector projecting non-trivially to V/V_1,

$$\overline{\rho}(\sigma) = \begin{pmatrix} \overline{\epsilon\eta}^{-1}(\sigma) & \beta(\sigma) \\ 0 & \overline{\eta}(\sigma) \end{pmatrix}$$

with $\overline{\epsilon}$ and $\overline{\eta}$ unramified. Then

$$W_1 = \{w \in W |\, w(V_1) = 0,\, w(V) \subset V_1\} = W^{I_p}$$

is 1-dimensional and is equal to the I_p-invariants of W. In the above basis,

$$W_1 = \left\{ \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \right\} \subset \left\{ \begin{pmatrix} a & b \\ c & -a \end{pmatrix} \right\} = W.$$

Since any "type A" deformation to $k[\varepsilon]$ is represented by a 1-cocycle in W whose restriction to I_p has values in W_1 (and vice versa) we get that $L_p = H^1_{\mathcal{D}_p}(G_p, W)$ consists of the classes whose restriction to I_p is in the image of $H^1(I_p, W_1)$, namely

$$H^1_{\mathcal{D}_p}(G_p, W) = \ker(H^1(G_p, W) \xrightarrow{r} H^1(I_p, W/W_1)).$$

Here r is the map "restrict to I_p and project modulo W^{I_p}". The key point is that

$$H^1(I_p, W) \to H^1(I_p, W/W_1)$$

is injective, as follows from the long exact sequence of I_p-cohomology attached to

$$0 \to W_1 \to W \to W/W_1 \to 0.$$

(Here the fact that both $H^1(I_p, W_1) = Hom(I_p, W_1)$ and $H^0(I_p, W/W_1)$ are 1-dimensional plays a role.) Therefore $\ker(r)$ is the same as

$$\ker(H^1(G_p, W) \xrightarrow{r'} H^1(I_p, W)) = H^1(G_p/I_p, W^{I_p}),$$

as had to be proved.

- If $p \neq \ell$ is finite and $p \in \Sigma$ then $\mathcal{D}_p$ is non-restricted (except for the condition on the determinant), $L_p = H^1(G_p, W)$, so

$$\frac{\#L_p}{\#H^0(G_p, W)} = \frac{\#H^1(G_p, W)}{\#H^0(G_p, W)} = \#H^2(G_p, W) = \#H^0(G_p, W(1))$$

by the local Euler characteristic formula and local Tate duality.

3.5. **Computation of local terms at ℓ (Week 7).**

3.5.1. *Flat, ordinary and semistable representations.* Let us generalize a little and consider $\rho : G_\ell \to GL_2(R)$ where R is a finite local ring of cardinality a power of ℓ. We let M_ρ be the underlying module, free of rank 2 over R.

Recall that in general, a finite G_ℓ-module M is called *flat* if there exists a finite flat group scheme $\mathscr{G}$ over $Spec\mathbb{Z}_\ell$ such that $M \simeq \mathscr{G}(\overline{\mathbb{Q}}_\ell)$ as a G_ℓ-module. Equivalently, since finite flat group schemes in characteristic 0, being étale, are nothing else but finite Galois modules, M may be *identified* with the generic fiber $\mathscr{G}_\eta$ of $\mathscr{G}$ and the condition then becomes that it *extends* to a finite flat group scheme over $Spec\mathbb{Z}_\ell$. Thanks to the fact that $e(\mathbb{Q}_\ell) = 1 < \ell - 1$ Raynaud's theorem guarantees that $\mathscr{G}$ is unique up to a unique isomorphism. We also note that when this notion is applied to M_ρ, the $\mathcal{O}$-module structure on M_ρ is an extra structure, but because of the full-faithfullness in Raynaud's theorem, this $\mathcal{O}$-structure extends uniquely to $\mathscr{G}$.

Likewise, we say that M is *ordinary* if it admits a G_ℓ-stable filtration

$$0 \subset M_1 \subset M$$

such that I_ℓ (inertia) acts trivially on M/M_1, and via the cyclotomic character ϵ on M_1. If M is ordinary, so is $M^* = Hom(M, \mu)$. Note that the definition so far does not exclude the possibility that $M_1 = M$ or 0.

Exercise: (i) If M is ordinary, M_1 is uniquely determined as

$$M_1 = \{x \in M | \forall \tau \in I_\ell \ \rho(\tau)x = \epsilon(\tau)x\}.$$

(ii) It is enough to stipulate that the filtration is I_ℓ-stable, since it is then also G_ℓ-stable. (iii) If $M \simeq M'$ as I_ℓ-modules, then M is ordinary if and only if M' is. (iv) If M admits an $\mathcal{O}$-structure, M_1 is an $\mathcal{O}$-submodule.

We say that ρ is flat or ordinary if M_ρ is, and in addition $\det \rho|_{I_\ell} = \epsilon$. In the ordinary case this implies that M_1 is free of rank 1 over R. By what we have seen in the exercise above, and previously in the flat case, being flat or ordinary depends only on $\rho|_{I_\ell}$. These notions are also stable under the operations of sub-objects, quotients and finite direct sums.

We say that ρ is *semistable* if it is flat or ordinary. It may be then (a) flat and ordinary, (b) flat non-ordinary, or (c) ordinary non-flat. The ℓ-torsion of elliptic curves already give examples of all three possibilities.

We recall two classical constructions from the theory of local fields.

(a) **Fundamental characters.** Recall that the tame inertia group T_ℓ is a quotient of I_ℓ, and has for any $r \geq 1$ a unique cyclic quotient isomorphic to $\mathbb{F}_{\ell^r}^\times$. It is the quotient by the kernel of the *r-th fundamental character*

$$\varepsilon_r : I_\ell \twoheadrightarrow \mathbb{F}_{\ell^r}^\times, \quad \varepsilon_r(\sigma) = \sigma(\sqrt[\ell^r-1]{\ell})/\sqrt[\ell^r-1]{\ell}.$$

(When $r = 1$ this is the cyclotomic character.) The characters ε_r^i, $0 \leq i < \ell^r - 1$ are the $\ell^r - 1$ distinct characters of this quotient, with values in $\mathbb{F}_{\ell^r}^\times$. The Galois group of the unramified extension of $\mathbb{Q}_\ell$ acts on these characters via $\sigma(\chi)(\tau) = \chi(\widetilde{\sigma}\tau\widetilde{\sigma}^{-1})$, and the Frobenius σ_ℓ sends ε_r to ε_r^ℓ.

(b) **Kummer Theory.** Let R be a finite ring of cardinality a power of ℓ. Suppose $\rho : G_\ell \to GL_2(R)$ is ordinary and

$$0 \to R(1) \to M_\rho \to R \to 0$$

is the corresponding filtration of $\rho|_{I_\ell}$. This gives a class

$$c_\rho \in H^1(I_\ell, R(1)).$$

On the other hand (this is Kummer Theory), the exact sequence

$$0 \to \mu_{\ell^n} \to \overline{\mathbb{Q}}_l^\times \xrightarrow{\ell^n} \overline{\mathbb{Q}}_l^\times \to 0,$$

together with Hilbert's Theorem 90, gives an isomorphism $H^1(I_\ell, \mu_{\ell^n}) \simeq K^\times/K^{\times\ell^n}$ where $K = \mathbb{Q}_\ell^{nr}$. Following it by the valuation $ord_\ell : K^\times/K^{\times\ell^n} \to \mathbb{Z}/\ell^n\mathbb{Z}$ and taking $\ell^n = \#R$ we get a map

$$v : H^1(I_\ell, R(1)) \to R.$$

Parts (a)-(c) of the following Lemma follow from standard facts on finite flat group schemes over $\mathbb{Z}_\ell$. For example, (b) follows from the existence of the connected-étale exact sequence. Part (d) follows from the theory of Fontaine-Laffaille modules, explained below.

Lemma 46. *(a) (Shape of residual flat representations) Let $\overline{\rho} : G_\ell \to GL_2(k)$ be a flat representation over k such that $\det(\overline{\rho})|_{I_\ell} = \overline{\epsilon}$. Then either $\overline{\rho}$ is ordinary or*

$$\overline{\rho}|_{I_\ell} \otimes_k \overline{k} \simeq \varepsilon_2 \oplus \varepsilon_2^\ell$$

(isomorphism over the algebraic closure of k).

(b) (When is "flat" also "ordinary"?) If $R \in \mathcal{C}_{\mathcal{O}}$, $\rho : G_\ell \to GL_2(R)$ is flat and $\overline{\rho} = \rho \mod \mathfrak{m}_R$ is ordinary, then ρ is ordinary.

(c) (When is "ordinary" also "flat"?) If $\rho : G_\ell \to GL_2(R)$ is ordinary, then ρ is also flat if and only if $v(c_\rho) = 0$. (Note that the notion of flatness only refers to the structure of M_ρ as a $\mathbb{Z}_\ell[G_\ell]$-module, so we may assume that $R = \mathbb{Z}/\ell^n\mathbb{Z}$). Serre called in this case ρ "peu ramifié", and "très ramifié" otherwise.

(d) (When does M_ρ flat imply ρ flat?) Let $\rho : G_\ell \to GL_2(R)$ and suppose M_ρ is flat. Then either $\det(\rho)|_{I_\ell} = \epsilon$ or $\rho|_{I_\ell} = \epsilon$ or $\rho|_{I_\ell} = 1$. In particular if $\overline{\rho}$ is flat (i.e., satisfies also the condition on the determinant restricted to inertia), so is ρ.

3.5.2. *Infinitesimal deformations.* Let $R = \mathcal{O}/\lambda^n$, and suppose that

$$\rho : G_\ell \to GL_2(R)$$

is semistable (i.e. flat or ordinary and $\det \rho|_{I_\ell} = \epsilon$). Recall that $H^1(G_\ell, Ad\rho)$ classified the infinitesimal deformations in the category of profinite $\mathcal{O}/\lambda^n[G_\ell]$-modules, i.e. deformations to $\rho' : G_\ell \to GL_2(R[\varepsilon])$. We remark that there are other deformations, e.g. to $\rho' : G_\ell \to GL_2(\mathcal{O}/\lambda^{n+1})$ that are not classified by $H^1(G_\ell, Ad\rho)$. In particular, when $n = 1$, we are talking only about "equicharacteristic deformations". Recall also that $H^1(G_\ell, Ad^0\rho)$ classified only the deformations with fixed determinant.

Although we shall use the language of Galois cohomology, and not extensions, the reader may note that

$$H^1(G_\ell, Ad\rho) \simeq \mathrm{Ext}^1(M_\rho, M_\rho),$$

the extensions taken in the category of $\mathcal{O}/\lambda^n[G_\ell]$-modules. This can be seen either by using the interpretation of both H^1 and Ext^1 as appropriate derived functors, or directly, by associating to a free rank-2 $\mathcal{O}/\lambda^n[\varepsilon]$-module $\widetilde{M}$ with a G_ℓ-action the extension

$$0 \to M \to \widetilde{M} \to M \to 0$$

with $M = \varepsilon\widetilde{M} \simeq \widetilde{M}/\varepsilon\widetilde{M}$.

We now define subspaces of $H^1(G_\ell, Ad\rho)$ that will turn out to be the (reduced) tangent spaces to the local deformation problems "with local conditions" as discussed above.

- If ρ is flat, we let

  $$H^1_f(G_\ell, Ad\rho) \subset H^1(G_\ell, Ad\rho)$$

  be the subgroup classifying infinitesimal deformations in the category of profinite $\mathcal{O}/\lambda^n[G_\ell]$-modules that are also flat. Note that $H^1_f(G_\ell, Ad\rho)$ is a functor of ρ. We have not associated any meaning to $H^1_f(G_\ell, M)$ for an arbitrary (ℓ-torsion or finite) G_ℓ-module M. (For M a $\mathbb{Q}_\ell$-vector space this may be done using Fontaine's ring B_{cris} and more generally, this is the subject of *integral p-adic Hodge theory*, but we do not go into it.)
- If ρ is ordinary, we let

  $$H^1_{ord}(G_\ell, Ad\rho) \subset H^1(G_\ell, Ad\rho)$$

  be the subgroup classifying infinitesimal deformations in the category of profinite $\mathcal{O}/\lambda^n[G_\ell]$-modules that remain ordinary. Again, this is a functor of ρ.
- If ρ is semistable (ordinary or flat) we let

  $$H^1_{ss}(G_\ell, Ad\rho) \subset H^1(G_\ell, Ad\rho)$$

  be the subgroup classifying infinitesimal deformations in the category of profinite $\mathcal{O}/\lambda^n[G_\ell]$-modules that are also semistable.

Regarding the relation between these three "tangent spaces", we have:

- If ρ is both ordinary and flat,

  $$H^1_f \subset H^1_{ss} = H^1_{ord}$$

because a flat deformation of an ordinary representation is ordinary. If ρ is flat but not ordinary,

$$H^1_{ss} = H^1_f$$

and H^1_{ord} does not make sense. If ρ is ordinary but not flat,

$$H^1_{ss} = H^1_{ord}$$

and H^1_f does not make sense.

- We define the same subgroups with coefficients in $Ad^0\rho$ by intersecting with $H^1(G_\ell, Ad^0\rho)$.
- Finally, let E be the field of fractions of $\mathcal{O}$. If $\rho : G_\ell \to GL_2(\mathcal{O})$ we define

$$H^1(G_\ell, Ad\rho \otimes E/\mathcal{O}) := \lim_{\rightarrow} H^1(G_\ell, Ad\rho \otimes \lambda^{-n}\mathcal{O}/\mathcal{O})$$

and similarly the H^1_f, H^1_{ord}, H^1_{ss} and the same groups for $Ad^0\rho$, under the usual assumptions.

3.5.3. *Calculations.* Suppose that $\rho : G_\ell \to GL_2(\mathcal{O}/\lambda^n)$ is semistable (this includes the assumption $\det \rho|_{I_\ell} = \epsilon$).

Proposition 47. *(i) If ρ is flat*

$$\#H^1_f(G_\ell, Ad^0\rho) = \#H^0(G_\ell, Ad^0\rho) \cdot \#(\mathcal{O}/\lambda^n).$$

If $n = 1$ and ρ is flat non-ordinary we have $H^0(G_\ell, Ad^0\overline{\rho}) = 0$ and $\dim H^1_f(G_\ell, Ad^0\overline{\rho}) = 1$.

(ii) If ρ is ordinary, let χ_1 and χ_2 be the unramified characters such that

$$\rho \sim \begin{pmatrix} \chi_1\epsilon & * \\ 0 & \chi_2 \end{pmatrix}.$$

Then

$$\#H^1_{ord}(G_\ell, Ad^0\rho) \le \#H^0(G_\ell, Ad^0\rho) \cdot \#(\mathcal{O}/\lambda^n) \cdot \#(\mathcal{O}/(\lambda^n, \chi_1\chi_2^{-1}(\sigma_\ell) - 1)).$$

If ρ is also flat, equality holds.

(iii) If ρ is ordinary non-flat and $n = 1$ (so $\rho = \overline{\rho}$)

$$\#H^1_{ord}(G_\ell, Ad^0\overline{\rho}) = \#k$$

and $H^0(G_\ell, Ad^0\overline{\rho}) = 0$.

Proof. (i) The first statement will be proved in the next section, using Fontaine-Laffaille theory. See also [Co], Main Theorem 3.3.

If $n = 1$ and $\overline{\rho}$ is not ordinary, then by Lemma 46(a) it is absolutely irreducible, so $H^0(G_\ell, Ad^0\overline{\rho}) = 0$ follows from Schur's lemma.

(ii) Let V be the underlying space of ρ and $W = Ad^0\rho = End^0(V)$. Let V_1 be the unique line in V on which I_ℓ acts via ϵ. Let

$$W_1 = \{w \in W \,|\, w(V_1) = 0,\ w(V) \subset V_1\} \simeq Hom(V/V_1, V_1).$$

The group G_ℓ acts on W_1 by the character $\epsilon\chi_1\chi_2^{-1}$. We claim that

$$H^1_{ord}(G_\ell, W) = \ker\left(H^1(G_\ell, W) \to H^1(I_\ell, W/W_1)\right).$$

Indeed, an infinitesimal deformation $\rho' : G_\ell \to GL_2(\mathcal{O}/\lambda^n[\varepsilon])$ of ρ belongs to the subspace on the right if and only if, up to a strict equivalence, $\rho'|_{I_\ell}$ is of the form

$$\begin{pmatrix} \epsilon & * \\ & 1 \end{pmatrix},$$

i.e. only the "upper right corner" gets deformed when we restrict to inertia at ℓ, or the deformation remains ordinary.

Consider the long exact sequence of G_ℓ-cohomology associated with

$$0 \to W_1 \to W \to W/W_1 \to 0.$$

It gives

$$0 \to H^0(G_\ell, W_1) \to H^0(G_\ell, W) \xrightarrow{\alpha} H^0(G_\ell, W/W_1) \xrightarrow{\beta}$$

$$H^1(G_\ell, W_1) \xrightarrow{\gamma} H^1(G_\ell, W) \xrightarrow{\delta} H^1(G_\ell, W/W_1),$$

from where we get

$$\frac{\#H^1_{ord}(G_\ell, W)}{\#H^0(G_\ell, W)} = \frac{\#H^1_{ord}(G_\ell, W)}{\#H^0(G_\ell, W_1)\#\mathrm{Im}\alpha} = \frac{\#H^1_{ord}(G_\ell, W) \cdot \#H^1(G_\ell, W_1)}{\#H^0(G_\ell, W_1)\#\mathrm{Im}\alpha \cdot \#\ker\gamma\#\mathrm{Im}\gamma}$$

$$= \frac{\#H^1(G_\ell, W_1)}{\#H^0(G_\ell, W_1)} \cdot \frac{\#H^1_{ord}(G_\ell, W)}{\#\ker\beta\#\mathrm{Im}\beta\#\mathrm{Im}\gamma} = \frac{\#H^1(G_\ell, W_1)}{\#H^0(G_\ell, W_1)} \cdot \frac{\#H^1_{ord}(G_\ell, W)}{\#H^0(G_\ell, W/W_1)\#\mathrm{Im}\gamma}$$

$$= \frac{\#H^1(G_\ell, W_1)}{\#H^0(G_\ell, W_1)} \cdot \frac{\#H^1_{ord}(G_\ell, W)}{\#H^1(G_\ell/I_\ell, (W/W_1)^{I_\ell})\#\ker\delta}.$$

However, $\ker\delta \subset H^1_{ord}(G_\ell, W)$, so

$$\frac{\#H^1_{ord}(G_\ell, W)}{\#H^1(G_\ell/I_\ell, (W/W_1)^{I_\ell})\#\ker\delta} = \frac{\#\delta(H^1_{ord}(G_\ell, W))}{\#H^1(G_\ell/I_\ell, (W/W_1)^{I_\ell})} \leq 1$$

because

$$\delta(H^1_{ord}(G_\ell, W)) \subset H^1(G_\ell/I_\ell, (W/W_1)^{I_\ell})$$

by the very definition of $H^1_{ord}(G_\ell, W)$.

We conclude that (writing $R = \mathcal{O}/\lambda^n$ for brevity)

$$\frac{\#H^1_{ord}(G_\ell, W)}{\#H^0(G_\ell, W)} \leq \frac{\#H^1(G_\ell, W_1)}{\#H^0(G_\ell, W_1)} = \#R \cdot \#H^2(G_\ell, W_1) =$$

$$= \#R \cdot \#H^0(G_\ell, W_1^*) = \#R \cdot \# \frac{R}{(\chi_1\chi_2^{-1}(\sigma_\ell) - 1)R}.$$

Here the three equalities after the inequality follow from the local Euler characteristic formula, Tate's local duality and the easy fact that

$$\#H^0(G_\ell, W_1^*) = \# \frac{R}{(\chi_1\chi_2^{-1}(\sigma_\ell) - 1)R}.$$

This gives the first statement.

If ρ is also flat, it can be shown that the inequality is an equality, as in [D-D-T], end of section 2.4. This, however, will not be used (and in fact is not proved in [W95]), so we skip it.

(iii) The proof is again somewhat technical. See Proposition 1.9(iv) of [W95], "Choice 2" on p. 116 of [Wa], or section 4.3, p.440, in [dS] for a full proof. Wiles calls "ordinary" by the name "Selmer", and "ordinary non-flat" he calls "strict". $\square$

3.6. **Fontaine-Laffaille theory.** Fontaine-Laffaille theory was an early attempt (from 1982) to establish an *integral p-adic Hodge theory.* It worked under stringent conditions on the absolute ramification index of the ground field, and the Hodge-Tate numbers (required to be, up to a shift, in the range $[0,p-1)$, or, with a litte more care, $[0,p-1]$). In our case $p=\ell$, the ground field is $\mathbb{Q}_\ell$ (so there are no problems) and the Hodge numbers are $\{0,1\}$ (or rather $\{-1,0\}$ since we talk about homology, not cohomology). This excludes only $\ell=2$, that was already excluded for other reasons. Today, integral p-adic Hodge theory has been developed to its full capacity by Breuil and Kisin, and from a perfectoid perspective by Bhatt, Morrow and Scholze.

Recall that (rational) p-adic Hodge theory classifies "good" p-adic representations of $G_p = Gal(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ by semi-linear objects. The basic example is that of crystalline representations, i.e. $\mathbb{Q}_p$-representations of G_p "whose p-adic periods belong to the ring B_{cris}", encompassing all p-adic étale cohomologies of proper smooth varieties with good reduction over $\mathbb{Q}_p$. On the semilinear algebra side Fontaine constructed a category MF_K^φ (here $K=\mathbb{Q}_p$) of filtered φ-modules and an exact functor

$$D_{cris} : Rep_K^{cris} \to MF_K^\varphi.$$

The essential image of the functor are the so-called *admissible* filtered φ-modules. A theorem of Colmez and Fontaine identified them as the weakly admissible modules, defined in terms of a condition on the Hodge and Newton polygons of all sub-objects. Almost by definition of what it means to be "crystalline", D_{cris} becomes an equivalence of categories between Rep_K^{cris} and the full subcategory of weakly admissible modules, ${}^{w.a.}MF_K^\varphi$. It should be remarked that while MF_K^φ is only an additive category, ${}^{w.a.}MF_K^\varphi$ is abelian, and closed under tensor products (an analogue of Totaro's theorem).

Integral p-adic Hodge theory tries to work integrally, not rationally. Assume that the Hodge-Tate numbers (the breaks in the filtration) are in the interval $[0,n]$. A *strongly divisible* lattice in a module $D \in MF_K^\varphi$ is a $\mathbb{Z}_p$-lattice $L \subset D$ such that $\varphi(Fil^i D \cap L) \subset p^i L$ and

$$\sum p^{-i}\varphi(Fil^i D \cap L) = L.$$

The prototypical example is this. Let X be a proper smooth scheme over $\mathcal{O}_K = \mathbb{Z}_p$ and $V = H^n_{et}(X_{\overline{K}}, \mathbb{Q}_p)$. Let $\kappa = \mathbb{F}_p$ so that X_κ is the special fiber. Let $D = H^n_{cris}(X_\kappa/W(\kappa)) \otimes_{W(\kappa)} K$. For φ_D take the crystalline Frobenius. The filtration is induced from the Hodge filtration on $H^n_{dR}(X_K/K)$ and the canonical isomorphism:

$$H^n_{cris}(X_\kappa/W(\kappa)) \otimes_{W(\kappa)} K \simeq H^n_{dR}(X_K/K).$$

The p-adic comparison isomorphism is the statement that

$$D_{cris}(V) = D$$

canonically. Letting $L = H^n_{cris}(X/W(\kappa))/\text{torsion}$ (i.e. the image of the integral crystalline cohomology in D), we get, *under the assumption* $n < p-1$, a strongly divisible lattice in D. An integral comparison isomorphism should relate lattices in the representation V with strongly divisible lattices in D.

In this case the Fontaine-Laffaille modules would be the groups $L/p^r L$. Since, working with torsion coefficients, we can no longer divide Frobenius by p^i on the ith step of the filtration (even if its image is in $p^i\times$ the module), we have to stipulate that the "$p^{-i}\varphi_D$" are part of the given structure. This leads to the following definition.

Definition 48. Let κ be a perfect field and $W = W(\kappa)$ the ring of Witt vectors. A Fontaine-Laffaille module D over W is a W-module *of finite length,* equipped with a descending separated filtration $Fil^\bullet$ with $Fil^0 D = D$, and semilinear maps $\varphi_D^i : Fil^i D \to D$ satisfying (1) $\varphi_D^i|_{Fil^{i+1}D} = p\varphi_D^{i+1}$ and (2) $\sum \varphi_D^i(Fil^i D) = D$.

In their paper [F-L] Fontaine and Laffaille consider the category of Fontaine-Laffaille modules with Hodge-Tate weights $\{0,1\}$. Thus they are looking (writing back ℓ for p) at pairs $(D, D^1 = Fil^1 D)$ and semilinear maps $\varphi_D : D \to D$ and $\varphi_D^1 : D^1 \to D$ such that (1) $\varphi_D|_{D^1} = \ell\varphi_D^1$ and $\varphi_D(D) + \varphi_D^1(D^1) = D$. It is easily seen that these axioms imply that D^1 is in fact a direct summand of D (as an abelian group). It is also clear how to endow everything with a structure of $\mathcal{O}$-modules in case $\mathcal{O}$ is a finite extension of $\mathbb{Z}_\ell$. Furthermore, if D is a Fontaine-Laffaille module with Hodge-Tate numbers $\{0,1\}$ its Cartier dual D^* is defined by letting

$$D^* = Hom(D, \mathbb{Q}_\ell/\mathbb{Z}_\ell), \quad (D^*)^1 = (D^1)^\perp,$$

$$\left\langle \varphi_{D^*}(h), \varphi_D(x) + \varphi_D^1(y) \right\rangle = \langle h, \ell x + y\rangle \quad (x \in D,\ y \in D^1,\ h \in D^*)$$

$$\left\langle \varphi_{D^*}^1(h), \varphi_D(x) \mod \varphi_D^1(D^1) \right\rangle = \langle h, x\rangle \quad (x \in D,\ h \in (D^*)^1).$$

We denote the category of all such modules by $\mathcal{MF}_\mathcal{O}{}^{[0,1]}$.

Theorem 49 (Fontaine-Laffaille)**.** *There are $\mathcal{O}$-additive equivalences between the following categories:*

(a) Finite flat group schemes $\mathscr{G}$ over $\mathbb{Z}_\ell$ with $\mathcal{O}$-action,

(b) Flat $\mathcal{O}[G_\ell]$-modules M of finite cardinality,

(c) $D \in \mathcal{MF}_\mathcal{O}{}^{[0,1]}$.

The equivalences preserve orders (in (b) and (c) the order is just the cardinality) and are compatible with Cartier duality $\mathscr{G} \mapsto \mathscr{G}^*$, $M \mapsto M^*$, $D \mapsto D^*$. If M and D correspond to each other, then M is unramified if and only if $D = D^1$.

Example. Take E, an elliptic curve with good supersingular reduction over $\mathbb{Z}_\ell$, $\mathscr{G} = \mathcal{E}[\ell]$ where $\mathcal{E}$ is its Néron model. Then the restriction of $M_\rho = E(\overline{\mathbb{Q}}_\ell)[\ell]$ to I_ℓ was described in Lemma 46(a). The Fontaine-Laffaille module must be (exercise!) of the shape

$$D = \mathbb{F}_\ell e_1 \oplus \mathbb{F}_\ell e_2, \quad D^1 = \mathbb{F}_\ell e_2$$

$$\varphi_D(e_1) = e_2, \ \ \varphi_D(e_2) = 0,$$

$$\varphi_D^1(e_2) = ce_1, \ \ c \neq 0.$$

In general, the special fiber of $\mathscr{G}$ is connected if and only if φ_D is nilpotent.

Using [F-L], Ramakrishna studied in his thesis [Ra] deformations of supersingular representations, and calculated the tangent space $H_f^1(G_\ell, Ad^0\rho)$. First, the previous theorem gives the following result.

Lemma 50. *Let $\rho : G_\ell \to GL_2(\mathcal{O}/\lambda^n)$ be a flat representation (recall that this means that M_ρ is flat and $\det(\rho) = \epsilon$). The following groups are then isomorphic.*

(a) $H_f^1(G_\ell, Ad\rho)$ (here we do not fix the determinant in the extension).

(b) The group $Ext^1_{\mathcal{O}/\lambda^n[G_\ell],f}(M_\rho, M_\rho)$ of flat extensions of M_ρ by itself in the category of $\mathcal{O}/\lambda^n$-Galois modules.

(c) The group $Ext^1_{\mathcal{MF}^{[0,1]}_{\mathcal{O}/\lambda^n}}(D_\rho, D_\rho)$ *where* D_ρ *is the Fontaine-Laffaille module associated to the finite flat group scheme whose generic fiber is* M_ρ*. Note that the extensions are in the category of Fontaine-Laffaille modules killed by* λ^n*.*

(d) Pairs (α, α^1) *where* $\alpha \in Hom_{\mathcal{O}}(D_\rho, D_\rho)$*,* $\alpha^1 \in Hom_{\mathcal{O}}(D^1_\rho, D_\rho)$*,* $\ell\alpha^1 = \alpha|_{D^1_\rho}$*, taken modulo the group of pairs of the form* $(a \circ \varphi_D - \varphi_D \circ a, a \circ \varphi^1_D - \varphi^1_D \circ a)$*, where* $a \in Hom(D_\rho, D_\rho)$ *satisfies* $a(D^1_\rho) \subset D^1_\rho$*.*

The explicit description in (d) allows an explicit computation of the order of $H^1_f(G_\ell, Ad\rho)$ as

$$\#H^1_f(G_\ell, Ad\rho) = \#(\mathcal{O}/\lambda^n)^2 \cdot \#H^0(G_\ell, Ad^0\rho).$$

As

$$Ad(\rho) = Ad^0(\rho) \oplus \mathcal{O}/\lambda^n$$

and $H^1_f(G_\ell, \mathcal{O}/\lambda^n) = \mathcal{O}/\lambda^n$, one eventually gets part (i) of Proposition 47. Here $H^1_f(G_\ell, \mathcal{O}/\lambda^n)$ refers to the flat infinitesimal deformations of the character $\epsilon \mod \ell^n$. By the results of Raynaud quoted above they are all of the form $(1 + \varepsilon\theta(\sigma))\epsilon(\sigma)$ with $\theta : G_\ell/I_\ell \simeq \widehat{\mathbb{Z}} \to \mathcal{O}/\lambda^n$.

3.7. **A bound on the number of generators.** We can now specify the local conditions $\mathcal{L}_\Sigma = \{L_{\Sigma,v}\}$ figuring in deformations of type $\mathcal{D}_\Sigma$.

- At $p = \ell$, $L_{\Sigma,\ell} = H^1_{ord}(G_\ell, Ad^0\overline{\rho})$ if $\ell \in \Sigma$ (so $\overline{\rho}$ is flat and ordinary) or if $\ell \notin \Sigma$ and $\overline{\rho}$ is not flat.
- $L_{\Sigma,\ell} = H^1_f(G_\ell, Ad^0\overline{\rho})$ if $\ell \notin \Sigma$ and $\overline{\rho}$ is flat.
- At $p \neq \ell$, $L_{\Sigma,p} = H^1(G_p, Ad^0\overline{\rho})$ if $p \in \Sigma$.
- $L_{\Sigma,p} = H^1(G_p/I_p, (Ad^0\overline{\rho})^{I_p})$ if $p \neq \ell$ and $p \notin \Sigma$.

Let $\mathcal{L}^*_\Sigma$ be the dual set of conditions. Note that if $p \neq \ell$ then

- $L^\perp_{\Sigma,p} = H^1(G_p/I_p, (Ad^0\overline{\rho})(1)^{I_p})$ if $p \notin \Sigma$
- $L^\perp_{\Sigma,p} = 0$ if $p \in \Sigma$.

Finally, if $\rho : G_{\mathbb{Q}} \to GL_2(\mathcal{O})$ is a lifting of $\overline{\rho}$ of type Σ we write $H^1_{\mathcal{L}_\Sigma}(G_{\mathbb{Q}}, Ad^0\rho \otimes E/\mathcal{O})$ for the direct limit of the groups $H^1_{\mathcal{L}_\Sigma}(G_{\mathbb{Q}}, Ad^0\rho \otimes \lambda^{-n}/\mathcal{O})$ and similarly for the dual Selmer group.

Theorem 51. *There exists a universal deformation* $(R_\Sigma, \rho^{univ}_\Sigma)$ *of* $\overline{\rho}$ *of type* Σ*. Moreover:*

(a) If E'/E *is a finite extension and* $\mathcal{O}'$ *its ring of integers, then* $R'_\Sigma = R_\Sigma \otimes_{\mathcal{O}} \mathcal{O}'$*.*

(b) The universal deformation ring R_Σ *can be generated as an* $\mathcal{O}$*-algebra by* $\dim H^1_{\mathcal{L}_\Sigma}(G_{\mathbb{Q}}, Ad^0\overline{\rho})$ *elements.*

(c) If $\phi : R_\Sigma \twoheadrightarrow \mathcal{O}$ *is a homomorphism and* $\rho = \phi \circ \rho^{univ}_\Sigma$*,* $\mathfrak{p} = \ker(\phi)$*, then*

$$Hom(\mathfrak{p}/\mathfrak{p}^2, E/\mathcal{O}) \simeq H^1_{\mathcal{L}_\Sigma}(G_{\mathbb{Q}}, Ad^0\rho \otimes_{\mathcal{O}} E/\mathcal{O}).$$

Proof. The representability of the deformation problem $\mathcal{D}_\Sigma$ by $(R_\Sigma, \mathfrak{m}_\Sigma) \in \mathcal{C}_{\mathcal{O}}$ was proved in §3.2. That the tangent space is

$$\boldsymbol{t}_\Sigma = \mathcal{D}_\Sigma(k[\varepsilon]) \simeq H^1_{\mathcal{L}_\Sigma}(G_{\mathbb{Q}}, Ad^0\overline{\rho})$$

was proved in Proposition 43. Since its dual is $\mathfrak{m}_\Sigma/(\mathfrak{m}^2_\Sigma, \lambda)$ and has the same dimension, point (b) follows from Nakayama's lemma.

Finally, for (c) it is enough to show that for $n \geq 1$

$$Hom(\mathfrak{p}/\mathfrak{p}^2, \mathcal{O}/\lambda^n) \simeq H^1_{\mathcal{L}_\Sigma}(G_{\mathbb{Q}}, Ad^0\rho \otimes_{\mathcal{O}} \mathcal{O}/\lambda^n).$$

This is done in a similar manner to the case $n = 1$. □

If $\ell \in \Sigma$ we let

$$d_\ell = \dim H^1_{ss}(G_\ell, Ad^0\overline{\rho}) - \dim H^1_f(G_\ell, Ad^0\overline{\rho}).$$

Recall that $\ell \in \Sigma$ only if $\overline{\rho}|_{G_\ell}$ is flat and ordinary, and we consider deformations that are ordinary but not necessarily flat. The integer d_ℓ measures then the discrepancy between the tangent space of all ordinary deformations and the tangent space of those that are flat (and ordinary).

If $\ell \notin \Sigma$, or if $\ell \in \Sigma$ and $\overline{\rho}$ is not flat, we let $d_\ell = 0$.

Proposition 52. *Assume that, if $\ell = 3$, $\overline{\rho}$ is absolutely irreducible even when it is restricted to the absolute Galois group of $L = \mathbb{Q}(\sqrt{-3})$. The deformation ring R_Σ can be topologically generated, as an $\mathcal{O}$-algebra, by*

$$r_\Sigma = \dim H^1_{\mathcal{L}^*_\Sigma}(G_\mathbb{Q}, Ad^0\overline{\rho}(1)) + d_\ell + \sum_{\ell \neq p \in \Sigma} \dim H^0(G_p, Ad^0\overline{\rho}(1))$$

elements.

Proof. Let $W = Ad\overline{\rho}$. We proved that R_Σ can be generated by $r_\Sigma = \dim H^1_{\mathcal{L}_\Sigma}(G_\mathbb{Q}, W)$ elements. By the Greenberg-Wiles formula (3.1),

$$r_\Sigma - \dim H^1_{\mathcal{L}^*_\Sigma}(G_\mathbb{Q}, W(1)) = \dim H^0(G_\mathbb{Q}, W) - \dim H^0(G_\mathbb{Q}, W(1)) + \\ + \sum_{v \in S} \left(\dim L_v - \dim H^0(G_v, W)\right).$$

(we should include $v = \infty$ in S). We have:

- $H^0(G_\mathbb{Q}, W) = 0$. Indeed, by Schur's lemma, an endomorphism commuting with the Galois action is a scalar, by the absolute irreducibility of $\overline{\rho}$. But there are no scalars of trace 0, since the characteristic is not 2.
- $H^0(G_\mathbb{Q}, W(1)) = 0$. Here, if $\ell = 3$, we need the irreducibility of $\overline{\rho}|_{G_L}$. See the proof of Lemma 40.
- If $v = p \neq \ell$ is finite and $p \notin \Sigma$ then $\dim(L_p) = \dim H^0(G_p, W)$ because $L_p = H^1(G_p/I_p, W^{I_p})$. See §3.4.
- If $v = p \neq \ell$ is finite and $p \in \Sigma$ then $\mathcal{D}_p$ is non-restricted (except for the condition on the determinant), $L_p = H^1(G_p, W)$, so

  $$\dim(L_p) - \dim H^0(G_p, W) = \dim H^2(G_p, W) = \dim H^0(G_p, W(1))$$

  by the local Euler characteristic formula and local Tate duality.
- If $v = \infty$, then $L_\infty = 0$ (because G_∞ has order 2 and $\ell > 2$) and $\dim H^0(G_\infty, W) = 1$ since $\overline{\rho}$ is odd.
- When $v = \ell$ and $\ell \notin \Sigma$ we have

  $$\dim L_\ell - \dim H^0(G_\ell, W) = 1.$$

  Indeed, either $\overline{\rho}$ is flat and $L_\ell = H^1_f(G_\ell, W)$, or $\overline{\rho}$ is ordinary and not flat, in which case $L_\ell = H^1_{ord}(G_\ell, W)$. In both cases, the formula follows from Proposition 47. The "1" from this local computation cancels the "-1" contribution from $v = \infty$.
- Finally, if $v = \ell \in \Sigma$ so $\overline{\rho}$ is flat and ordinary, but $L_\ell = H^1_{ss}(G_\ell, W)$ and not $H^1_f(G_\ell, W)$, we should add d_ℓ to the previous computation.

□

3.8. **Taylor-Wiles primes (week 8).**

3.8.1. *Special auxiliary primes and deformations of type Q.* We introduce a set of auxiliary primes

$$Q = \{q_1, \dots, q_r\}$$

satisfying:

- $q \equiv 1 \mod \ell$
- $\overline{\rho}$ is unramified at q and $\overline{\rho}(\sigma_q)$ has distinct eigenvalues $\overline{\alpha}, \overline{\beta} \in k$.

If k is too small to contain the eigenvalues of $\overline{\rho}(\sigma_q)$ for some $q \in Q$, replace it by its quadratic extension and replace E and $\mathcal{O}$ by the correspodning unramified extension and its ring of integers. More assumptions on the set Q will be imposed later on.

Lemma 53. *Let ρ be a deformation of $\overline{\rho}|_{G_q}$ to a homomorphism $G_q \to GL_2(R)$, $R \in \mathcal{C}_{\mathcal{O}}$. Then there are tamely ramified characters $\xi_1, \xi_2 : G_q \to R^\times$ such that*

$$\rho \sim \begin{pmatrix} \xi_1 & \\ & \xi_2 \end{pmatrix}.$$

Proof. We may assume that

$$\overline{\rho}(\sigma_q) = \begin{pmatrix} \overline{\alpha} & \\ & \overline{\beta} \end{pmatrix}.$$

Since $\overline{\rho}$ is unramified, $\rho(I_q) \subset 1 + M_2(\mathfrak{m}_R)$, so $\rho|_{I_q}$ factors through the maximal pro-ℓ quotient $T_q^{(\ell)}$ of I_q, which is pro-cyclic. Let τ be a topological generator of $T_q^{(\ell)}$ and $\sigma \in G_q$ a lifting of the Frobenius σ_q. Recall that $\sigma\tau\sigma^{-1} = \tau^q$. Since $\overline{\alpha} \neq \overline{\beta}$, by a version of Hensel's lemma, we may assume that our basis for ρ has been chosen so that

$$\rho(\sigma) = \begin{pmatrix} \alpha & \\ & \beta \end{pmatrix}$$

is diagonal, with α and β in R lifting $\overline{\alpha}$ and $\overline{\beta}$. Write

$$\rho(\tau) = 1 + \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in 1 + M_2(\mathfrak{m}_R).$$

Calculating $\rho(\sigma)\rho(\tau)\rho(\sigma)^{-1} = \rho(\tau)^q$ we get that $(\alpha\beta^{-1} - q)b$ and $(\alpha^{-1}\beta - q)c$ lie in $\mathfrak{m}_R \cdot (b, c)$. Since $q \equiv 1 \mod \mathfrak{m}_R$ and $\alpha\beta^{-1} - 1 \notin \mathfrak{m}_R$ we get that

$$(b, c) = \mathfrak{m}_R \cdot (b, c).$$

By Nakayama's lemma $b = c = 0$. It follows that $\rho(\tau)$ is also diagonal, hence ρ is given by two characters ξ_1, ξ_2 as above. □

Let Δ_q be the ℓ-Sylow subgroup of $(\mathbb{Z}/q\mathbb{Z})^\times$ and

$$\chi_q : G_{\mathbb{Q}} \twoheadrightarrow Gal(\mathbb{Q}(\zeta_q)/\mathbb{Q}) \simeq (\mathbb{Z}/q\mathbb{Z})^\times \twoheadrightarrow \Delta_q.$$

Let Δ_Q be the product of the Δ_q for $q \in Q$, and χ_Q the product of the χ_q. Let $\mathfrak{a}_Q$ be the augmentation ideal in $\mathcal{O}[\Delta_Q]$. Observe that

$$\mathcal{O}[\Delta_Q] \simeq \mathcal{O}[S_1, \dots S_r]/((1 + S_1)^{\ell^{n_1}} - 1, \dots, (1 + S_r)^{\ell^{n_r}} - 1)$$

where $|\Delta_{q_i}| = \ell^{n_i}$, and then $\mathfrak{a}_Q = (S_1, \dots, S_r)$.

Corollary 54. *$\xi_1|_{I_q} = \xi_2|_{I_q}^{-1}$ factors through $\chi_q|_{I_q}$: there exists a unique character $\phi_q : \Delta_q \to R^\times$ so that*

$$\xi_1|_{I_q} = \phi_q \circ \chi_q, \;\; \xi_2|_{I_q} = (\phi_q \circ \chi_q)^{-1}.$$

Proof. That $\xi_1|_{I_q} = \xi_2|_{I_q}^{-1}$ follows from the fact that $\epsilon = \det \rho$ is unramified at q (since $q \neq \ell$). Now $\xi_1(I_q) \subset 1 + \mathfrak{m}_R$ since $\overline{\rho}$ is unramified at q; it is therefore pro-ℓ. But $\xi_1|_{I_q}$ factors through the inertia subgroup of $Gal(\mathbb{Q}_q^{ab}/\mathbb{Q}_q)$, which is isomorphic to $\mathbb{Z}_q^\times$ via the q-adic cyclotomic character, by local class field theory (or the local Kronecker-Weber theorem). Since its image is pro-ℓ, it in fact factors through the ℓ-Sylow of $\mathbb{Z}_q^\times$, namely Δ_q. □

Apply all this to the universal deformation of type Q, ρ_Q^{univ}. We get the existence of a unique character $\phi_q : \Delta_q \to R_Q^\times$ such that $\rho_Q^{univ}|_{G_q}$ has the shape given by the lemma, with the $\xi_{i,q}$ as in the corollary. Grouping the r primes $q \in Q$ we get a character $\phi_Q : \Delta_Q \to R_Q^\times$, which we use to give R_Q the structure of a $\mathcal{O}[\Delta_Q]$-algebra.

Proposition 55. *Via ϕ_Q the universal deformation ring R_Q is an $\mathcal{O}[\Delta_Q]$-algebra, and*

$$R_Q/\mathfrak{a}_Q R_Q = R_\emptyset.$$

Proof. Here $R_\emptyset$ is the minimal deformation ring, when Q is empty. Consider the image of ρ_Q^{univ} in $GL_2(R_Q/\mathfrak{a}_Q R_Q)$. By the definition of $\mathfrak{a}_Q$ and the previous corollary, it is unramified at each $q \in Q$. It is therefore "a deformation of type $\emptyset$". By the universal property of $R_\emptyset$ there exists a unique homomorphism $R_\emptyset \to R_Q/\mathfrak{a}_Q R_Q$ bringing $\rho_\emptyset^{univ}$ to $\rho_Q^{univ} \mod \mathfrak{a}_Q$. On the other hand, $\rho_\emptyset^{univ}$ is clearly a "deformation of type Q", so by the universal property of R_Q there exists a unique homomorphism $R_Q \to R_\emptyset$ bringing ρ_Q^{univ} to $\rho_\emptyset^{univ}$. Since $\rho_\emptyset^{univ}$ is unramified at each $q \in Q$, this homomorphism factors through $R_Q/\mathfrak{a}_Q R_Q$. These two homomorphisms are inverse to each other and yield the desired isomorphism between $R_Q/\mathfrak{a}_Q R_Q$ and $R_\emptyset$. □

3.8.2. *A bound on the number of generators of R_Q.*

Proposition 56. *(a) If $q \in Q$ then the spaces $H^0(G_q, Ad^0\overline{\rho})$, $H^0(G_q, Ad^0\overline{\rho}(1))$, $H^1(G_q/I_q, Ad^0\overline{\rho})$ and $H^1(G_q/I_q, Ad^0\overline{\rho}(1))$ are all 1-dimensional.*

(b) Let $r = |Q|$. Then the universal deformation ring R_Q can be topologically generated as an $\mathcal{O}$-algebra by

$$r + \dim H^1_{\mathcal{L}_Q^*}(G_\mathbb{Q}, Ad^0\overline{\rho}(1))$$

elements.

(c) If the set Q is chosen so that, in addition, localization at the primes $q \in Q$ induces an isomorphism

$$H^1_{\mathcal{L}_\emptyset^*}(G_\mathbb{Q}, Ad^0\overline{\rho}(1)) \simeq \prod_{q \in Q} H^1(G_q/I_q, Ad^0\overline{\rho}(1))$$

then $r = |Q| = \dim H^1_{\mathcal{L}_\emptyset^*}(G_\mathbb{Q}, Ad^0\overline{\rho}(1))$ and R_Q can be generated as an $\mathcal{O}$-algebra by r elements.

Proof. In view of the explicit shape of $\overline{\rho}$, we know that $\overline{\rho}$ is unramified and σ_q acts on $Ad^0\overline{\rho}$ with eigenvalues $x, 1, x^{-1}$ for some $1 \neq x \in k^\times$. The same is true for $Ad^0\overline{\rho}(1)$ because $q \equiv 1 \mod \ell$ so the twist by $\overline{\epsilon}$, the cyclotomic character mod ℓ, does not change $Ad^0\overline{\rho}|_{G_q}$. The calculations of the cohomology classes in (a) become an easy exercise.

Part (b) follows now from (a) and Proposition 52. Note that $d_\ell = 0$ since $\ell \notin Q$.

For (c) note that in $\mathcal{L}_\emptyset$ we had (for $q \in Q$) $L_q = H^1(G_q/I_q, Ad^0\overline{\rho})$ and $L_q^\perp = H^1(G_q/I_q, Ad^0\overline{\rho}(1))$, while in $\mathcal{L}_Q$ we relaxed the condition of being unramified at q to $L_q = H^1(G_q, Ad^0\overline{\rho})$, so $L_q^\perp = 0$. Thus the dual Selmer group "with conditions at Q" is

$$\ker\left(H^1_{\mathcal{L}_\emptyset^*}(G_{\mathbb{Q}}, Ad^0\overline{\rho}(1)) \overset{loc_Q}{\to} \prod_{q\in Q} H^1(G_q/I_q, Ad^0\overline{\rho}(1))\right). \tag{3.2}$$

If loc_Q is an isomorphism then the kernel vanishes, $H^1_{\mathcal{L}_Q^*}(G_{\mathbb{Q}}, Ad^0\overline{\rho}(1)) = 0$ and (c) follows from (b). Since each $H^1(G_q/I_q, Ad^0\overline{\rho}(1))$ is one-dimensional, we get also that $H^1_{\mathcal{L}_\emptyset^*}(G_{\mathbb{Q}}, Ad^0\overline{\rho}(1))$ was r-dimensional. □

3.8.3. *On the choice of Q: an application of Čebotarev's density theorem and some group theory.* We are left with the task of proving that a set Q as above, satisfying also the condition

$$loc_Q : H^1_{\mathcal{L}_\emptyset^*}(G_{\mathbb{Q}}, Ad^0\overline{\rho}(1)) \simeq \prod_{q\in Q} H^1(G_q/I_q, Ad^0\overline{\rho}(1))$$

can be chosen. For the application we would like also that each $q \in Q$ satisfies $q \equiv 1 \mod \ell^n$ for a fixed $n \geq 1$. This will guarantee that the ring $\mathcal{O}[\Delta_Q]$ is large. In fact, "in the limit" on n, it will become $\mathcal{O}[[S_1, \ldots, S_r]]$, the formal power series ring in r variables over $\mathcal{O}$.

Theorem 57 (Existence of Taylor-Wiles primes)**.** *Assume that $\overline{\rho}$ remains absolutely irreducible when restricted to G_L, $L = \mathbb{Q}(\sqrt{(-1)^{(\ell-1)/2}\ell})$, the quadratic subfield of $\mathbb{Q}(\zeta_\ell)$. Consider the minimal deformation problem "of type $\emptyset$" and let*

$$r = \dim H^1_{\mathcal{L}_\emptyset^*}(G_{\mathbb{Q}}, Ad^0\overline{\rho}(1))$$

be the dimension of its "dual Selmer group". Fix $n \geq 1$. Then there exists a set Q_n of r primes q such that

1. *Each $q \equiv 1 \mod \ell^n$,*
2. *If $q \in Q_n$ then $\overline{\rho}$ is unramified at q and $\overline{\rho}(\sigma_q)$ has distinct eigenvalues (which we may assume, belong to k),*
3. *The universal deformation ring R_{Q_n} can be topologically generated as an $\mathcal{O}$-algebra by r elements.*

Proof. It is enough to find a set Q satisfying the first two conditions, such that

$$H^1_{\mathcal{L}_Q^*}(G_{\mathbb{Q}}, Ad^0\overline{\rho}(1)) = 0.$$

Since this dual Selmer group is given by (3.2) we may inductively find r primes q such that the condition $loc_q([\psi]) = 0$ imposes each time a non-empty condition on the common kernel of loc_q for the previous q's. Since $H^1(G_q/I_q, Ad^0\overline{\rho}(1)) = 1$, the vanishing of loc_q for the new prime q will decrease the dimension of the kernel by 1, and after r steps we will be done.

Fix $[\psi] \in H^1_{\mathcal{L}_\emptyset^*}(G_{\mathbb{Q}}, Ad^0\overline{\rho}(1))$. Here ψ is a 1-cocycle representing the cohomology class $[\psi]$. We may assume that $[\psi]$ is in the common kernel of loc_q for the q's found so far, and look for a new q satisfying (1) and (2) such that $loc_q([\psi]) \neq 0$. Write

$W = Ad^0\overline{\rho}$, $W^* = Ad^0\overline{\rho}(1)$. By Čebotarev's density theorem it is enough to find a $\sigma \in G_{\mathbb{Q}}$ such that

(1) $\sigma|_{\mathbb{Q}(\zeta_{\ell^n})} = 1$,
(2) The eigenvalues of $\overline{\rho}(\sigma)$ are distinct,
(3) $\psi_\sigma \notin (\sigma - 1)W^*$.

Let M be a finite Galois extension of $\mathbb{Q}$, containing ζ_{ℓ^n}, which is a splitting field for $\overline{\rho}$ and the cocycle ψ. Let q be a prime, unramified in M, such that $\sigma|_M = (\mathfrak{Q}, M/\mathbb{Q})$ for a suitable prime $\mathfrak{Q}|q$ of M. Then the first two conditions on σ imply the first two conditions on q, and the third implies that $\psi|_{G_q}$ is not a coboundary, because $\sigma \in G_q$ (more precisely, in the decomposition group of $\mathfrak{Q}/q$).

Consider the tower of fields

$$\begin{array}{c} F \\ | \\ K \\ |H \\ \mathbb{Q}(\zeta_{\ell^n}) \end{array}$$

where F is the splitting field of $\overline{\rho}|_{Gal(\overline{\mathbb{Q}}/\mathbb{Q}(\zeta_{\ell^n}))}$, and K is the splitting field of $Ad^0\overline{\rho}|_{Gal(\overline{\mathbb{Q}}/\mathbb{Q}(\zeta_{\ell^n}))}$. Note that F and K are Galois over $\mathbb{Q}$, since they are the compositums of $\mathbb{Q}(\zeta_{\ell^n})$ with the splitting fields of the corresponding representations on the full $G_{\mathbb{Q}}$. We write

$$H = Gal(K/\mathbb{Q}(\zeta_{\ell^n})) \simeq Ad^0\overline{\rho}(G_{\mathbb{Q}(\zeta_{\ell^n})}), \;\; \widetilde{H} = Gal(F/\mathbb{Q}(\zeta_{\ell^n})) \simeq \overline{\rho}(G_{\mathbb{Q}(\zeta_{\ell^n})}).$$

We similarly write G and $\widetilde{G}$ for $Ad^0\overline{\rho}(G_{\mathbb{Q}})$ and $\overline{\rho}(G_{\mathbb{Q}})$. Since $\overline{\rho}$ is absolutely irreducible, Schur's lemma implies that $G \simeq \widetilde{G}k^\times/k^\times \subset PGL_2(k)$ (when we regard $\widetilde{G} \subset GL_2(k)$). The group $\widetilde{H}$ is a subgroup of $\widetilde{G}$ and H is again its projective image

$$H = \widetilde{H}k^\times/k^\times \subset PGL_2(k).$$

Lemma 58. $H^1(Gal(K/\mathbb{Q}), W^*) = 0$.

We postpone the proof of the lemma, and conclude the proof of the theorem. By the lemma, and the inflation-restriction exact sequence, the non-vanishing of $[\psi]$ implies that $Res_K^{\mathbb{Q}}[\psi] \neq 0$. But over K the Galois action on W^* is trivial, so

$$0 \neq \psi|_{G_K} \in Hom(G_K, W^*)^{Gal(K/\mathbb{Q})}.$$

It follows that $\psi(G_K)$ is a $G_{\mathbb{Q}}$-submodule of W^*. The absolute irreducibility of $\overline{\rho}|_{G_L}$ implies the irreducibility of W^* (see the argument in the proof of Proposition 52). Thus, $\langle\psi(G_K)\rangle = W^*$. Here, for a subgroup A, we denote by $\langle A\rangle$ its k-linear span.

We claim that there exists a $\sigma_0 \in G_{\mathbb{Q}(\zeta_{\ell^n})}$ such that $\overline{\rho}(\sigma_0)$ has distinct eigenvalues. If not, $\overline{\rho}(G_{\mathbb{Q}(\zeta_{\ell^n})})$ is contained in a group conjugate to

$$\left\{ \begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \right\}.$$

If $\overline{\rho}(G_{\mathbb{Q}(\zeta_{\ell^n})})$ consist of scalar matrices only, it is easily seen that $\overline{\rho}$ can not be absolutely irreducible: any eigenvector of $\overline{\rho}(\gamma)$, where γ projects to a generator of the cyclic group $Gal(\mathbb{Q}(\zeta_{\ell^n})/\mathbb{Q})$, would span a line invariant under $\overline{\rho}(G_{\mathbb{Q}})$. If, on the other hand, $\overline{\rho}(G_{\mathbb{Q}(\zeta_{\ell^n})})$ contains non-scalar matrices, it has a unique invariant line, which must then be invariant also under $\overline{\rho}(G_{\mathbb{Q}})$. In both cases, this contradicts the absolute irreducibility of $\overline{\rho}$.

Let α, β be the eigenvalues of $\overline{\rho}(\sigma_0)$. The eigenvalues of $Ad^0\overline{\rho}(\sigma_0)$ are α/β, 1 and β/α. These are also the eigenvalues of $Ad^0\overline{\rho}(1)(\sigma_0)$, since $q \equiv 1 \mod \ell$. As one of the eigenvalues is 1,

$$0 \neq (\sigma_0 - 1)W^* \neq W^* = \langle \psi(G_K) \rangle .$$

Let $\tau \in G_K$. It acts trivially on W and on W^*, so $\overline{\rho}(\tau)$ is a scalar matrix. It follows that $\sigma = \tau\sigma_0 \in G_{\mathbb{Q}(\zeta_{\ell^n})}$ still has distinct eigenvalues under $\overline{\rho}$. But τ acts trivially on W^* so

$$\psi_\sigma = \tau\psi_{\sigma_0} + \psi_\tau = \psi_{\sigma_0} + \psi_\tau .$$

As the ψ_τ, for $\tau \in G_K$, span W^* over k, we can find a τ so that $\psi_\sigma \notin (\sigma_0 - 1)W^*$. However, $(\sigma_0 - 1)W^* = (\sigma - 1)W^*$ since τ acts trivially on W^*. This shows that (3) can be guaranteed too. □

The proof of the lemma is group-theoretic. It relies on the classification of finite subgroups of $PGL_2(\overline{k})$. According to a classical theorem of Dickson, every such finite group is one of the following:

- Contained in a Borel subgroup of $PGL_2(\overline{k})$,
- Conjugate to $PGL_2(k')$ or $PSL_2(k')$ for a finite field k',
- Isomorphic to the dihedral group D_{2n} for $(n, \ell) = 1$, or
- Isomorphic to A_4, S_4 or A_5.

Let $Z = \ker(\widetilde{G} \to G)$, the scalar matrices in $Im(\overline{\rho})$. If $Z \neq \{\pm 1\}$ $\det(Z) \neq 1$, so, $W = Ad^0\overline{\rho}$ being invariant under Z, $W^{*Z} = 0$. Note Z is cyclic of order prime to ℓ, so in particular $H^1(Z, W^*) = H^2(Z, W^*) = 0$. If we denote by $M \subset F$ the splitting field of $\overline{\rho}$, so that $\widetilde{G} = Im(\overline{\rho}) = Gal(M/\mathbb{Q})$, then

$$\mathbb{Q}(\zeta_\ell) \subset M \subset F \subset M(\zeta_{\ell^n}).$$

It follows that $Gal(F/M)$ is a normal ℓ-subgroup of $Gal(F/\mathbb{Q})$. Thus $Z \subset Gal(M/\mathbb{Q})$, an abelian group whose order is prime to ℓ, lifts to a subgroup of $Gal(F/\mathbb{Q})$, which we still denote by Z. From the inflation-restriction exact sequence, the vanishing of $H^i(Z, W^*)$ for $i = 1, 2$ and the vanishing of W^{*Z},

$$0 = H^1(Gal(F/\mathbb{Q})/Z, W^{*Z}) \simeq H^1(Gal(F/\mathbb{Q}), W^*).$$

A fortiori, $H^1(Gal(K/\mathbb{Q}), W^*) = 0$.

When $Z = \{\pm 1\}$ but $\ell > 3$, Z fixes $\mathbb{Q}(\zeta_\ell)$. The group $\Delta = Gal(\mathbb{Q}(\zeta_\ell)/\mathbb{Q})$ is a quotient of $\widetilde{G}$, because M, the splitting field of $\overline{\rho}$, must contain $\mathbb{Q}(\zeta_\ell)$, the splitting field of $\det \overline{\rho} = \overline{\epsilon}$. As Z acts trivially on $\mathbb{Q}(\zeta_\ell)$, Δ is in fact a quotient of $G = \widetilde{G}/Z \subset PGL_2(k)$. Using Dickson's classification theorem and the fact that $\ell \geq 5$, we see that $G = Im(Ad^0\overline{\rho})$ must be a subgroup of a Borel, or of order prime to ℓ. (The other subgroups do not have a cyclic quotient of order $\ell - 1$.) The first option contradicts the irreducibility of $\overline{\rho}$. The second implies that H, too, has order prime to ℓ. Inflation-restriction yields

$$H^1(Gal(K/\mathbb{Q}), W^*) \simeq H^1(Gal(\mathbb{Q}(\zeta_{\ell^n})/\mathbb{Q}, W^{*H}).$$

However, $W^{*H} = 0$, or else W^* would be reducible (the invariants would be stable under the full $Gal(K/\mathbb{Q})$), contradicting the absolute irreducibility of $\overline{\rho}|_{G_L}$, as before.

There remains the case $Z = \{\pm 1\}$ and $\ell = 3$ (the prime ℓ that we end up using!). Here one must resort to a case-by-case study and to a theorem of Cline, Parshal and Scott from 1975 on cohomology of finite groups of Lie type. See the original proof in Wiles' paper or [dS], Theorem 20, p.443-444, for details.

4. The Hecke algebra $\mathbb{T}_\Sigma$ and the proof of $R_\Sigma \simeq \mathbb{T}_\Sigma$

4.1. **Modularity of the residual representation (Langlands-Tunnell) (week 9).**

4.1.1. *Artin conductors.* Fix a residual mod-ℓ representation

$$\overline{\rho} : G_{\mathbb{Q}} \to GL_d(k).$$

For $p \neq \ell$, let $\{I_p^u | -1 < u < \infty\}$ be the (decreasing) upper filtration of the inertia group at p, so that $I_p^u = I_p$ for $-1 < u \leq 0$, $I_p^v = \bigcap_{u<v} I_p^u$, and the wild inertia $P_p = \bigcup_{0<u} I_p^u$. The prime-to-$\ell$ Artin conductor of $\overline{\rho}$ is

$$N(\overline{\rho}) = \prod_{\ell \neq p} p^{m_p(\overline{\rho})}$$

where the exponent at p is given by

$$m_p(\overline{\rho}) = \int_{-1}^{\infty} \mathrm{codim} V_{\overline{\rho}}^{I_p^u} du = \mathrm{codim} V_{\overline{\rho}}^{I_p} + \int_0^{\infty} \mathrm{codim} V_{\overline{\rho}}^{I_p^u} du.$$

It is known that $m_p(\overline{\rho})$ is an integer, and vanishes if and only if $\overline{\rho}$ is unramified at p. If $\rho : G_{\mathbb{Q}} \to GL_d(\mathcal{O})$ is a lifting of $\overline{\rho}$ then $m_p(\rho)$ is defined by the same formula. As the kernel of $GL_d(\mathcal{O}) \to GL_d(k)$ is pro-ℓ, $\rho(P_p) \simeq \overline{\rho}(P_p)$ is finite, so the integral defining $m_p(\rho)$ is finite. Moreover, it follows from here that

$$m_p(\rho) = m_p(\overline{\rho}) + \left(\dim V_{\overline{\rho}}^{I_p} - \dim V_{\rho}^{I_p}\right),$$

because the part of the integral for $u > 0$ is the same for ρ and $\overline{\rho}$.

Suppose now $d = 2$ and $\overline{\rho}$ satisfies the "running assumptions" of §3.1.1. It follows that $N(\overline{\rho})$ is square-free, i.e. if $\overline{\rho}$ is ramified ("type A") at $p \neq \ell$, then $m_p(\overline{\rho}) = 1$, because it is then tamely ramified and the inertia invariants are 1- dimensional. The same holds true with any ℓ-adic deformation ρ which is "minimal" (i.e. again type A) at p. Thus if ρ is a "type Σ" ℓ-adic deformation and $p \notin \Sigma$ then $m_p(\rho) = m_p(\overline{\rho})$: either both are 0, in the unramified case, or both are 1, in the ramified type A case.

If, on the other hand, $p \in \Sigma$, so ρ is not minimal at p, then $0 \leq m_p(\rho) - m_p(\overline{\rho}) \leq 2$.

These remarks become important when we look for newforms f that might give rise to ℓ-adic deformations ρ "of type Σ" of $\overline{\rho}$. If there is such an f, and λ is the prime above ℓ in $\mathbb{Q}(a_n(f))$ giving rise to the deformation $\rho = \rho_{f,\lambda}$, then by Carayol's theorem the analytic conductor N_f (i.e. the level of f) will be equal to the Artin conductor of $\rho_{f,\lambda}$. Thus, for $\ell \neq p \notin \Sigma$, $p \nmid N_f$ if $\overline{\rho}$ were unramified at p and $p || N_f$ if $\overline{\rho}$ were ramified there. On the other hand, if $p \in \Sigma$ we would have $ord_p N_f \leq 2$ if $\overline{\rho}$ were unramified at p and ≤ 3 if $\overline{\rho}$ were ramified there. In practice, for the application to modularity of elliptic curves, we shall only have to include p in Σ if $\overline{\rho}$ is unramified there.

A similar analysis takes place at ℓ, where by assumption $\overline{\rho}$ is semistable (flat or ordinary). Let $\delta(\overline{\rho}) = 1$ if $\overline{\rho}$ is (ordinary) non-flat at ℓ, and 0 otherwise. Then, by Deligne's theorem (based on the work of Deligne and Rapoport) we would have $\ell || N_f$ if either $\delta(\overline{\rho}) = 1$, or if $\overline{\rho}$ were flat and ordinary at ℓ, but $\ell \in \Sigma$ and $\rho_{f,\lambda}$ is non-flat. At all other cases (where $\rho_{f,\lambda}$ stays flat), $\ell \nmid N_f$.

4.1.2. *Modularity of* $\overline{\rho}$. Fix a (finite, possibly empty) set of finite primes Σ such that

(a) if $\ell \in \Sigma$ then $\overline{\rho}$ is flat and ordinary at ℓ,

(b) if $\ell \neq p \in \Sigma$ then $\overline{\rho}$ is unramified at p.

The second assumption is not essential, but since it suffices for the application to modularity of elliptic curves, and it makes life somewhat easier, we impose it.

Let $\mathcal{N}_\Sigma$ be the collection of triples (f, λ_f, ι_f) consisting of a newform f of weight 2, level N_f and trivial nebentypus, a prime λ_f above ℓ in $\mathcal{O}_f$, the ring of integers of $K_f = \mathbb{Q}(a_n(f))$, an embedding

$$\iota_f : \mathcal{O}_{f,\lambda_f} \hookrightarrow \mathcal{O}'_f$$

into a finite extension of $\mathcal{O}$, with uniformizer λ'_f and residue field k'_f containing k, such that:

(i) $\overline{\rho}$ and $\iota_f \circ \overline{\rho}_{f,\lambda_f}$ become isomorphic over k'_f,

(ii) over $\mathcal{O}'_f$, $\iota_f \circ \rho_{f,\lambda_f}$ is a deformation of type Σ of $\overline{\rho}$.

By the discussion of Artin conductors and assumptions (a) and (b), for any $f \in \mathcal{N}_\Sigma$ we have $ord_p(N_f) \leq 2$ for $p \neq \ell$, $ord_p(N_f) = 0$ if $\overline{\rho}$ is unramified at p and $p \notin \Sigma$, and $ord_\ell(N_f) \leq 1$. Since the determinant of a deformation of type Σ is the cyclotomic character, the nebentypus of f is trivial. We conclude that the collection $\mathcal{N}_\Sigma$ is finite, at most.

Once we know that $\mathcal{N}_\Sigma$ is finite, we may assume, enlarging $E, \mathcal{O}$ and k, that $\mathcal{O} = \mathcal{O}'_f$ for every f in $\mathcal{N}_\Sigma$, so that ι_f becomes an embedding $\mathcal{O}_{f,\lambda_f} \hookrightarrow \mathcal{O}$, inducing $\mathcal{O}_f/\lambda_f \hookrightarrow k$.

The following theorem follows from the work of Langlands and Tunnell on the Artin conjecture, the paper [De-Se74], and Ribet's theorem [Ri90].

Theorem 59. *Assume that $k = \mathbb{F}_3$. Then the collection $\mathcal{N}_\Sigma$ is not empty. In other words, the representation $\overline{\rho}$ is modular, and moreover it is modular of level $\ell^{\delta(\overline{\rho})} N(\overline{\rho})$, weight 2 and trivial nebentypus, as predicted by Serre.*[6]

Proof. (Sketch) As $GL_2(\mathbb{F}_3)$ is solvable, so is the image of $\overline{\rho}$. Consider the reduction map

$$GL_2(\mathbb{Z}[\sqrt{-2}]) \twoheadrightarrow GL_2(\mathbb{F}_3)$$

modulo $\varpi = (1 + \sqrt{-2})$ (one of the primes above 3). One can check directly that it admits a section

$$s : GL_2(\mathbb{F}_3) \hookrightarrow GL_2(\mathbb{Z}[\sqrt{-2}]) \subset GL_2(\mathbb{C}).$$

In fact, this s is one of the three cuspidal representations of $GL_2(\mathbb{F}_3)$. The representation $s \circ \overline{\rho}$ is odd, irreducible (otherwise it would be abelian, and so would be $\overline{\rho}$), and its image is solvable. Applying the Langlands-Tunnell theorem on the Artin conjecture to $s \circ \overline{\rho}$ one obtains a weight 1 newform g (of some level and nebentypus) whose associated Galois representation ρ_g is $s \circ \overline{\rho}$. In other words, for all but finitely many primes p we have $a_p(g) = tr(s \circ \overline{\rho}(\sigma_p))$, so

$$a_p(g) \mod \varpi = tr(\overline{\rho}(\sigma_p)).$$

[6] We might need here the assumption that $\overline{\rho}|_{G_L}$ remains absolutely irreducible. See remark in the proof.

The problem is that g has weight 1, not 2. Here comes an idea of Shimura. Let $\chi(d) = \left(\frac{-3}{d}\right)$ (Legendre symbol) and

$$E_{1,\chi} = 1 + 6\sum_{n=1}^{\infty}(\sum_{d|n}\chi(d))q^n \in M_1(\Gamma_0(3),\chi).$$

Then $gE_{1,\chi}$ is a weight 2 cusp-form, and since the q-expansion of $E_{1,\chi}$ is $1 \mod 3$,

$$a_p(gE_{1,\chi}) \mod \varpi = tr(\overline{\rho}(\sigma_p)).$$

However, $gE_{1,\chi}$ is not an eigenform. The Deligne-Serre Lemma solves this issue: it guarantees the existence of a newform f (of weight 2, some level and nebentypus) and a prime λ above ϖ in K_fK_g such that

$$a_p(f) \mod \lambda = tr(\overline{\rho}(\sigma_p)).$$

Finally, Ribet's theorem on lowering the level (*we might need the assumption on $\overline{\rho}|_{G_L}$ being absolutely irreducible; it was needed in Ribet's original theorem, and it is not clear to me if Diamond's work on the refined Serre conjecture really removed it in the form needed here*) guarantees that we can take f of weight 2, level $\ell^{\delta(\overline{\rho})}N(\overline{\rho})$, and trivial nebentypus. □

4.1.3. *The Hecke algebra $\mathbb{T}_\Sigma$ and the map $R_\Sigma \to \mathbb{T}_\Sigma$.* As remarked above, since $\mathcal{N}_\Sigma$ is finite, we may enlarge $\mathcal{O}$ and k and assume that for each $(f,\lambda_f,\iota_f) \in \mathcal{N}_\Sigma$, $\iota_f : \mathcal{O}_{f,\lambda_f} \hookrightarrow \mathcal{O}$ induces $\mathcal{O}_f/\lambda_f \hookrightarrow k$. Associated to it we get a representation

$$\iota_f \circ \rho_{f,\lambda_f} : G_{\mathbb{Q}} \to GL_2(k).$$

Let $\mathfrak{T}$ be the abstract polynomial algebra generated over $\mathcal{O}$ by the variables T_p for p a prime different from ℓ, the primes dividing $N(\overline{\rho})$ or the primes in Σ. For $(f,\lambda_f,\iota_f) \in \mathcal{N}_\Sigma$ the representation $\iota_f \circ \rho_{f,\lambda_f}$ is unramified at p and we consider the homomorphism

$$\mathfrak{T} \to \widetilde{\mathbb{T}}_\Sigma = \prod_{f\in\mathcal{N}_\Sigma}\mathcal{O},$$

sending T_p to $(...,\iota_f(a_p(f)),...)$. Let $\mathbb{T}_\Sigma$ be its image. Since the reduction of $\iota_f(a_p(f))$ modulo λ is $tr(\overline{\rho}(\sigma_p))$, independently of f, the ring $\mathbb{T}_\Sigma$ is a *local* ring, with residue field k, and maximal ideal generated by λ and $T_p - a_p$, where $a_p \in \mathcal{O}$ is any lifting of $tr(\overline{\rho}(\sigma_p))$, for all the "good" primes p as above. The ring $\mathbb{T}_\Sigma$ is evidently finite flat (free as a module) over $\mathcal{O}$, and belongs to $\mathcal{C}_{\mathcal{O}}$. It is the ring obtained by "gluing" the $\mathcal{O}_{f,\lambda_f}$, and the higher the congruences between the various ρ_{f,λ_f}, the more "gluing" there is. When we tensor with $\mathbb{Q}$ we get

$$\mathbb{T}_{\Sigma,\mathbb{Q}} = \widetilde{\mathbb{T}}_{\Sigma,\mathbb{Q}} = \prod_{f\in\mathcal{N}_\Sigma} E,$$

because the $\mathbb{Q}$-algebra generated by $(a_p(f_1),...,a_p(f_n))$ for distinct newforms $f_1,...,f_n$ and all $p \notin S$ (S finite) is $K_{f_1}\times\cdots\times K_{f_n}$.

The next lemma, due to Carayol, shows that not only the integral Hecke rings $\mathcal{O}_{f,\lambda_f}$ glue, but the representations glue as well.

Lemma 60 (Carayol's Lemma). *There is a continuous representation*

$$\rho_\Sigma^{mod} : G_{\mathbb{Q}} \to GL_2(\mathbb{T}_\Sigma)$$

such that if $p \nmid \ell N(\overline{\rho})\Sigma$ then ρ_Σ^{mod} is unramified at p, and $tr(\rho_\Sigma^{mod}(\sigma_p)) = T_p$. Moreover, *(a)* ρ_Σ^{mod} *is a lift of type* Σ *of* $\overline{\rho}$*, and there is a unique surjection*

$$\phi_\Sigma : R_\Sigma \twoheadrightarrow \mathbb{T}_\Sigma$$

bringing ρ_Σ^{univ} *to* ρ_Σ^{mod} *(up to strict equivalence).*

(b) If $\Sigma \subset \Sigma'$ *there is a unique surjection* $\mathbb{T}_{\Sigma'} \twoheadrightarrow \mathbb{T}_\Sigma$ *bringing* $\rho_{\Sigma'}^{mod}$ *to* ρ_Σ^{mod}*, and compatible with the images of* T_p *for* $p \nmid \ell N(\overline{\rho})\Sigma'$*.*

(c) The formation of $\mathbb{T}_\Sigma$ *is compatible with extensions of* $\mathcal{O}$*.*

Proof. Everything hinges on showing that the canonical representation

$$\widetilde{\rho}_\Sigma^{mod} : G_\mathbb{Q} \to GL_2(\widetilde{\mathbb{T}}_\Sigma),$$

whose f-coordinate is $\iota_f \circ \rho_{f,\lambda_f}$, can be conjugated so that it factors through $\mathbb{T}_\Sigma$. Let c be a complex conjugation. It is possible to conjugate $\widetilde{\rho}_\Sigma^{mod}$ so that

$$\widetilde{\rho}_\Sigma^{mod}(c) = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}.$$

Let $\{e_+, e_-\}$ be the corresponding basis. For any $\gamma \in G_\mathbb{Q}$, both $tr(\gamma)$ and $tr(c\gamma)$ belong to $\mathbb{T}_\Sigma$ (which is generated by the traces), so if

$$\widetilde{\rho}_\Sigma^{mod}(\gamma) = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

both $a \pm d$, hence also $a, d \in \mathbb{T}_\Sigma$.

By the irreducibility of $\overline{\rho}$ we can find a $\sigma \in G_\mathbb{Q}$ with $\overline{\rho}(\sigma)$ having $\overline{b} \neq 0$. Rescaling e_+ by a unit of $\widetilde{\mathbb{T}}_\Sigma$ we may assume that $\widetilde{\rho}_\Sigma^{mod}(\sigma)$ has $b = 1$. For any $\gamma \in G_\mathbb{Q}$ we have

$$\widetilde{\rho}_\Sigma^{mod}(\sigma\gamma) = \begin{pmatrix} a_\sigma a_\gamma + c_\gamma & * \\ * & * \end{pmatrix}$$

so also $c_\gamma \in \mathbb{T}_\Sigma$. Similarly, $b_\gamma \in \mathbb{T}_\Sigma$ and we are done. $\square$

4.2. **Some results on the Hecke algebra.** We shall need two deep results on the structure of $\mathbb{T}_\Sigma$. They will be proved later on, and in the meanwhile we assume them and continue with the proof of "$R = T$".

4.2.1. *Freeness over the diamond operators.*

Theorem 61. *Let* Q *be a set of Taylor-Wiles primes, consider the homomorphism* $\phi_Q : R_Q \to \mathbb{T}_Q$ *and equip* $\mathbb{T}_Q$ *with a structure of an* $\mathcal{O}[\Delta_Q]$*-algebra via* ϕ_Q*. Then* $\mathbb{T}_Q$ *is free of finite rank over* $\mathcal{O}[\Delta_Q]$*.*

Corollary 62. $\mathbb{T}_\emptyset = \mathbb{T}_Q/\mathfrak{a}_Q\mathbb{T}_Q$.

Proof. By the theorem, if $\mathbb{T}_Q \simeq \mathcal{O}[\Delta_Q]^m$, then $\mathbb{T}_Q/\mathfrak{a}_Q\mathbb{T}_Q \simeq \mathcal{O}^m$ is $\mathcal{O}$-torsion free. It is therefore enough to show that $\mathbb{T}_{Q,\mathbb{Q}}/\mathfrak{a}_Q\mathbb{T}_{Q,\mathbb{Q}} = \mathbb{T}_{\emptyset,\mathbb{Q}}$ when we consider the *rational* Hecke algebras as modules over $E[\Delta_Q]$. From the diagram

$$\begin{array}{ccc} R_{Q,\mathbb{Q}} & \twoheadrightarrow & R_{Q,\mathbb{Q}}/\mathfrak{a}_Q R_{Q,\mathbb{Q}} = R_{\emptyset,\mathbb{Q}} \\ \downarrow & & \downarrow \\ \mathbb{T}_{Q,\mathbb{Q}} = \prod_{f \in \mathcal{N}_Q} E & \twoheadrightarrow & \mathbb{T}_{Q,\mathbb{Q}}/\mathfrak{a}_Q\mathbb{T}_{Q,\mathbb{Q}} \end{array}$$

of E-algebra homomorphisms, where the vertical arrows are surjective, we get that a direct factor E labeled by an $f \in \mathcal{N}_Q$ survives in the map to $\mathbb{T}_{Q,\mathbb{Q}}/\mathfrak{a}_Q\mathbb{T}_{Q,\mathbb{Q}}$ (the corresponding idempotent maps to a non-zero idempotent) if and only if the corresponding ρ_{f,λ_f} factors through $\rho_\emptyset^{univ}$, if and only if ρ_{f,λ_f} is unramified at the primes of Q. But this holds if and only if $f \in \mathcal{N}_\emptyset$. Thus

$$\mathbb{T}_{Q,\mathbb{Q}}/\mathfrak{a}_Q\mathbb{T}_{Q,\mathbb{Q}} = \prod_{f\in\mathcal{N}_\emptyset} E = \mathbb{T}_{\emptyset,\mathbb{Q}}.$$

□

4.2.2. *The congruence ideal $\eta_{\Sigma,f}$.* Consider a homomorphism

$$\pi_{\Sigma,f} : \mathbb{T}_\Sigma \to \mathcal{O}.$$

Such a homomorphism extends to a homomorphism $\mathbb{T}_{\Sigma,\mathbb{Q}} \to E$, so is equivalent to giving the newform $f \in \mathcal{N}_\Sigma$, for which $\pi_{\Sigma,f}(T_p) = a_p(f)$. If $\Sigma \subset \Sigma'$ is enlarged, then $\pi_{\Sigma',f}$ is obtained from $\pi_{\Sigma,f}$ by composing it with the canonical projection $\mathbb{T}_{\Sigma'} \twoheadrightarrow \mathbb{T}_\Sigma$.

Recall the homomorphism

$$\phi_\Sigma : R_\Sigma \twoheadrightarrow \mathbb{T}_\Sigma$$

between the universal deformation ring and the Hecke algebra, and the prime ideal

$$\mathfrak{p}_{\Sigma,f} = \ker(\pi_{\Sigma,f} \circ \phi_\Sigma : R_\Sigma \to \mathcal{O}).$$

Let $\rho_{f,\lambda} = \pi_{\Sigma,f} \circ \rho_\Sigma^{mod} = \pi_{\Sigma,f} \circ \phi_\Sigma \circ \rho_\Sigma^{univ}$. We observed before that the tangent space of the deformation problem $\mathcal{D}_\Sigma$ "along $\rho_{f,\lambda}$" is given by

$$Hom_{\mathcal{O}}(\mathfrak{p}_{\Sigma,f}/\mathfrak{p}_{\Sigma,f}^2, E/\mathcal{O}) \simeq H^1_{\mathcal{L}_\Sigma}(G_\mathbb{Q}, Ad^0\rho_{f,\lambda} \otimes_{\mathcal{O}} E/\mathcal{O}). \tag{4.1}$$

Define

$$\eta_{\Sigma,f} = \pi_{\Sigma,f}(\mathrm{Ann}_{\mathbb{T}_\Sigma}(\ker \pi_{\Sigma,f})).$$

This is called the *congruence ideal (of f).* To understand the terminology, suppose for simplicity that $\mathcal{O} = \mathbb{Z}_\ell$, that $\mathcal{N}_\Sigma$ consists of only two newforms f, g and that $n \geq 1$ is the highest power of ℓ such that $a_p(f) \equiv a_p(g) \mod \ell^n$, or equivalently, that $\rho_{f,\ell}$ and $\rho_{g,\ell}$ (in appropriate bases) are congruent modulo ℓ^n. Then

$$\mathbb{T}_\Sigma = \{(a,b) \in \mathbb{Z}_\ell^2 | \, a \equiv b \mod \ell^n\},$$

$\pi_{\Sigma,f}$ is the projection on the first copy of $\mathbb{Z}_\ell$, its kernel is $0 \times \ell^n\mathbb{Z}_\ell$, its annihilator is $\ell^n\mathbb{Z}_\ell \times 0$, and $\eta_{\Sigma,f} = (\ell^n)$.

As a more sophisticated example, suppose $\mathcal{N}_\Sigma = \{f_1, f_2, f_3, f_4\}$, all congruent modulo ℓ and no higher power, but suppose that in addition $f_1 + f_4 \equiv f_2 + f_3 \mod \ell^2$. If "no further congruences exist", then we might have

$$\mathbb{T}_\Sigma = \{(a,b,c,d) \in \mathbb{Z}_\ell^4 | \, a \equiv b \equiv c \equiv d \mod \ell, \; a + d \equiv b + c \mod \ell^2\}.$$

Check that this is a ring! Then $\eta_{\Sigma,f_1} = (\ell^2)$.

4.2.3. *The quantities c_p.* Suppose $\Sigma \subset \Sigma'$. For every $p \in \Sigma' - \Sigma$, we define canonical elements $c_p \in \mathbb{T}_\Sigma$. The importance of these elements is twofold, and serves to relate the change in R_Σ to the change in $\mathbb{T}_\Sigma$ when we enlarge Σ. As such, these elements become indispensable when Wiles boosts up his "$R = T$" theorem from the minimal case ($\Sigma = \emptyset$) to the general case.

On the one hand, $\pi_{\Sigma,f}(c_p)$ gives an *upper bound* for the growth of the tangent space of the deformation problem "along $\rho_{f,\lambda}$", when we relax the minimality condition at p. More precisely, these elements (for all $p \in \Sigma' - \Sigma$) control (from above) the difference between the Selmer groups $H^1_{\mathcal{L}_\Sigma}(G_\mathbb{Q}, Ad^0\rho_{f,\lambda} \otimes E/\mathcal{O})$ and the same group with Σ' replacing Σ.

On the other hand, $\pi_{\Sigma,f}(c_p)$ gives a *lower bound* for the growth of the congruence ideal $\eta_{\Sigma,f}$, when we change Σ to Σ'.

Taken together, we shall deduce that the change in the Selmer group is bounded above by the change in the congruence ideal. A fairly general commutative algebra criterion will allow then to propagate the "$R = T$" result from Σ to Σ'.

We shall define the c_p now, and establish the relation to the tangent space of the deformation problem. This is relatively easy, and boils down to calculations in local cohomology groups, made possible by the relation between $\rho_{f,\lambda}|_{G_p}$ and $a_p(f)$. [We use this relation at ramified primes $p \neq \ell$ and at ℓ as well!]

The relation between the same c_p's and the congruence ideal is more subtle, and is the topic of the second deep result we shall need about the Hecke algebra.

Finally we stress that for proving the main theorem in the minimal case only, this whole section is unnecessary.

Define:

- If $p \neq \ell$ and $\overline{\rho}$ is unramified at p, then $c_p = (p-1)(T_p^2 - (p+1)^2)$.
- If $p \neq \ell$ and $\overline{\rho}$ is ramified at p, then $c_p = (p^2 - 1)$.
- If $\overline{\rho}$ is flat and ordinary at ℓ, let $c_\ell = T_\ell^2 - (\ell+1)^2$.
- If $\overline{\rho}$ is either non-flat or non-ordinary, $c_\ell = 1$.

Fix $f \in \mathcal{N}_\Sigma$, and let $\rho = \rho_{f,\lambda}$ for brevity. If $\ell \neq p \in \Sigma' - \Sigma$ let

$$H_p = H^1(G_p, Ad^0\rho \otimes E/\mathcal{O})/H^1(G_p/I_p, (Ad^0\rho \otimes E/\mathcal{O})^{I_p}).$$

If $\ell \in \Sigma' - \Sigma$ (recall that then $\overline{\rho}$ is both flat and ordinary)

$$H_\ell = H^1_{ss}(G_\ell, Ad^0\rho \otimes E/\mathcal{O})/H^1_f(G_\ell, Ad^0\rho \otimes E/\mathcal{O}).$$

Lemma 63. *The groups H_p and H_ℓ are finite,*

$$\#H_p = \#\mathcal{O}/\pi_{\Sigma,f}(c_p),$$

and

$$\#H_\ell = \#\mathcal{O}/\pi_{\Sigma,f}(c_\ell).$$

Proof. Since, by our convention, only unramified $p \neq \ell$ may appear in Σ', and if $\ell \in \Sigma'$ then $\overline{\rho}$ is both flat and ordinary, we shall give the proof only in these cases, although it is valid also for ramified p (and holds trivially at ℓ if it is either non-flat or non-ordinary). Observe that we may take $c_\ell = (\ell-1)(T_\ell^2 - (\ell+1)^2)$, just like c_p, because $\ell - 1$ is a unit.

Assume $p \neq \ell$ and $\overline{\rho}$ is unramified at p. Since $p \notin \Sigma$ and $\rho_{f,\lambda}$ is of type $\mathcal{D}_\Sigma$, it is also unramified at p (in the ramified case, the minimality condition $p \notin \Sigma$ would mean that it stays "type A" at p, but as agreed above, we shall not need to relax this condition; practically, because in the application to semistable elliptic curves this situation will occur only if the elliptic curve had multiplicative reduction at p, and then the ℓ-adic representation stays "type A".) Let α, β be the two eigenvalues of $\rho_{f,\lambda}(\sigma_p)$. We first note that $H^1(G_p/I_p, (Ad^0\rho \otimes \lambda^{-n}/\mathcal{O})^{I_p})$ has the same cardinality as $H^0(G_p, Ad^0\rho \otimes \lambda^{-n}/\mathcal{O})$, so by the Euler characteristic formula $H_p^{(n)}$ (where $E/\mathcal{O}$ is replaced by $\lambda^{-n}/\mathcal{O}$) has the same cardinality as

$$H^2(G_p, Ad^0\rho \otimes \lambda^{-n}/\mathcal{O}),$$

or, by Tate local duality, of $H^0(G_p, Ad^0\rho \otimes \lambda^{-n}/\mathcal{O}(1))$. To prove that this cardinality is bounded in n we observe that $p\alpha/\beta, p$ and $p\beta/\alpha$ are all different from 1. Indeed, if not, since $\alpha\beta = p$, we must have $\{\alpha, \beta\} = \pm\{1, p\}$. This would violate, however, Hasse's theorem that α and β are p-Weil numbers. The same argument shows that $c_p \neq 0$. It is now an easy matter to show (writing

$$c_{p,f} = \pi_{\Sigma,f}(c_p) = (p-1)(a_p(f)^2 - (p+1)^2)$$

that

$$\#H^0(G_p, Ad^0\rho \otimes \lambda^{-n}/\mathcal{O}(1)) = \#\mathcal{O}/(\lambda^n, c_{p,f}).$$

It boils down to the identity

$$(p-1)(p\alpha\beta^{-1} - 1)(p\beta\alpha^{-1} - 1) = (p-1)((\alpha+\beta)^2 - (p+1)^2).$$

The computation at ℓ is identical, because Proposition 47(i),(ii) gives

$$\#H_\ell^{(n)} = \#\mathcal{O}/(\lambda^n, \chi_2\chi_1^{-1}(\sigma_\ell) - 1))$$

where χ_i are the unramified characters figuring in

$$\rho|G_\ell \sim \begin{pmatrix} \epsilon\chi_1 & * \\ & \chi_2 \end{pmatrix}.$$

Note that $\chi_1 = \chi_2^{-1}$. As we have seen before in the flat and ordinary case ($\delta = 0$ in previous notations), $a_\ell(f) \in \mathcal{O}^\times$, and the work of Deligne and Rapoport implies that $u = \chi_2(\sigma_\ell)$ is the *unit root* (in $\mathcal{O}$) of

$$X^2 - a_\ell(f)X + \ell = 0.$$

The lemma therefore boils down, if $\ell \in \Sigma' - \Sigma$, to the identity

$$a_\ell(f)^2 - (\ell+1)^2 = (u+\ell u^{-1})^2 - (\ell+1)^2 = (u^2-1)(1-u^{-2}\ell^2) \sim u^2 - 1 = \chi_2\chi_1^{-1}(\sigma_\ell) - 1.$$

Note that, as in the case of $p \neq \ell$, $u^2 \neq 1$, because $a_\ell(f) \neq \pm(\ell+1)$. This is because $\ell \notin \Sigma$ is a prime of good reduction for f, where $\rho_{f,\nu}$ for some prime ν *not* above ℓ is unramified, so the roots of $X^2 - a_\ell(f)X + \ell$ are ℓ-Weil numbers and can not be $\pm\{1, \ell\}$. □

Corollary 64. *There is an exact sequence*

$$0 \to H^1_{\mathcal{L}_\Sigma}(G_\mathbb{Q}, Ad^0\rho_{f,\lambda} \otimes E/\mathcal{O}) \to H^1_{\mathcal{L}_{\Sigma'}}(G_\mathbb{Q}, Ad^0\rho_{f,\lambda} \otimes E/\mathcal{O}) \to \prod_{p \in \Sigma' - \Sigma} H_p.$$

We have

$$\#\left(H^1_{\mathcal{L}_{\Sigma'}}(G_\mathbb{Q}, Ad^0\rho_{f,\lambda} \otimes E/\mathcal{O}) / H^1_{\mathcal{L}_\Sigma}(G_\mathbb{Q}, Ad^0\rho_{f,\lambda} \otimes E/\mathcal{O})\right) \leq \#\mathcal{O}/(\prod_{p \in \Sigma' - \Sigma} c_{p,f})$$

with equality if ad only if the sequence is also exact on the right.

We now state the second theorem on the structure of the Hecke algebras that we shall prove later.

Theorem 65. *Let $\Sigma \subset \Sigma'$ be finite sets of primes such that if $\ell \neq p \in \Sigma'$ then $\overline{\rho}$ is unramified at p, and if $\ell \in \Sigma'$ then $\overline{\rho}$ is flat and ordinary at ℓ. Let $f \in \mathcal{N}_\Sigma$. Then*

$$\eta_{\Sigma',f} \subset \eta_{\Sigma,f} \cdot (\prod_{p \in \Sigma' - \Sigma} c_{p,f}).$$

Corollary 66. *With the above notation,*

$$\#\left(H^1_{\mathcal{L}_{\Sigma'}}(G_\mathbb{Q}, Ad^0\rho_{f,\lambda} \otimes E/\mathcal{O}) / H^1_{\mathcal{L}_\Sigma}(G_\mathbb{Q}, Ad^0\rho_{f,\lambda} \otimes E/\mathcal{O})\right) \leq \#(\eta_{\Sigma,f}/\eta_{\Sigma',f}).$$

4.3. **The two commutative algebra criteria (week 10).**

4.3.1. *Local complete intersections and the first criterion.* Let $A \in \mathcal{C}_{\mathcal{O}}$ be finite free as an $\mathcal{O}$-module. A is called a local complete intersection (l.c.i.) if

$$A \simeq \mathcal{O}[[X_1, \ldots, X_r]]/(f_1, \ldots, f_r)$$

(same number of variables and relations). This notion is clearly invariant under finite base change $\mathcal{O}'/\mathcal{O}$, but it is also true that if $A \otimes_{\mathcal{O}} \mathcal{O}'$ is a l.c.i., so is A. We shall study l.c.i.'s and their relation to singualrity types (Cohen Macaulay-ness and Gorenstein-ness) in the later chapter on commutative algebra, and also see some examples.

The following commutative algebra criterion will serve to pass from the proof of the Main Theorem in the minimal case, to a proof in the general case.

Theorem 67 (Wiles' first criterion)**.** *Suppose that*

$$\phi : R \twoheadrightarrow T$$

is a surjection of $\mathcal{O}$-algebras in $\mathcal{C}_{\mathcal{O}}$. Suppose also that T is finite free as an $\mathcal{O}$-module and is equipped with a homomorphism $\pi : T \twoheadrightarrow \mathcal{O}$. Let $\mathfrak{p} = \ker(\pi \circ \phi)$, so that $R \simeq \mathcal{O} \oplus \mathfrak{p}$ as an $\mathcal{O}$-module. Let

$$\eta = \pi(\mathrm{Ann}_T(\ker \pi)) \subset \mathcal{O},$$

and suppose that $\eta \neq 0$. Then the following are equivalent:
(i) $\phi : R \simeq T$ and these rings are l.c.i.
(ii) $\#\mathfrak{p}/\mathfrak{p}^2 = \#\mathcal{O}/\eta$.
(iii) $\#\mathfrak{p}/\mathfrak{p}^2 \leq \#\mathcal{O}/\eta$.

A-priori, we do not assume that R is finite over $\mathcal{O}$, nor that it is $\mathcal{O}$-torsion free.

4.3.2. *J-structures and the second criterion.* Fix an integer $r \geq 1$ (in practice, $r = \#Q$ for a set of Taylor-Wiles primes). Let $J \lhd \mathcal{O}[[S_1, \ldots, S_r]]$ be an ideal contained in $(S_1, \ldots, S_r)$ (in practice, the ideal generated by $(1+S_i)^{\ell^n} - 1$ for some large n, and $\mathcal{O}[[S_1, \ldots, S_r]]/J$ will be a quotient of the rings of diamond operators $\mathcal{O}[\Delta_Q]$ if $q \equiv 1 \mod \ell^n$ for all $q \in Q$). By a *J-structure* for the surjection $R \twoheadrightarrow T$ in $\mathcal{C}_{\mathcal{O}}$ we mean a commutative diagram

$$\begin{array}{ccccc} & & \mathcal{O}[[S_1, \ldots, S_r]] & & \\ & & \downarrow & \searrow & \\ \mathcal{O}[[X_1, \ldots X_r]] & \twoheadrightarrow & R' & \twoheadrightarrow & T' \\ & & \downarrow & & \downarrow \\ & & R & \twoheadrightarrow & T \end{array}$$

in $\mathcal{C}_{\mathcal{O}}$ satisfying:

- T' is finite and free as an $\mathcal{O}$-module,
- $T'/(S_1, \ldots, S_r)T' = T$ and $R'/(S_1, \ldots, S_r)R' = R$,
- For any ideal $I \supset J$, the map from $\mathcal{O}[[S_1, \ldots, S_r]]/I$ to T'/IT' is injective.

We make a few remarks concerning the definition. First, there is no relation between the S_i and the X_i, and the X_i do not figure out in the properties of the J-structure, except for the fact that R' can be generated as an $\mathcal{O}$-algebra by r variables. Instead of specifiying the homomorphism from $\mathcal{O}[[X_1, \ldots X_r]]$ we may simply say that the k-dimension of the reduced cotangent space $\mathfrak{m}_{R'}/(\mathfrak{m}_{R'}^2, \lambda)$ is $\leq r$, where r is the number of S_i. The second remark is that we may replace R' and T' by R'/JR' and T'/JT', so without loss of generality we may assume that J is the kernel of both homomorphisms $\mathcal{O}[[S_1,\ldots,S_r]] \to T'$ and $\mathcal{O}[[S_1,\ldots,S_r]] \to R'$. Finally, for any ideal $J' \supset J$, a J-structure is clearly also a J'-structure.

As in the first criterion, we do not know a-priori that R, let alone R', is finite or torsion-free over $\mathcal{O}$. These assumptions are only made on T and T'. The following commutative algebra criterion of Taylor-Wiles, slightly improved by Faltings, will serve to prove the Main Theorem in the minimal case.

Theorem 68 (Faltings-Taylor-Wiles second criterion)**.** *Suppose that there exists a sequence of ideals $J_n \lhd \mathcal{O}[[S_1,\ldots,S_r]]$ such that $J_0 = (S_1,\ldots,S_r)$, $J_n \supset J_{n+1}$ and $\bigcap J_n = 0$. Suppose that for each n there exists a J_n-structure for $R \twoheadrightarrow T$. Then $R \simeq T$ and both are l.c.i.*

4.4. **The proof of the Main Theorem.**

Theorem 69 (Main Theorem "$R = T$")**.** *Let $\overline{\rho}$ be as in §3.1.1 and assume, in addition, that $\overline{\rho}$ is modular. Let Σ be a finite set of finite primes such that, if $\ell \neq p \in \Sigma$ then $\overline{\rho}$ is unramified at p, and if $\ell \in \Sigma$ then $\overline{\rho}$ is flat and ordinary at ℓ. Let R_Σ be the universal deformation ring of type $\mathcal{D}_\Sigma$ and $\mathbb{T}_\Sigma$ the Hecke algebra constructed in §4.1.3. Let*

$$\phi_\Sigma : R_\Sigma \twoheadrightarrow \mathbb{T}_\Sigma$$

be the surjective homomorphism bringing ρ_Σ^{univ} to ρ_Σ^{mod}. Then ϕ_Σ is an isomorphism, and $R_\Sigma \simeq \mathbb{T}_\Sigma$ is a l.c.i..

Corollary 70. *Let E be an elliptic curve defined over $\mathbb{Q}$. Assume that E is everywhere semistable and that $\overline{\rho}_{E,3}$ is irreducible. Then E is modular.*

Proof. Let $\ell = 3$, $k = \mathbb{F}_3$, $\mathcal{O} = \mathbb{Z}_3$ and observe that $\overline{\rho} = \overline{\rho}_{E,3}$ satisfies the running assumptions: it is odd and irreducible, $\det \overline{\rho} = \overline{\epsilon}$, it is "type A" at the $p \neq \ell$ where it is ramified (thanks to the assumption that E has multiplicative reduction), and is flat if E has good reduction at ℓ, or ordinary if E has multiplicative reduction there. By Theorem 59 $\overline{\rho}$ is modular of weight 2, level $\ell^{\delta(\overline{\rho})}N(\overline{\rho})$ and trivial nebentypus, where $\delta(\overline{\rho}) = 0$ if $\overline{\rho}$ is flat at ℓ and $= 1$ if it is ordinary non-flat.

Let Σ be the set of primes p where $\overline{\rho}$ is unramified but $\rho = \rho_{E,3}$ is ramified (if $p \neq \ell$), as well as $\ell = 3$ if $\overline{\rho}$ is flat there but ρ is not (i.e. E has bad reduction at 3). Enlarge $\mathcal{O}$ and k to contain all the $\mathcal{O}_{f,\lambda_f}$ for $(f,\lambda_f,\iota_f) \in \mathcal{N}_\Sigma$ as before. By the semi-stability assumption ρ is a deformation of type $\mathcal{D}_\Sigma$ of $\overline{\rho}$, so there exists a unique homomorphism $\pi : R_\Sigma \to \mathcal{O}$ bringing ρ_Σ^{univ} to ρ. The homomorphism $\pi \circ \phi_\Sigma^{-1} : \mathbb{T}_\Sigma \to \mathcal{O}$ corresponds to an $(f,\lambda_f,\iota_f) \in \mathcal{N}_\Sigma$ such that

$$\rho \simeq \iota_f \circ \rho_{f,\lambda_f},$$

as desired. $\square$

We now prove the main theorem.

Proof. Assume first that $\Sigma = \emptyset$. According to Proposition 39, since $\overline{\rho}$ is modular, it satisfies condition (L), namely $\overline{\rho}|G_L$ is absolutely irreducible, where $L = \mathbb{Q}(\sqrt{(-1)^{(\ell-1)/2}\ell})$. Let

$$r = \dim H^1_{\mathcal{L}_\emptyset^*}(G_\mathbb{Q}, Ad^0\overline{\rho}(1)).$$

By Theorem 57, for each $n \geq 1$ there exists a set Q_n of r "Taylor-Wiles primes" $q \equiv 1 \mod \ell^n$, and R_{Q_n} is topologically generated as an $\mathcal{O}$-algebra by r elements, i.e. is a quotient of $\mathcal{O}[[X_1,\ldots,X_r]]$. Let n_i be the highest power of ℓ dividing q_i-1, so that $n_i \geq n$. Fix an isomorphism as before

$$\mathcal{O}[\Delta_{Q_n}] \simeq \mathcal{O}[[S_1,\ldots,S_r]]/(\ldots,(1+S_i)^{\ell^{n_i}}-1,\ldots)$$

by mapping $1+S_i$ to a generator of the ℓ-Sylow subgroup of $\Delta_{q_i} \simeq (\mathbb{Z}/q_i\mathbb{Z})^\times$. Then for each $n \geq 1$ we get a diagram

$$\begin{array}{ccccc} & & \mathcal{O}[[S_1,\ldots,S_r]] & & \\ & & \downarrow & \searrow & \\ \mathcal{O}[[X_1,\ldots X_r]] & \twoheadrightarrow & R_{Q_n} & \twoheadrightarrow & \mathbb{T}_{Q_n} \\ & & \downarrow & & \downarrow \\ & & R_\emptyset & \twoheadrightarrow & \mathbb{T}_\emptyset \end{array}$$

where:

- $R_{Q_n}/(S_1,\ldots,S_r)R_{Q_n} = R_{Q_n}/\mathfrak{a}_{Q_n}R_{Q_n} = R_\emptyset$,
- $\mathbb{T}_{Q_n}$ is finite free over $\mathcal{O}[\Delta_{Q_n}]$ (see Theorem 61),
- $\mathbb{T}_{Q_n}/(S_1,\ldots,S_r)\mathbb{T}_{Q_n} = \mathbb{T}_{Q_n}/\mathfrak{a}_{Q_n}\mathbb{T}_{Q_n} = \mathbb{T}_\emptyset$ (see Corollary 62).

We conclude that $\phi_\emptyset : R_\emptyset \twoheadrightarrow \mathbb{T}_\emptyset$ admits a J_n-structure, where $J_n = (\ldots,(1+S_i)^{\ell^n}-1,\ldots)$. Indeed, letting

$$J_n' = (\ldots,(1+S_i)^{\ell^{n_i}}-1,\ldots),$$

$\mathbb{T}_{Q_n}$ is finite free as a module over $\mathcal{O}[[S_1,\ldots,S_r]]/J_n'$, so, all the more so, is finite free as an $\mathcal{O}$-module. In addition, for any ideal $I \supset J_n \supset J_n'$ we get that $\mathbb{T}_{Q_n}/I\mathbb{T}_{Q_n}$ is free over $\mathcal{O}[[S_1,\ldots,S_r]]/I$, so the latter injects into the first. The statement of the theorem follows now from the second commutative-algebra criterion, Theorem 68.

Next, we assume the theorem is proved for $\Sigma=\emptyset$, and prove the general case. Let $f \in \mathcal{N}_\emptyset$, and consider the homomorphism $\pi_f : \mathbb{T}_\emptyset \to \mathcal{O}$. From the fact that $\mathbb{T}_\Sigma \otimes_{\mathcal{O}} E \simeq E^n$ it follows easily that we always have

$$\eta_\Sigma \neq 0,$$

so the first commutative-algebra criterion (Theorem 67) applies. Let $\mathfrak{p} = \ker(\pi_f \circ \phi)$ where $\phi : R_\emptyset \simeq \mathbb{T}_\emptyset$. Since we proved that $R_\emptyset \simeq \mathbb{T}_\emptyset$ are l.c.i., we know that

$$\#\mathfrak{p}/\mathfrak{p}^2 = \#\mathcal{O}/\eta_\emptyset.$$

In particular, $\mathfrak{p}/\mathfrak{p}^2$ is a finite group and not only a finite $\mathcal{O}$-module, something that is not a-priori clear at all. By Corollary 66 and formula 4.1 we obtain that

$$\#\mathfrak{p}_\Sigma/\mathfrak{p}_\Sigma^2 \leq \#\mathcal{O}/\eta_\Sigma < \infty$$

where η_Σ and $\mathfrak{p}_\Sigma$ refer now to the same f, but to the rings $\mathbb{T}_\Sigma$ and R_Σ. A second application of Theorem 67 shows that $\phi_\Sigma : R_\Sigma \simeq \mathbb{T}_\Sigma$ and the rings are l.c.i.. □

4.5. **The 3-5 trick.** The proof of modularity of semistable elliptic curves relied, so far, on the irreducibility of $\overline{\rho}_{E,3}$. To conclude the proof Wiles used a trick that became known as the “3-5” trick.

Theorem 71. *Let E be a semistable elliptic curve defined over $\mathbb{Q}$. Then E is modular.*

Proof. If $\overline{\rho}_{E,3}$ is irreducible then this follows from Corollary 70. Suppose $\overline{\rho}_{E,3}$ was reducible. We claim that $\overline{\rho}_{E,5}$ is then irreducible. For otherwise, E would have a rational subgroup of order 15 defined over $\mathbb{Q}$, and would give rise to a non-cuspidal rational point of $X_0(15)$. This curve is of genus 1, and is known to have only 4 non-cuspidal rational points, which do not correspond to semi-stable elliptic curves (and in any case correspond to modular elliptic curves).

We now prove that $\overline{\rho} = \overline{\rho}_{E,5}$ is modular. Consider the modular curve $X(\overline{\rho})$ parametrizing generalized elliptic curves A with $A[5] \simeq E[5]$ (as finite flat group schemes over $\mathbb{Q}$, i.e. as Galois modules), compatible with the Weil pairing. It is a twisted form of a certain connected component of $X(5)$, which is known to have genus 0. (Recall that $X(5)$ classifies $A[5] \simeq \mu_5 \times \mathbb{Z}/5\mathbb{Z}$.) Since it has a rational point (correspodning to E), it is isomorphic to $\mathbb{P}^1_{\mathbb{Q}}$. Consider the modular curve $X'(\overline{\rho})$ which is the covering of $X(\overline{\rho})$ classifying, in addition, a rational subgroup of order 3. It is a twisted form of $X_{\Gamma(5)\cap\Gamma_0(3)}$, which has genus greater than 1, so by Mordell's conjecture (Faltings' theorem) has finitely many $\mathbb{Q}$-rational points. Let $x \in X(\overline{\rho})(\mathbb{Q})$ be a rational point above which there does not lie any rational point of $X'(\overline{\rho})$. Let A be the elliptic curve represented by x. By definition, $A[5] \simeq E[5]$, and A does not admit a rational subgroup of order 3, so $\overline{\rho}_{A,3}$ is irreducible. For a prime $q \neq 5$ (including $q = 3$, if needed), $\overline{\rho}_{A,5} \simeq \overline{\rho}_{E,5}$. It follows from the lemma below that A too is semistable at q. If we choose x, in addition, close to the point representing E in the 5-adic topology, we can guarantee that A is also semistable at 5. We can now apply the main theorem to A, and the prime $\ell = 3$, to conclude that A is modular. However, this shows that $\overline{\rho}_{A,5} = \overline{\rho}_{E,5}$ is modular.

We now know that $\overline{\rho}_{E,5}$ is both irreducible and modular. Applying the main theorem to E and $\ell = 5$, concludes the proof. □

Lemma 72. *Let A and E be two elliptic curves over $\mathbb{Q}_q$ ($q \neq 5$), such that $A[5]$ and $E[5]$ are isomorphic as G_q-modules. If E is semistable, so is A.*

Proof. We use the fact that an elliptic curve over $\mathbb{Q}_q$ is semistable if and only if for every $\tau \in I_q$

$$(\rho_{E,5}(\tau) - I)^2 = 0,$$

if and only if 1 is the only eigenvalue of $\rho_{E,5}(\tau)$. By our assumption, $(\rho_{A,5}(\tau) - I)^2 \in 5M_2(\mathbb{Z}_5)$, so for any eigenvalue α of $\rho_{A,5}(\tau)$, $(\alpha-1)^2 \equiv 0 \mod 5\overline{\mathbb{Z}}_5$. However, by potential semistability, α is a root of unity. Let m be its exact order, and assume that $m > 1$. If m is not a power of 5, then $\alpha - 1$ is a 5-adic unit. If m is a power of 5, then $v_5(\alpha-1) = 1/n$ for $n = [\mathbb{Q}(\alpha) : \mathbb{Q}] \geq 4$, so we can not have $v_5((\alpha-1)^2) \geq 1$. Therefore $m = 1$ and $\alpha = 1$.

Note: The argument works with any prime $r \geq 5$ different from q replacing 5. The same argument, with any prime $r \geq 3$ different from q replacing 5, shows that if E has good reduction, so does A (instead of arguing on $(\rho_{A,r}(\tau) - I)^2 \equiv 0 \mod r$, argue, more simply, on $\rho_{A,r}(\tau) - I \equiv 0 \mod r$.) □

5. Complements on the Hecke algebra (weeks 11,12)

5.1. The geometry behind $\mathbb{T}_Q$.

5.1.1. *Passing to the full Hecke algebra.* Our construction of the Hecke algebra $\mathbb{T}_\Sigma$ was *representation-theoretic*, and included only the Hecke operators at good primes. Let us recall it. We determined the collection $\mathcal{N}_\Sigma$ of weight 2 newforms that give rise to deformations $\rho : G_{\mathbb{Q}} \to GL_2(\mathcal{O})$ of type Σ by looking at the prime-to-ℓ Artin conductors of such deformations. We noted that the power of any $p \neq \ell$ in it, under our assumptions on $\overline{\rho}$ and Σ, is equal to 1 if $\overline{\rho}$ was ramified at p, or bounded by 2, if $p \in \Sigma$ (in which case $\overline{\rho}$ was unramified at p, but ρ was allowed to be ramified there). At ℓ, we let $\delta_\Sigma(\overline{\rho}) = 0$ if $\overline{\rho}$ was flat at ℓ and $\ell \notin \Sigma$, and $\delta_\Sigma(\overline{\rho}) = 1$ otherwise (in which case ρ could be ordinary but non-flat). This gave us, in view of Carayol's theorem "arithmetic conductor = analytic conductor", the level

$$N_\Sigma = \ell^{\delta_\Sigma(\overline{\rho})} N(\overline{\rho}) \prod_{\ell \neq p \in \Sigma} p^2.$$

We took $(f, \lambda_f, \iota_f) \in \mathcal{N}_\Sigma$ if and only if f was a newform of weight 2, trivial nebentypus, and level $N_f | N_\Sigma$, such that $\iota_f \circ \overline{\rho}_{f,\lambda_f} = \overline{\rho}$.

Enlarging $\mathcal{O}$, if necessary, to contain all the images $\iota_f(\mathcal{O}_{f,\lambda_f})$, the Hecke algebra $\mathbb{T}_\Sigma$ was defined to be the $\mathcal{O}$-subalgebra of $\widetilde{\mathbb{T}}_\Sigma = \prod_{f \in \mathcal{N}_\Sigma} \mathcal{O}$ generated by the images $T_p \mapsto (..., a_p(f), ...)$ for good p ($p \nmid \ell N_\Sigma$). Note that since we did not include Hecke operators for primes dividing N_Σ, the question of oldforms did not arise; every $f \in \mathcal{N}_f$ was an eigenform of all the good T_p. In fact, we could omit any finite set of p's from the list of good primes, and still obtain exactly the same $\mathbb{T}_\Sigma$.

The ring $\mathbb{T}_\Sigma$ was automatically local, since for any good p and all $f \in \mathcal{N}_\Sigma$,

$$a_p(f) \equiv tr\overline{\rho}(\sigma_p) \mod \lambda,$$

independently of f. The ideal $\mathfrak{M}_\Sigma \subset \mathbb{T}_\Sigma$ consisting of vectors all of whose f-coordinates are divisible by λ is therefore maximal, with residue field k. To show that any element of $\mathbb{T}_\Sigma$ outside $\mathfrak{M}_\Sigma$ is invertible, it is enough to show that any element of the form $1 - x$, with $x \in \mathfrak{M}_\Sigma$, is invertible. But this follows from the formula $(1-x)^{-1} = 1 + x + x^2 + \cdots$, valid in $\widetilde{\mathbb{T}}_\Sigma$, and the fact that $\mathbb{T}_\Sigma$ is closed in $\widetilde{\mathbb{T}}_\Sigma$.

Alternatively, we may look at the ring of endomorphisms $\mathbb{T}_\mathcal{O} = \mathfrak{T}(S_2(\Gamma_0(N_\Sigma))) \otimes \mathcal{O}$ which is the image of the *full* abstract Hecke algebra $\mathfrak{T} = \mathbb{Z}[..., T_p, ...]$ (all p's !) in $\mathrm{End}(S_2(\Gamma_0(N_\Sigma)))$, base changed to $\mathcal{O}$. If $p | N_\Sigma$ we denote by T_p (at level N_Σ) the Atkin-Lehner operator U_p, as usual. This $\mathbb{T}_\mathcal{O}$ is a complete semi-local ring, the direct product of its localizations at maximal ideals. However, because of the existence of oldforms in $S_2(\Gamma_0(N_\Sigma))$, it may not be reduced.

Proposition 73. *There exists a maximal ideal* $\mathfrak{m}$ *of* $\mathbb{T}_\mathcal{O}$ *and an isomorphism between* $\mathbb{T}_\Sigma$ *and the localization* $\mathbb{T}_\mathfrak{m}$ *of* $\mathbb{T}_\mathcal{O}$ *at* $\mathfrak{m}$, *sending "T_p to T_p" for* $p \nmid \ell N_\Sigma$.

Proof. We refer to [D-D-T], Proposition 4.7 and the Lemmas preceding it, or to [W95], Proposition 2.15, for the construction of $\mathfrak{m}$ and the proof of the Proposition. It relies on the theory of old-forms, and the specifics of our situation, where the primes $p | N(\overline{\rho})$ must divide the level N_f of every newform $f \in \mathcal{N}_\Sigma$, ℓ appears to order ≤ 1 in N_Σ, and the primes $p \in \Sigma$ to order at most 2. The essential part of the proof is the surjectivity. For this one has to show that if $p | \ell N_\Sigma$ then the image of T_p in $\mathbb{T}_\mathfrak{m}$ is already in the image of $\mathbb{T}_\Sigma$, although the latter is generated only by the good Hecke operators. □

5.1.2. *Trading off the level and the nebentypus.* When $\Sigma = Q$ was a set of Taylor-Wiles primes, we analyzed the universal deformation ρ_Q^{univ} locally at $q \in Q$ and obtained that R_Q, hence also $\mathbb{T}_Q$, acquired a structure of an $\mathcal{O}[\Delta_Q]$-algebra, and the Δ_Q-coinvariants were $R_\emptyset$ and $\mathbb{T}_\emptyset$. We called the Δ_Q "diamond operators", but their origin was not geometric, and as all our forms had trivial nebentypus, this needs justification and explanation.

To get the finer structure theorems on $\mathbb{T}_Q$ we give another, more geometric, construction of it, which also justifies the name "diamond operators". There is a trade-off between the level and the nebentypus. *We shall sacrifice the "trivial nebentypus" assumption (or the cyclotomic determinant condition on the representation) and gain that the level becomes square free.* In other words, let

$$M_Q = \prod_{p|N_Q} p$$

be the radical of

$$N_Q = \ell^{\delta(\overline{\rho})} N(\overline{\rho}) \prod_{q \in Q} q^2.$$

Then

$$(\mathbb{Z}/M_Q\mathbb{Z})^\times \twoheadrightarrow \prod_{q \in Q} (\mathbb{Z}/q\mathbb{Z})^\times \twoheadrightarrow \Delta_Q$$

where Δ_Q is the ℓ-Sylow subgroup of $\prod_{q \in Q}(\mathbb{Z}/q\mathbb{Z})^\times$. Let

$$\Gamma_1(M_Q) \subset \Gamma_Q \subset \Gamma_0(M_Q)$$

be the subgroup for which $\Gamma_0(M_Q)/\Gamma_Q$ is Δ_Q. Consider $\mathfrak{T}(S_2(\Gamma_Q)) \otimes \mathcal{O}$ where now we include in the Hecke algebra, besides *all* the T_p, also the diamond operators $\langle \delta \rangle$ for $\delta \in \Delta_Q$.

If f is a newform of weight 2 and trivial nebentypus, whose level N_f divides N_Q, giving rise to a representation of type $\mathcal{D}_Q$ (i.e. if $f \in \mathcal{N}_Q$), we saw that for $q \in Q$

$$\rho_{f,\lambda}|G_q \sim \begin{pmatrix} \epsilon\xi_q & 0 \\ 0 & \xi_q^{-1} \end{pmatrix}$$

where $\xi_q : G_q \to \mathcal{O}^\times$ is a (ramified, in general) character such that $\theta_q = \xi_q|I_q$ factors through $I_q \twoheadrightarrow \Delta_q$. Let $\theta_Q : \Delta_Q \to \mathcal{O}^\times$ be the product of the θ_q and view it also as a Dirichlet character, and as a Galois character of $G_\mathbb{Q}$ by composition with the cyclotomic character. As a Galois character, it is tamely ramified at every $q \in Q$. Consider the newform

$$f' = f \otimes \theta_Q$$

whose Fourier coefficients $a_p(f') = a_p(f)\theta_Q(p)$ $(p \nmid Q)$. It has weight 2 and nebentypus $\chi_Q = \theta_Q^2$. From now on we write $\theta_Q = \chi_Q^{1/2}$ since the square root (in a cyclic ℓ-group) is uniquely defined.

The Galois representation associated to f' is $\rho_{f',\lambda} = \rho_{f,\lambda} \otimes \chi_Q^{1/2}$, so that

$$\rho_{f',\lambda}|G_q \sim \begin{pmatrix} \epsilon\xi_q\chi_Q^{1/2} & 0 \\ 0 & \xi_q^{-1}\chi_Q^{1/2} \end{pmatrix}.$$

Since $\xi_q^{-1}\chi_Q^{1/2}$ is *unramified* at q, the conductor of $\rho_{f',\lambda}$ is divisible by q to the first power only (the dimension of the I_q-coinvariants is 1 and not 2). Thus the level of f' divides M_Q and in fact $f' \in S_2(\Gamma_Q)$. Since $\chi_Q^{1/2}$ takes values in ℓ-power roots of unity, its image in $k^\times$ is trivial, so the residual representation is unchanged by twisting.

This procedure is reversible. Start with a newform f' in $S_2(\Gamma_Q)$ giving rise to a representation of the above shape at $q \in Q$, so that its nebentypus is χ_Q. Then $f = f' \otimes \chi_Q^{-1/2}$ has a trivial nebentypus, gives rise to a deformation of $\overline{\rho}$ of type $\mathcal{D}_Q$, and its level divides N_Q. It therefore lies in $\mathcal{N}_Q$.

Let $\mathbb{T}'_{\mathcal{O}} = \mathfrak{T}(S_2(\Gamma_Q), \mathcal{O})$ be the image of the full abstract Hecke algebra $\mathfrak{T} \otimes \mathcal{O}$ in $End(S_2(\Gamma_Q, \mathcal{O}))$. This is the same as the image of $\mathfrak{T} \otimes \mathcal{O}$ in $End(\mathcal{T}_\ell J_{\Gamma_Q}) \otimes_{\mathbb{Z}_\ell} \mathcal{O}$ (use the identification of $S_2(\Gamma)$ with the cotangent space at 0 of the Jacobian J_Γ). We have proved the following.

Proposition 74. *Suppose that $\overline{\rho}$ is modular of type $\mathcal{D}_\emptyset$, let Q be a set of Taylor-Wiles primes for $\overline{\rho}$, and let $\mathbb{T}_\mathfrak{m}$ be the local component of the full Hecke algebra $\mathbb{T}_{\mathcal{O}} = \mathfrak{T}(S_2(\Gamma_0(N_Q), \mathcal{O})$ associated with $\overline{\rho}$ and the set $\Sigma = Q$ as in the previous Proposition. Then there exists a maximal ideal $\mathfrak{m}'$ of the full Hecke algebra $\mathbb{T}'_{\mathcal{O}} = \mathfrak{T}(S_2(\Gamma_Q), \mathcal{O})$ and an isomorphism*

$$\mathbb{T}_\mathfrak{m} \simeq \mathbb{T}'_{\mathfrak{m}'}$$

carrying T_p on the left ($p \notin Q$) to $T_p \cdot \langle p \rangle^{-1/2}$ on the right.

The reason for the square root is that $a_p(f') = a_p(f)\chi_Q^{1/2}(p)$ while the nebentypus of f' is χ_Q. [Check: if $f \leftrightsquigarrow f'$ then $T_p f' = a_p(f')f' = a_p(f)\chi_Q^{1/2}(p)f'$, so

$$T_p \cdot \langle p \rangle^{-1/2} f' = a_p(f)f'.]$$

It can be checked that $\mathfrak{m}'$ is generated by the following. (The appearance of quantities from k means: substitute any lift to $\mathcal{O}$; since $\lambda \in \mathfrak{m}'$, it does not matter which lift we choose.)

- λ
- $T_p - tr(\overline{\rho}(\sigma_p))$ and $\langle p \rangle - 1$ for $p \nmid N_Q$ (including $p = \ell$ if $\delta(\overline{\rho}) = 0$),
- $U_p - \overline{\rho}_{I_p}(\sigma_p)$ for $p|N(\overline{\rho})$, $p \neq \ell$. Here $\overline{\rho}_{I_p}$ is the character of G_p/I_p on the rank 1 I_p-coinvariants (recall that the local representation is "type A"),
- $U_\ell - \overline{\rho}_{I_\ell}(\sigma_\ell)$ if $\ell|N(\overline{\rho})$ with the same convention as before (recall that if $\ell|N(\overline{\rho})$ then $\overline{\rho}$ is ordinary non-flat at ℓ),
- $U_q - \beta_q$ for $q \in Q$. Here β_q is the eigenvalue of $\overline{\rho}(\sigma_q)$ which was used to define the action of Δ_q on R_Q, hence the structure of a module over $\mathcal{O}[\Delta_Q]$.

The following Corollary, which we leave out as an exercise, follows easily from the discussion above.

Corollary 75. *Under the isomorphism constructed between $\mathbb{T}_Q \simeq \mathbb{T}_\mathfrak{m} \simeq \mathbb{T}'_{\mathfrak{m}'}$ the action of Δ_Q on $\mathbb{T}_Q$ gets translated to the standard diamond operators action on $\mathbb{T}'_{\mathfrak{m}'}$.*

From now on we forget the two steps taken in the two propositions, namely (1) including the Hecke operators for the bad primes, and (2) twisting to replace N_Q by M_Q, at the expense of allowing a "partial" Γ_1-level, via the action of the diamond operators Δ_Q. We write $\mathbb{T}_\mathfrak{m}$ for $\mathbb{T}'_{\mathfrak{m}'}$ and record the isomorphism

$$\mathbb{T}_\Sigma \simeq \mathbb{T}_\mathfrak{m}$$

resulting from the two propositions.

5.1.3. *The geometry of J_{Γ_Q} and the proof of Theorem 61.* Having given an alternative construction of $\mathbb{T}_Q$ that sheds light on the geometric origin of the diamond operators, we prove the freeness of $\mathbb{T}_Q$ over $\mathcal{O}[\Delta_Q]$. We follow the method of [T-W95], although an alternative approach, based on q-expansions, was suggested later by F. Diamond and is used in [D-D-T].

Write, for simplicity, $J_Q = J_{\Gamma_Q}$, and observe that it lies between $J_0(M_Q)$ and $J_1(M_Q)$. Similarly, $X_Q = X(\Gamma_Q)$, and Y_Q is the corresponding open modular curve.

Write $\widetilde{\Gamma}_Q = \Gamma_0(M_Q)$, $\widetilde{X}_Q = X_0(M_Q)$ etc. We let $\widetilde{\mathfrak{m}}$ be the maximal ideal of the Hecke algebra of $S_2(\Gamma_0(M_Q), \mathcal{O})$ which is the image of $\mathfrak{m}$ (the map being restriction of Hecke operators), and $\widetilde{\mathbb{T}}_{\mathfrak{m}}$ the corresponding localization.

When we change Q we add the subscript Q (or $\emptyset$ if Q is the empty set) to the maximal ideal and the Hecke algebra.

The following theorem is deep, and relies on work of Mazur and Tilouine. The corresponding theorem for the rational Tate module $V_\ell J_Q$ or the rational cohomology is easy, but the integral statement needed here invokes a multiplicity-one statement for mod-ℓ representations that appear in $J_Q[\ell]$, and is delicate. It relies, crucially, on the irreducibility of $\overline{\rho}$, which implies that the maximal ideal $\mathfrak{m}$ is "non-Eisenstein" in Mazur's language.

Theorem 76. *(i) The ℓ-adic Tate module $(\mathcal{T}_\ell J_Q)_{\mathfrak{m}}$ (completed at $\mathfrak{m}$) is free of rank 2 over $\mathbb{T}_{\mathfrak{m}}$.*

(ii) $H^1(X_Q, \mathcal{O})^{\pm}_{\mathfrak{m}} = H^1(Y_Q, \mathcal{O})^{\pm}_{\mathfrak{m}}$ ($\pm$ refers to complex conjugation) are equal and free of rank 1 each.

Similar statements hold for localizations at $\widetilde{\mathfrak{m}}$ of $\mathcal{T}_\ell \widetilde{J}_Q$, or the cohomologies of the modular curves $\widetilde{X}_Q$, as modules over $\widetilde{\mathbb{T}}_{\mathfrak{m}}$.

Recall that $\mathbb{T}_\emptyset = \widetilde{\mathbb{T}}_\emptyset$ is the localization of $\mathfrak{T}(S_2(\Gamma_0(N), \mathcal{O}))$ at $\mathfrak{m} = \mathfrak{m}_\emptyset$, where $N = \ell^{\delta(\overline{\rho})} N(\overline{\rho})$ (when Q is empty there is no difference between the tilde and non-tilde versions). While there is no map from $\mathfrak{T}(S_2(\Gamma_0(M_Q), \mathcal{O}))$ to $\mathfrak{T}(S_2(\Gamma_0(N), \mathcal{O}))$ (the Hecke operators U_q for $q \in Q$ do not preserve $S_2(\Gamma_0(N), \mathcal{O}) \subset S_2(\Gamma_0(M_Q), \mathcal{O})$), the next lemma shows that after localizing at $\mathfrak{m}$ such a map exists, and in fact is an isomorphism.

Lemma 77. *There exists an isomorphism $\widetilde{\mathbb{T}}_Q \simeq \widetilde{\mathbb{T}}_\emptyset$ mapping the Hecke operators T_p ($p \nmid NQ$) and U_p ($p|N$) in $\widetilde{\mathbb{T}}_Q$ to the corresponding operators in $\widetilde{\mathbb{T}}_\emptyset$.*

Proof. See [dS], Lemma 13. In the first step one uses the assumption that the two eigenvalues α_q and β_q of $\overline{\rho}(\sigma_q)$ are distinct, to show that $(\mathcal{T}_\ell \widetilde{J}_Q)_{\widetilde{\mathfrak{m}}_Q}$ is "Q-old". By this we mean that this direct summand of $\mathcal{T}_\ell \widetilde{J}_Q \otimes \mathcal{O}$ is contained in the Tate module of the Q-old subvariety of $\widetilde{J}_Q$, which is isogenous to a product of 2^r copies of $J_0(N)$.

One way to prove this statement is to compute the module of fusion between the Q-old and the Q-new parts of $\widetilde{J}_Q$. This computation, based on $\alpha_q \neq \beta_q$ tells us that

$$\widetilde{J}_Q^{old}[\widetilde{\mathfrak{m}}_Q] \cap \widetilde{J}_Q^{new}[\widetilde{\mathfrak{m}}_Q] = \{0\}.$$

However, it follows from Theorem 76 that $\dim_k \widetilde{J}_Q[\widetilde{\mathfrak{m}}_Q] = 2$, or that the multiplicity of $\overline{\rho}$ in it is 1. Since $\overline{\rho}$ appears in $\widetilde{J}_Q^{old}[\widetilde{\mathfrak{m}}_Q]$ it can not appear in $\widetilde{J}_Q^{new}[\widetilde{\mathfrak{m}}_Q]$. Therefore $\overline{\rho}$ is not a constituent (i.e. a subquotient as a Galois module) of $\widetilde{J}_Q^{new}[\ell^\infty]_{\widetilde{\mathfrak{m}}_Q}$, so this ℓ-divisible group, and its Tate module as well, are 0.

Assume, for simplicity, that $Q = \{q\}$ consists of a single prime. The Q-old Hecke algebra is then isomorphic to

$$\mathbb{T}_0(N)[u_q]/(u_q^2 - T_q u_q + \langle q \rangle\, q).$$

Since the roots of the quadratic polynomial are distinct modulo $\mathfrak{m} = \mathfrak{m}_\emptyset$, and since $U_q - \beta_q \in \widetilde{\mathfrak{m}}_Q$, Hensel's lemma shows that after we localize the Q-old Hecke algebra at $\widetilde{\mathfrak{m}}_Q$, we get $\mathbb{T}_\emptyset = T_0(N)_{\mathfrak{m}}$. $\square$

We can now prove Theorem 61. Since $H^1(Y_Q, \mathcal{O})^{\pm}_{\mathfrak{m}}$ is free of rank 1 over $\mathbb{T}_Q$, it is enough to show that it is free over $\mathcal{O}[\Delta_Q]$. We shall in fact prove the stronger claim that $H^1(Y_Q, \mathcal{O})^-$ is free over $\mathcal{O}[\Delta_Q]$. Assume, for simplicity, that $\widetilde{\Gamma}_Q$ had no elliptic elements, or more generally, that the orders of its elliptic elements are invertible in $\mathcal{O}$ (unfortunately, this might not be the case when $\ell = 3$). Then

$$H^1(Y_Q, \mathcal{O}) \simeq H^1(\Gamma_Q, \mathcal{O}) \simeq H^1(\widetilde{\Gamma}_Q, \mathcal{O}[\Delta_Q]).$$

The first isomorphism comes from the relation between singular cohomology of curves and group cohomology of their fundamental group. The second isomorphism stems from Shapiro's lemma. The Δ_Q action on the cohomology on the left gets translated to its action on the coefficients. The key point, now, is that $\widetilde{\Gamma}_Q$ is a *free* group, since it has no elliptic elements. The abelian group $Z^1(\widetilde{\Gamma}_Q, \mathcal{O}[\Delta_Q])$ of 1-cocycles with values in $\mathcal{O}[\Delta_Q]$ is therefore free over $\mathcal{O}[\Delta_Q]$. So are its $\pm$ parts (this needs to be checked). If we focus on $H^1(Y_Q, \mathcal{O})^-$ we need not worry about the coboundarys because $B^1(\widetilde{\Gamma}_Q, \mathcal{O}[\Delta_Q])$ all lie in the $+$ eigenspace for complex conjugation.

In case $\widetilde{\Gamma}_Q$ has elliptic elements we have to introduce an auxiliary Γ_1-level to get rid of them, and then descend. See [dS], Proposition 14, how this is done.

In any case, we deduce that $H^1(Y_Q, \mathcal{O})^-$, and with it $\mathbb{T}_Q$, is free over $\mathcal{O}[\Delta_Q]$. Moreover, Shapiro's lemma tells us that the Δ_Q-coinvariants, i.e. the module obtained after we divide by the augmentation ideal $\mathfrak{a}_Q$, is identified with

$$H^1(\widetilde{\Gamma}_Q, \mathcal{O})^- \simeq H^1(\widetilde{Y}_Q, \mathcal{O})^-.$$

When we localize at $\mathfrak{m}_Q$ we get

$$\mathbb{T}_Q/\mathfrak{a}_Q\mathbb{T}_Q \simeq \mathbb{T}_{\mathfrak{m}_Q}/\mathfrak{a}_Q\mathbb{T}_{\mathfrak{m}_Q} \simeq \widetilde{\mathbb{T}}_{\mathfrak{m}_Q} \simeq \mathbb{T}_{\emptyset}.$$

The last isomorphism is a consequence of the last Lemma. We therefore conclude that

$$\mathrm{rk}_{\mathcal{O}[\Delta_Q]}\mathbb{T}_Q = \mathrm{rk}_{\mathcal{O}}\mathbb{T}_{\emptyset},$$

which gives another proof of Corollary 62. (One may say that the previous proof was representation-theoretic, while the new one is based on the geometry of modular Jacobians.)

5.2. **Congruence ideals and Hecke algebras.** We turn our attention to the second major result about the Hecke algebra $\mathbb{T}_\Sigma$, Theorem 65. Recall the statement.

Theorem. *Let $\Sigma \subset \Sigma'$ be finite sets of primes such that if $\ell \neq p \in \Sigma'$ then $\overline{\rho}$ is unramified at p, and if $\ell \in \Sigma'$ then $\overline{\rho}$ is flat and ordinary at ℓ. Let $f \in \mathcal{N}_\Sigma$. Then*

$$\eta_{\Sigma',f} \subset \eta_{\Sigma,f} \cdot (\prod_{p \in \Sigma' - \Sigma} c_{p,f}).$$

It is enough to prove the theorem, of course, when $\Sigma' = \Sigma \cup \{p\}$, which we assume from now on. Recalling that

$$\mathbb{T}_\Sigma \subset \widetilde{\mathbb{T}}_\Sigma = \prod_{g \in \mathcal{N}_\Sigma} \mathcal{O},$$

and assuming that our f is the first "g" in $\mathcal{N}_\Sigma$, we let

$$\wp = \ker(\pi_{\Sigma,f}), \quad I = \mathrm{Ann}_{\mathbb{T}_\Sigma}(\wp)$$

so that $\eta_{\Sigma,f} = \pi_{\Sigma,f}(I)$. Clearly

$$\wp = \{(0, *, ..., *) \in \mathbb{T}_\Sigma\}, \ \ I = \{(*, 0, ..., 0) \in \mathbb{T}_\Sigma\}.$$

Note that

$$\mathcal{O}/\eta_{\Sigma,f} \simeq (\mathcal{O} \oplus \wp)/(\eta_{\Sigma,f} \oplus \wp) \simeq \mathbb{T}_\Sigma/(I \oplus \wp).$$

Regarding the ideal $\mathfrak{p}_{\Sigma,f} \subset R_\Sigma$, we have

$$\mathfrak{p}_{\Sigma,f} = \phi_\Sigma^{-1}(\wp).$$

Let $\mathbb{T}$ be the Hecke algebra generated by *all* the Hecke operators acting on $S_2(\Gamma_0(N_\Sigma), \mathcal{O})$ and $\mathfrak{m}$ its maximal ideal for which we constructed an isomorphism $T_\Sigma \simeq \mathbb{T}_\mathfrak{m}$ sending "T_p to T_p" for $p \nmid N_\Sigma$ (see Proposition 73). Let $W_\Sigma = (\mathcal{T}_\ell J_0(N_\Sigma) \otimes \mathcal{O})_\mathfrak{m}$, so that by Theorem 76, W_Σ is free of rank 2 over $\mathbb{T}_\Sigma$. Fix an isomorphism $\mathbb{Z}_\ell(1) \simeq \mathbb{Z}_\ell$. The Weil pairing, and the fact that the Hecke operators at Γ_0-level are self-adjoint, implies that there is an alternating pairing

$$\langle , \rangle_\Sigma : W_\Sigma \times W_\Sigma \to \mathcal{O},$$

inducing an isomorphism of $\mathbb{T}_\Sigma$-modules

$$W_\Sigma \simeq Hom_\mathcal{O}(W_\Sigma, \mathcal{O}).$$

Incidentally note that if we break W_Σ into its $\pm$-eigenspaces for complex conjugation, then Mazur's theorem 76 implies that each is free of rank 1 over $\mathbb{T}_\Sigma$, and the Weil pairing induces a duality between W_Σ^+ and W_Σ^-. It follows that

$$\mathbb{T}_\Sigma \simeq Hom_\mathcal{O}(\mathbb{T}_\Sigma, \mathcal{O})$$

as a $\mathbb{T}_\Sigma$-module, which is the *Gorenstein property* of $\mathbb{T}_\Sigma$.

At any rate, the Weil pairing induces a perfect pairing

$$\langle , \rangle_\Sigma : W_\Sigma[\wp] \times W_\Sigma/\wp W_\Sigma \to \mathcal{O}.$$

Fix a symplectic isomorphism $W_\Sigma \simeq \mathbb{T}_\Sigma^2$, where the pairing on the right is the determinant pairing coupled with the self-duality of $\mathbb{T}_\Sigma$. Since $W_\Sigma[\wp] = I^2 \simeq \eta_\Sigma^2$ and $W_\Sigma/\wp W_\Sigma \simeq \mathcal{O}^2$, it is easy to obtain from the above the following lemma (see [D-D-T], Lemma 4.17, and put $d = 2$ there).

Lemma 78. *The submodule $W_\Sigma[\wp]$ is free of rank 2 over $\mathcal{O} = \mathbb{T}_\Sigma/\wp$. Let $\{x, y\}$ be a basis of $W_\Sigma[\wp]$ over $\mathcal{O}$. Then*

$$\eta_{\Sigma,f} = (\langle x, y\rangle_\Sigma).$$

We now compare the quantities $\eta_{\Sigma,f}$ and $\eta_{\Sigma',f}$ by comparing $W_\Sigma[\wp]$ and $W_{\Sigma'}[\wp']$. Recall $\Sigma' = \Sigma \cup \{p\}$ and $N_{\Sigma'} = N_\Sigma p^2$ if $p \neq \ell$ (in which case we assumed that $\overline{\rho}$ was unramified at p), or $N_{\Sigma'} = N_\Sigma \ell$ if $p = \ell$ (in which case $\overline{\rho}$ was flat and ordinary at ℓ). There are 3 (if $p \neq \ell$) or 2 (if $p = \ell$) degeneracy maps

$$\delta_i : X_0(N_{\Sigma'}) \to X_0(N_\Sigma)$$

coming from $\tau \mapsto p^i\tau$ ($i = 0, 1, 2$ or $i = 0, 1$ respectively) on $\mathfrak{H}$. They induce, by Albanese functoriality, similar maps on Jacobians, hence maps

$$\delta_i : \mathcal{T}_\ell J_0(N_{\Sigma'}) \otimes \mathcal{O} \to \mathcal{T}_\ell J_0(N_\Sigma) \otimes \mathcal{O} \twoheadrightarrow (\mathcal{T}_\ell J_0(N_\Sigma) \otimes \mathcal{O})_\mathfrak{m} = W_\Sigma,$$

compatible with all the Hecke operators T_r (including the $r|N_\Sigma$) except for T_p. It follows from the way the ideals $\mathfrak{m}$ and $\mathfrak{m}'$ were constructed that the homomorphism

$$\beta = \delta_0 - p^{-1}T_p \circ \delta_1 + p^{-1}\delta_2 \quad (p \neq \ell)$$

$$\beta = \delta_0 - u_\ell^{-1} \circ \delta_1 \quad (p = \ell),$$

where u_ℓ is the "unit root" in $\mathbb{T}_\Sigma$ of $X^2 - T_\ell X + \ell$ (note that if $p = \ell$ it follows from our running assumptions that $\overline{\rho}$ was flat and *ordinary* at ℓ, hence all the $a_\ell(f)$, for $f \in \mathcal{N}_\Sigma$, are in $\mathcal{O}^\times$), is a homomorphism

$$\beta : W_{\Sigma'} \to W_\Sigma,$$

commuting with all the good Hecke operators, i.e. a $\mathbb{T}_{\Sigma'}$-homomorphism, where we let $\mathbb{T}_{\Sigma'}$ act on the target via the canonical homomorphism $\mathbb{T}_{\Sigma'} \to \mathbb{T}_\Sigma$.

Let $\beta' : W_\Sigma \to W_{\Sigma'}$ be the dual of β with respect to the Weil pairings $\langle x, y\rangle_\Sigma$ and $\langle x, y\rangle_{\Sigma'}$. A computation of 3×3 or 2×2 determinants yields (in all cases) that $\beta\beta' \in End(W_\Sigma)$ is, up to a unit, equal to the Hecke operator

$$c_p = (p-1)((1+p)^2 - T_p^2) \in \mathbb{T}_\Sigma$$

(note that p is a good prime at level N_Σ!). See [D-D-T], top of p.133.

Lemma 79. *The homomorphism $\beta' : W_\Sigma \to W_{\Sigma'}$ has an $\mathcal{O}$-torsion free cokernel (i.e. it embeds W_Σ as an $\mathcal{O}$-direct summand of the larger module $W_{\Sigma'}$). As a consequence, it maps $W_\Sigma[\wp]$ isomorphically onto $W_{\Sigma'}[\wp']$.*

Recall that $W_{\Sigma'}[\wp']$ is the common kernel, in $W_{\Sigma'}$, of all the endomorphisms $T \in \mathbb{T}_{\Sigma'}$ which act trivially on f ($\wp'$ is the kernel of $\pi_{\Sigma',f} : \mathbb{T}_{\Sigma'} \to \mathcal{O}$). A similar interpretation exists for $W_\Sigma[\wp]$. Since β' is a $\mathbb{T}_{\Sigma'}$-homomorphism, it follows that $W_\Sigma[\wp]$ is in fact mapped by β' to $W_{\Sigma'}[\wp']$. It furthermore follows easily that after tensoring with E over $\mathcal{O}$ it is an isomorphism. The fact that β' has an $\mathcal{O}$-torsion free cokernel implies that it induces an isomorphism $W_\Sigma[\wp] \simeq W_{\Sigma'}[\wp']$.

Corollary 80. $\eta_{\Sigma',f} = c_{p,f}\eta_{\Sigma,f}$.

Proof. Let $\{x, y\}$ be a basis of $W_\Sigma[\wp]$. Then $\{\beta' x, \beta' y\}$ is a basis of $W_{\Sigma'}[\wp']$. We have (up to a unit)

$$\eta_{\Sigma',f} \sim \langle \beta' x, \beta' y\rangle_{\Sigma'} = \langle \beta\beta' x, y\rangle_\Sigma = \langle c_p x, y\rangle_\Sigma .$$

Since $x \in W_\Sigma[\wp]$, the Hecke operator c_p acts on it via $\pi_{\Sigma,f}(c_p) = c_{p,f} \in \mathcal{O}$. It follows that

$$\eta_{\Sigma',f} \sim c_{p,f} \langle x, y\rangle_\Sigma = c_{p,f}\eta_{\Sigma,f}.$$

□

It remains to prove the lemma (the fact that coker(β') is torsion-free). This follows from an argument from Ribet's theorem on "raising the level" known as Ihara's Lemma. See Lemma 4.6 of [Di-Ri], or [D-D-T], Lemma 4.24.

6. Commutative Algebra (weeks 13,14)

6.1. The cotangent space and the congruence ideal.

6.1.1. *The two invariants.* Let $\mathcal{O}$ be the ring of integers in a finite extension of $\mathbb{Q}_\ell$ and $\mathcal{C}_\mathcal{O}$ the category of local complete noetherian $\mathcal{O}$-algebras with residue field k, where the morphisms are local $\mathcal{O}$-homomorphisms inducing the identity on k. Let $\dot{\mathcal{C}}_\mathcal{O}$ be the category of pairs (A, π_A) where $\pi_A : A \to \mathcal{O}$ is a morphism (such a pair will be called a pointed, or augmented, $\mathcal{O}$-algebra). Morphisms are local homomorphims $f : A \to B$ of $\mathcal{O}$-algebras such that $\pi_B \circ f = \pi_A$. For example, using the notation of the previous chapters, if $f \in \mathcal{N}_\Sigma$ we may take $(\mathbb{T}_\Sigma, \pi_{\Sigma,f})$ or $(R_\Sigma, \pi_{\Sigma,f} \circ \phi_\Sigma)$.

With $(A, \pi_A) \in \mathcal{C}_{\mathcal{O}}^{\cdot}$ we associate two invariants. Let $I_A = \ker(\pi_A)$. Define

$$\Phi_A = I_A/I_A^2, \quad \eta_A = \pi_A(\mathrm{Ann}_A I_A).$$

The invariant Φ_A is an $\mathcal{O}$-module, the cotangent space along π_A. The ideal $\eta_A \subset \mathcal{O}$ is called the congruence ideal of π_A.

Example 81. (i) $A = \mathcal{O}[[X,Y]]/(XY, X(X-\lambda), Y(Y-\lambda))$. We have an $\mathcal{O}$-algebra isomorphism

$$A \simeq \{(a,b,c) \in \mathcal{O}^3 | \, a \equiv b \equiv c \mod \lambda\}$$

under $f \mapsto (f(0,0), f(0,\lambda), f(\lambda,0))$. To check that we get everything on the RHS use the polynomials $X+Y-\lambda$,X and Y to get $(-\lambda,0,0)$, $(0,0,\lambda)$ and $(0,\lambda,0)$. Let π_A be the projection to the first factor, i.e. $f \mapsto f(0,0)$. Then

$$\Phi_A \simeq \mathcal{O}/\lambda \times \mathcal{O}/\lambda, \;\; \eta_A = (\lambda).$$

(ii) $A = \mathcal{O}[[X,Y]]/(X(X-\lambda), Y(Y-\lambda))$. We have

$$A \simeq \{(a,b,c,d) \in \mathcal{O}^4 | \, a \equiv b \equiv c \equiv d \mod \lambda, \; a+d \equiv b+c \mod \lambda^2\}$$

under $f \mapsto (f(0,0), f(0,\lambda), f(\lambda,0), f(\lambda,\lambda))$. To check that we get everything on the RHS use the polynomials $X+Y-\lambda$, X, Y to get $(-\lambda,0,0,\lambda)$, $(0,0,\lambda,\lambda)$, $(0,\lambda,0,\lambda)$, and XY to get $(0,0,0,\lambda^2)$. Again, let π_A be the projection to the first coordinate.

Here, in contrast to the first example,

$$\Phi_A \simeq \mathcal{O}/\lambda \times \mathcal{O}/\lambda, \;\; \eta_A = (\lambda^2).$$

Note that in this example A is a l.c.i. and $\#\Phi_A = \#(\mathcal{O}/\eta_A)$. Both assertions fail for (i). Note also the the first A is a quotient of the second A.

(iii) $A = \mathcal{O}[[X]]/(X^2)$, $\pi_A f = f(0)$. Here

$$\Phi_A \simeq \mathcal{O}, \;\; \eta_A = \{0\}.$$

This example shows that Φ_A and $\mathcal{O}/\eta_A$ need not be finite, even if A is finite flat over $\mathcal{O}$.

(iv) $A = \mathcal{O}[[X]]/(X(X-\lambda^n)) \simeq \{(a,b) \in \mathcal{O}^2 | \, a \equiv b \mod \lambda^n\}$ under $f \mapsto (f(0), f(\lambda^n))$, π_A being the first projection. This is a "good" example, like (ii), in the sense that A is a l.c.i. and

$$\Phi_A \simeq \mathcal{O}/\lambda^n, \;\; \eta_A = (\lambda^n)$$

have $\#\Phi_A = \#(\mathcal{O}/\eta_A)$.

(v) Quite generally, we can always assume that

$$A = \mathcal{O}[[X_1, \dots, X_n]]/(f_1, \dots, f_r)$$

with f_i power series without constant term, and $\pi_A(h \mod (f_i)) = h(0,...,0)$. Then, letting $\overline{f}_i$ be the linear term in f_i we have

$$\Phi_A \simeq \left(\bigoplus_{i=1}^{n} \mathcal{O} X_i \right) / \left(\sum_{i=1}^{r} \mathcal{O}\overline{f_i} \right).$$

If $r < n$ it is of infinite length. On the other hand $I_A = (X_1, \dots, X_n)$, $\mathrm{Ann}_A I_A$ is the image in A of the ideal

$$J = \{h \in \mathcal{O}[[X_1, \dots, X_n]] | \, \forall i \, X_i h \in (f_1, \dots, f_r)\}$$

and η_A is the ideal of $\mathcal{O}$ generated by the constant terms of all such h, equivalently, $\eta_A = (\lambda^m)$ where m is the smallest integer such that there exists an h in J with $v_\lambda(h(0)) = m$.

6.1.2. *The First Criterion.* We state a version of Wiles' criterion that is due to Lenstra. Recall that $A \in \mathcal{C}_{\mathcal{O}}$ is a l.c.i. if it is isomorphic to a ring of the form $\mathcal{O}[[X_1, \ldots, X_n]]/(f_1, ..., f_n)$, where the f_i form a regular sequence. If the Krull dimension of A is 1, and it is given by such a presentation, with as many f_i as X_j, then the f_i necessarily form a regular sequence, and A is a l.c.i..

Theorem 82. *Let* (R, π_R), $(T, \pi_T) \in \mathcal{C}_{\mathcal{O}}^{\cdot}$ *and* $\phi : R \twoheadrightarrow T$ *a surjective morphism of pointed* $\mathcal{O}$*-algebras. Then:*

(i) $\eta_R \subset \eta_T$, *hence* $\#(\mathcal{O}/\eta_R) \geq \#(\mathcal{O}/\eta_T)$.

(ii) $\Phi_R \twoheadrightarrow \Phi_T$, *hence* $\#\Phi_R \geq \#\Phi_T$.

(iii) $\#\Phi_R \geq \#(\mathcal{O}/\eta_R)$ *(and similarly of course for* T*)*

(iv) (the main point) Assume that T *is finite and flat over* $\mathcal{O}$ *and that* $\eta_T \neq 0$. *Then* $\#\Phi_R \geq \#(\mathcal{O}/\eta_T)$ *and equality holds if and only if* ϕ *is an isomorphism and* $R \simeq T$ *is a l.c.i..*

Corollary 83. *Assume that* $\eta_R \neq 0$. *Then* R *is a l.c.i. if and only if* $\#\Phi_R = \#(\mathcal{O}/\eta_R)$. *(Take* $R = T$ *and* ϕ *the identity.)*

6.1.3. *Fitting ideals and the proof of (i)-(iii).* The assertion $\#\Phi_R \geq \#(\mathcal{O}/\eta_T)$ in (iv) is a direct consequence of (i) and (iii) (or of (ii) and (iii)). Points (i) and (ii) are easy. For (i) note that $\mathrm{Ann}_R(I_R)$ is mapped under ϕ to $\mathrm{Ann}_T(I_T)$, because $\phi : I_R \twoheadrightarrow I_T$ by the surjectivity of ϕ. Therefore

$$\eta_R = \pi_R(\mathrm{Ann}_R(I_R)) = \pi_T \circ \phi(\mathrm{Ann}_R(I_R)) \subset \pi_T(\mathrm{Ann}_T(I_T)) = \eta_T.$$

For (ii) note that Φ_R is functorial: ϕ induces a surjection $I_R \twoheadrightarrow I_T$, hence a surjection $\Phi_R \twoheadrightarrow \Phi_T$. In fact (this is not used here, but is good to know) by Nakayama, if $\phi : \Phi_R \twoheadrightarrow \Phi_T$ is surjective, so is $\phi : R \twoheadrightarrow T$.

Point (iii) is a little deeper, and requires the notion of (the zeroth) *Fitting ideals.*

Definition 84. Let A be a noetherian ring and M a finite A-module. If $\theta : A^n \twoheadrightarrow M$, consider the ideal in A generated by all the determinants $\det(v_1, \ldots, v_n)$ where $v_i \in \ker(\theta)$. Denote it by $\mathrm{Fit}_A(M)$.

The following are easy:

- $\mathrm{Fit}_A(M)$ is generated by the determinants $\det(v_1, \ldots, v_n)$ where the v_i range over a given set of generators of $\ker(\theta)$. In particular, if this kernel is generated by $< n$ vectors in A^n, then $\mathrm{Fit}_A(M) = 0$.
- If $m_1, \ldots, m_n$ are the generators of M corresponding to $\theta(e_i)$, and $m_{n+1} = \sum_{i=1}^n a_i m_i$, and if $\theta' : A^{n+1} \twoheadrightarrow M$ sends e_i to m_i ($1 \leq i \leq n+1$) then $\ker(\theta')$ is generated by $(v, 0)$ where $v \in \ker(\theta)$ and the extra vector $(a_1, \ldots, a_n, -1)$. Using the previous remark it follows that the Fitting ideal computed via θ is the same as the Fitting ideal computed via θ'.
- It follows that for any two $\theta : A^n \twoheadrightarrow M$ and $\theta' : A^m \twoheadrightarrow M$ the Fitting ideals computed via θ and θ' agree. (Compare both to the Fitting ideal of $\theta \oplus \theta' : A^{m+n} \twoheadrightarrow M$ and use, inductively, the previous remark.) Hence the Fitting ideal is well defined.

Proposition 85. *(i) If* M *is generated over* A *by* n *elements, then*

$$\mathrm{Ann}_A(M)^n \subset \mathrm{Fit}_A(M) \subset \mathrm{Ann}_A(M).$$

(ii) $\mathrm{Fit}_A(A/I) = I$.

(iii) If $0 \to M' \to M \to M'' \to 0$ *is an exact sequence then*

$$\mathrm{Fit}_A(M')\mathrm{Fit}_A(M'') \subset \mathrm{Fit}_A(M),$$

and if the exact sequence splits, so that $M \simeq M' \oplus M''$*, this is an equality.*

(iii) If A *is a PID and* $M = A^r \oplus A/(f_1) \oplus \cdots \oplus A/(f_s)$ *then* $\mathrm{Fit}_A(M) = 0$ *if* $r > 0$ *and* $\mathrm{Fit}_A(M) = (f_1 \cdots f_s)$ *otherwise. If* A *is a DVR* $\mathit{length}_A(M) = \mathit{length}_A(A/\mathrm{Fit}_A(M))$.

(iv) If B *is an* A*-algebra, then* $\mathrm{Fit}_B(B \otimes_A M) = B\mathrm{Fit}_A(M)$.

Proof. The only non-trivial claim is (i), from which (ii) follows letting $n = 1$. If $m_1, \dots, m_n$ are generators of M and $\theta : A^n \twoheadrightarrow M$ the corresponding surjection, let $\psi : A^n \to A^n$ be any map with $\theta \circ \psi = 0$, and $\psi^\dagger : A^n \to A^n$ the adjoint map, so that $\psi \circ \psi^\dagger$ is multiplication by $\det(\psi)$. Then

$$0 = \theta \circ \psi \circ \psi^\dagger = \theta \circ \det \psi = \det \psi \circ \theta$$

so by the surjectivity of θ, $\det(\psi) \in \mathrm{Ann}_A(M)$. But $\mathrm{Fit}_A(M)$ is generated by all such $\det(\psi)$. If $a_1, \dots, a_n \in \mathrm{Ann}_A(M)$ the map $\psi : A^n \to A^n$, $\psi(e_i) = a_i e_i$, satisfies $\theta \circ \psi = 0$, so

$$a_1 \cdots a_n = \det(\psi) \in \mathrm{Fit}_A(M)$$

and we get the other inclusion. □

We can now finish the proof of (iii). Since

$$\Phi_R = I_R/I_R^2 = R/I_R \otimes_R I_R = \mathcal{O} \otimes_{\pi_R, A} I_R$$

we have

$$\mathrm{Fit}_{\mathcal{O}}(\Phi_R) = \pi_R(\mathrm{Fit}_R(I_R)) \subset \pi_R(\mathrm{Ann}_R(I_R)) = \eta_R,$$

so, $\mathcal{O}$ being a DVR, $\#\Phi_R = \#\mathcal{O}/\mathrm{Fit}_{\mathcal{O}}(\Phi_R) \geq \#\mathcal{O}/\eta_R$.

6.1.4. *Koszul complexes.* We shall need to work with Koszul complexes. If R is a commutative ring and $f_1, ..., f_n \in R$ we define $K_i(f, R)$ to be the complex where for $0 \leq m \leq n$

$$K_m = \bigwedge^m R^n = \bigoplus_{1 \leq i_1 < \cdots < i_m \leq n} R e_{i_1} \wedge \cdots \wedge e_{i_m}$$

and where $d : K_m \to K_{m-1}$ is

$$d(e_{i_1} \wedge \cdots \wedge e_{i_m}) = \sum_{j=1}^m (-1)^{j-1} f_{i_j} e_{i_1} \wedge \cdots \wedge \widehat{e_{i_j}} \wedge \cdots \wedge e_{i_m}.$$

The K_m are free modules of rank $\binom{n}{m}$. The homologies of the complex are denoted $H_m(f, R)$. Clearly $H_0(f, R) = R/I$ where $I = (f_1, \dots, f_n)$.

If M is an R-module we let $K_m(f, M) = K_m(f, R) \otimes_R M$ and denote by $H_m(f, M)$ the corresponding homologies. Clearly $H_0(f, M) = M/IM$.

Proposition 86. *(i) The homologies* $H_m(f, M)$ *are annihilated by* I.

(ii) If $(f_1, \dots, f_n)$ *is an* M*-regular sequence (i.e. multiplication by* f_i *on the quotient* $M/(f_1, ..., f_{i-1})M$ *is injective for* $i = 1, ..., n$*), or, more generally, if* I *contains an* M*-regular sequence of length* n*, then* $H_i(f, M) = 0$ *for* $i > 0$.

(iii) If $(f_1, ..., f_n)$ *is a regular sequence in* R*, then* $K.(f, R)$ *is a free resolution of* R/I.

Proof. (i) Assume that $x = \sum x_{i_1 i_2 \ldots i_m} e_{i_1} \wedge \cdots \wedge e_{i_m}$ satisfies $dx = 0$. We must find a y with, say, $dy = f_1 x$. Write $x = e_1 \wedge x' - x''$ where x' and x'' are in K_{m-1} and are supported in indices $\{2, ..., n\}$. From $dx = 0$ we get (separating the index sets containing 1 from the rest) $dx' = 0$ and $dx'' = f_1 x'$. It follows that $y = -e_1 \wedge x''$ solves our problem: $dy = e_1 \wedge dx'' - f_1 x'' = f_1(e_1 \wedge x' - x'') = f_1 x$.

(ii) Let $p_1, ..., p_n$ be an M-regular sequence in I. For $0 \leq j \leq n$ we show, by decreasing induction on j, that

$$H_i(f, M/(p_1, ..., p_j)M) = 0$$

for all $i > j$. For $j = n$ this is trivial (all the homologies of $K.(f, N)$ with $i > n$ vanish, for any module N). For $j = 0$ this is part (ii). Assume therefore that the assertion had been proved for some $j \geq 1$, and let us prove it for $j - 1$. Let $M' = M/(p_1, ..., p_{j-1})M$. Since the p_i form an M-regular sequence,

$$0 \to M' \stackrel{p_j}{\to} M' \to M'/p_j M' \to 0$$

is a short exact sequence. Since $K.(f, R)$ is a complex of *free* modules we get, by tensoring, a short exact sequence of complexes (with *descending* indices)

$$0 \to K.(f, M') \stackrel{p_j}{\to} K.(f, M') \to K.(f, M'/p_j M') \to 0.$$

Since, by (i), p_j kills the homologies in positive degrees, we get from the long exact sequence in homology, a bunch of short exact sequences ($i \geq 2$)

$$0 \to H_i(f, M') \to H_i(f, M'/p_j M') \to H_{i-1}(f, M') \to 0.$$

By the induction hypothesis the middle term vanishes for $i > j$, so $H_{i-1}(f, M')$ also vanishes for $i > j$, i.e. $H_i(f, M')$ vanishes for $i > j - 1$.

(iii) This is a special case of (ii). □

6.2. **Complete intersections and the Gorenstein property.**

6.2.1. *Tate's theorem on l.c.i.* We continue to use the standard notation above. We assume now that $A \in \mathcal{C}_{\mathcal{O}}$ is a l.c.i. and is also finite and flat over $\mathcal{O}$.

Theorem 87. *Suppose that*

$$A = \mathcal{O}[[X_1, ..., X_n]]/(f_1, ..., f_n)$$

where the f_i have no constant term, and $\pi_A(h) = h(0)$, so that I_A is the image in A of $(X_1, ..., X_n)$. Assume that A is finite flat over $\mathcal{O}$. Write

$$f_j = \sum_{i=1}^{n} X_i g_{ij}$$

and let d be the image of $D = \det(g_{ij})$ in A. Then:

(i) $\mathrm{Fit}_A(I_A) = \mathrm{Ann}_A(I_A) = (d) \neq 0$,

(ii) As an $\mathcal{O}$-module, (d) is a rank 1 direct summand of A.

(iii) $\#\Phi_A = \#(\mathcal{O}/\eta_A)$.

Proof. Parts (ii) and (iii) are easy consequences of (i). The homomorphism π_A induces an isomorphism $A/I_A \simeq \mathcal{O}$, so $A \simeq \mathcal{O} \oplus I_A$ as $\mathcal{O}$-modules. Thus $(d) = Ad = \mathcal{O}d$ is rank-1 as an $\mathcal{O}$-module. Since A is $\mathcal{O}$-torsion free, $\mathrm{Ann}_A(I_A)$ is saturated as an $\mathcal{O}$-submodule, so it is a direct summand. This gives (ii). For (iii) note that the proof of (iii) in Theorem 82 showed that if $\mathrm{Fit}_A(I_A) = \mathrm{Ann}_A(I_A)$ holds, then (iii) holds too.

To prove (i) let $P = \mathcal{O}[[X_1, ..., X_n]]$, let f be the row vector $(f_1, ..., f_n)$ and X the row vector $(X, ..., X_n)$, so that $f : P^n \to P$ and $X : P^n \to P$ are linear transformations satisfying $f = X \circ G$ with $G = (g_{ij})$. Let $V = P^n$. By functoriality, we get a commutative diagram of augmented Koszul complexes

$$\begin{array}{cccccccccccl} 0 & \to & \bigwedge^n V & \to & \cdots & \to & V & \stackrel{f}{\to} & P & \stackrel{\varepsilon_A}{\to} & A & \to & 0 & \quad : K.(f,P) \\ & & D\downarrow & & & & G\downarrow & & || & & \pi_A\downarrow & & & \\ 0 & \to & \bigwedge^n V & \to & \cdots & \to & V & \stackrel{X}{\to} & P & \stackrel{\varepsilon_{\mathcal{O}}}{\to} & \mathcal{O} & \to & 0 & \quad : K.(X,P) \end{array}$$

where the horizontal arrows are the differentials constructed from the sequences f and X. Since the rows are free resolutions of A (resp. $\mathcal{O}$) as P-modules (f and X are regular sequences!) we may use them to compute $Tor_j(A, N)$ (resp. $Tor_j(\mathcal{O}, N)$) on the category of P-modules by tensoring the resolutions on the right $\otimes_P N$ and calculating the homology of the resulting complex. Taking $N = A$ and $j = n$ we get, from the left-most column, a commutative diagram with exact rows

$$\begin{array}{ccccccc} 0 & \to & Tor_n(A,A) & \to & \bigwedge^n V_A & \stackrel{0}{\to} & \bigwedge^{n-1} V_A \\ & & \pi_{A*}\downarrow & & d\downarrow & & \downarrow \\ 0 & \to & Tor_n(\mathcal{O},A) & \to & \bigwedge^n V_A & \stackrel{X^\dagger}{\to} & \bigwedge^{n-1} V_A \end{array}.$$

Here the top arrow between $\bigwedge^n V_A$ and $\bigwedge^{n-1} V_A$ is 0 because all the f_j map to 0 in A. The bottom arrow is the map

$$X^\dagger : e_1 \wedge \cdots \wedge e_n \mapsto \sum_{j=1}^{n} (-1)^{j-1} \overline{X}_j e_1 \wedge \cdots \wedge \widehat{e_j} \wedge \cdots \wedge e_n$$

where $\overline{X}_j$ is the image of X_j in A. It follows that $Tor_n(\mathcal{O}, A) \simeq \mathrm{Ann}_A(I_A)$. However, $\pi_A \circ \iota_A = id_{\mathcal{O}}$ where $\iota_A : \mathcal{O} \to A$ is the structure map. It follows that the composition of

$$Tor_n(\mathcal{O},A) \stackrel{\iota_{A*}}{\to} Tor_n(A,A) \stackrel{\pi_{A*}}{\to} Tor_n(\mathcal{O},A)$$

is the identity, and in particular that π_{A*} is *onto* $Tor_n(\mathcal{O}, A)$. Since $Tor_n(A, A) \simeq \bigwedge^n V_A \simeq A$ we conclude that

$$\mathrm{Ann}_A(I_A) = (d).$$

We claim that $d \neq 0$. In fact, the diagrams above can be reduced modulo λ, and yield a similar result for the reduction $\overline{A} = k \otimes_{\mathcal{O}} A$, with $\overline{d}$ replacing d and $\mathfrak{m}_{\overline{A}} = \overline{I}_A$ replacing I_A. But $\overline{A}$ is an artinian ring, so the annihilator of its maximal ideal is non-zero. We conclude that $\overline{d} \neq 0$, and a-fortiori $d \neq 0$.

But we have seen that, in general,

$$(d) = (\det(G)) \subset \mathrm{Fit}_A(I_A) \subset \mathrm{Ann}_A(I_A).$$

Thus equality holds throughout. This concludes the proof. □

6.2.2. *The Gorenstein condition.* Although this will not be needed in the sequel, let us draw the following conclusion. Recall that a ring $A \in \mathcal{C}_{\mathcal{O}}$ which is finite and flat over $\mathcal{O}$ is called *Gorenstein* if

$$A \simeq Hom_{\mathcal{O}}(A, \mathcal{O})$$

as A-modules.

Corollary 88. *If $A \in \mathcal{C}_{\mathcal{O}}$ is finite flat over $\mathcal{O}$ and is a l.c.i., then it is Gorenstein.*

[The example $A = \mathcal{O}[[X]]/(\lambda X) \simeq \mathcal{O} \oplus kt \oplus kt^2 \oplus \cdots$ (as an $\mathcal{O}$-module) shows that without the finite flat assumption, the self-duality need not hold. The definition of Gorenstein however, is more general, and a l.c.i. is always Gorenstein. See the Stacks project.]

Proof. By Tate's theorem, part (ii), there exists a map of $\mathcal{O}$-modules $t : A \to \mathcal{O}$ with $t(d) = 1$. We claim that

$$A \to Hom_{\mathcal{O}}(A, \mathcal{O}), \quad a \mapsto at,$$

where $at(x) = t(ax)$, is an isomorphism.

Both sides are finite free over $\mathcal{O}$ of the same rank. It is therefore enough to show that the map is surjective. By Nakayama, it is enough to show that the corresponding map is surjective after we reduce modulo λ, and since now we deal with finite dimensional k-vector spaces, it is enough to prove that it is injective, namely that

$$\mathrm{Ann}_{\overline{A}}(\bar{t}) = 0.$$

Clearly $(\bar{d}) \not\subseteq \mathrm{Ann}_{\overline{A}}(\bar{t})$. If $\mathrm{Ann}_{\overline{A}}(\bar{t}) \neq 0$ it must contain a minimal non-zero ideal $\mathfrak{a}$. This $\mathfrak{a}$ must annihilate $\overline{I}_A$ (otherwise $\overline{I}_A \mathfrak{a}$ is strictly smaller and still non-zero). Thus $\mathfrak{a} \subset (\bar{d})$, and since $\dim_k(\bar{d}) = 1$, $\mathfrak{a} = (\bar{d})$, contradicting $(\bar{d}) \not\subseteq \mathrm{Ann}_{\overline{A}}(\bar{t})$. We conclude that $\mathrm{Ann}_{\overline{A}}(\bar{t}) = 0$, as desired. □

6.3. **Proof of the first criterion.**

Lemma 89. *Let $f : A \to B$ be a homomorphism in $\mathcal{C}_{\mathcal{O}}$ and assume that B is finite flat over $\mathcal{O}$. Let $\overline{f} : \overline{A} \to \overline{B}$ be its reduction modulo λ. Then f is an isomorphism if and only if $\overline{f}$ is an isomorphism.*

Proof. Assume $\overline{f}$ is an isomorphism. By Nakayama (applied to A and B as $\mathcal{O}$-modules), since $\overline{f}$ is surjective, so is f. Suppose $J = \ker(f)$. From the exact sequence

$$0 \to J \to A \to B \to 0$$

and the fact that B is $\mathcal{O}$-torsion free, we get the exactness of

$$0 \to \overline{J} \to \overline{A} \to \overline{B} \to 0.$$

This shows that $\overline{J} = 0$, and again by Nakayama, $J = 0$. □

We shall deduce Wiles' numerial criterion, Theorem 82, from the following.

Theorem 90. *In the situation of Theorem 82, the map ϕ is an isomorphism between l.c.i. if and only if*

$$\phi(\mathrm{Fit}_R(I_R)) \not\subseteq \lambda T.$$

Proof. Suppose first that ϕ is an isomorphism of l.c.i.. By Tate's theorem

$$\phi(\mathrm{Fit_R(I_R)}) = \mathrm{Fit}_T(I_T) = \mathrm{Ann}_T(I_T)$$

is a rk 1 direct summand of T as an $\mathcal{O}$-module, hence $\not\subseteq \lambda T$.

For the converse, consider first the same statement with $\mathcal{O}$ replaced by k: a homomorphism

$$\phi : R \to T$$

in $\mathcal{C}_k^{\cdot}$ (commuting with $\pi_R : R \to k$ and $\pi_T : T \to k$), where $\dim_k T < \infty$, is an isomorphism between l.c.i. if (and only if)

$$\phi(\mathrm{Fit}_R(I_R)) \neq 0.$$

Note that now $I_R = \ker(\pi_R)$ is the maximal ideal of R.

Write $R = k[[X_1, ..., X_n]]/J_R$ in such a way that $\phi(X_i)$ generate I_T *as a k-vector space* (this is possible since $\dim_k T < \infty$). Let $J_T = \ker(k[[X_1, ..., X_n]] \twoheadrightarrow R \twoheadrightarrow T)$, so that $T = k[[X_1, \ldots, X_n]]/J_T$ and $J_R \subset J_T$. The ideals I_R and I_T are the images of $I = (X_1, ..., X_n)$ mod J_R and J_T respectively.

The assumption $\phi(\mathrm{Fit}_R(I_R)) \neq 0$ means that there are $g_{ij} \in k[[X_1, ..., X_n]]$ such that $\sum_{j=1}^n g_{ij} X_j \in J_R$ but $\det(g_{ij}) \notin J_T$.

Since the X_i span $I_T = I/J_T$ over k, the monomials $X_i X_j$ span I^2/IJ_T. This means that every element of $k[[X_1, ..., X_n]]/IJ_T$ is represented by a quadratic polynomial in the X_i. Let p_i and q_i be quadratic polynomials such that

$$p_i \equiv \sum_{j=1}^n g_{ij} X_j \mod IJ_T,$$

$$q_i \equiv X_i^3 \mod IJ_T.$$

Let

$$f_i = X_i^3 - q_i + p_i.$$

Note that $f_i \in IJ_T + J_R \subset J_T$ (since the $\sum_{j=1}^n g_{ij} X_j \in J_R$), and that $f_i = \sum_{j=1}^n G_{ij} X_j$ for $G_{ij} \equiv g_{ij} \mod J_T$ (since the difference $f_i - \sum_{j=1}^n g_{ij} X_j$, which lies in IJ_T, can be written as $\sum_{j=1}^n H_{ij} X_j$ with $H_{ij} \in J_T$, so we may put $G_{ij} = g_{ij} + H_{ij}$).

Consider

$$B = k[[X_1, ..., X_n]]/(f_1, ..., f_n) \overset{\psi}{\twoheadrightarrow} k[[X_1, ..., X_n]]/J_T = T.$$

Since every element of B is represented by a polynomial which is of degree ≤ 2 in each X_i (X_i^3 is expressible as a quadratic polynomial modulo $(f_1, ..., f_n)$), $\dim_k B < \infty$. It follows from Tate's theorem that B is a l.c.i. and $(d) = (\det(G_{ij}))$ is the *unique minimal ideal* of B. (It has dimension 1 over k and is the annihilator of the maximal ideal I_B; any minimal non-zero ideal must annihilate I_B, so is contained in (d), hence must be equal to it.)

Now $\psi(d) \neq 0$ by our assumption since $d = \det(G_{ij}) \equiv \det(g_{ij}) \mod J_T$. It follows that (d) is not contained in $\ker(\psi)$. As it is the unique minimal ideal in B, and must be contained in *any* non-zero ideal, $\ker \psi = 0$ and ψ is an isomorphism.

It follows that $B \simeq T$ is a l.c.i.. It also follows that

$$J_T = (f_1, ..., f_n) \subset IJ_T + J_R \subset J_T$$

so $IJ_T + J_R = J_T$. By Nakayama $J_R = J_T$ and ϕ is an isomorphism.

This concludes the proof of the theorem with $\mathcal{O}$ replaced by k. Getting back to the original formulation, assume that

$$\phi(\mathrm{Fit}_R(I_R)) \not\subseteq \lambda T.$$

Since $R = I_R \oplus \mathcal{O}$ as an $\mathcal{O}$-module, the kernel of $\overline{R} \to k$ is $\overline{I}_R$. We observe that $\overline{\phi}(\mathrm{Fit}_{\overline{R}}(\overline{I}_R)) \neq 0$, so $\overline{\phi}$ is an isomorphism between $\overline{R}$ and $\overline{T}$ and these rings are l.c.i.. It follows from the previous Lemma (thanks to the assumption that T is finite flat over $\mathcal{O}$) that ϕ is an isomorphism too. Pick an isomorphism

$$k[[X_1, ..., X_n]]/(f_1, ..., f_n) \simeq \overline{T}.$$

By Nakayama, we can lift it to a surjection $\mathcal{O}[[X_1,...,X_n]] \twoheadrightarrow T$ whose kernel, J_T, reduces to $(f_1,...,f_n)$. We may therefore lift f_i to $\widetilde{f_i} \in J_T$. Consider now the surjection

$$\mathcal{O}[[X_1,...,X_n]]/(\widetilde{f_1},...,\widetilde{f_n}) \twoheadrightarrow T.$$

Applying the Lemma, we find that it is an isomorphism, hence T is a l.c.i.. $\square$

We now complete the proof of Theorem 82. Assume that

$$\#\Phi_R = \#\mathcal{O}/\eta_T$$

and this number is finite. We have seen that this means

$$\pi_T \circ \phi \mathrm{Fit}_R(I_R) = \pi_R \mathrm{Fit}_R(I_R) = \pi_T \mathrm{Ann}_T(I_T).$$

Let us show first that $I_T \cap \mathrm{Ann}_T(I_T) = 0$. Pick a $y \in \mathrm{Ann}_T(I_T)$ with $\pi_T(y) \neq 0$. If $x \in I_T \cap \mathrm{Ann}_T(I_T)$, then $xy = 0$ and $x(y - \pi_T(y)) = 0$. This means $x\pi_T(y) = 0$, so $x = 0$ by our assumption that T is $\mathcal{O}$-torsion free.

It follows that $\pi_T : \mathrm{Ann}_T(I_T) \simeq \eta_T$. This means that $\phi\mathrm{Fit}_R(I_R) = \mathrm{Ann}_T(I_T)$. But $\mathrm{Ann}_T(I_T)$ is $\mathcal{O}$-saturated in T, so (as it is non-zero by the assumption that $\eta_T \neq 0$), $\phi\mathrm{Fit}_R(I_R) \nsubseteq \lambda T$. We therefore conclude from the previous theorem that ϕ is an isomorphism between l.c.i..

6.4. **J-structures and the second criterion.** We shall now prove the Taylor-Wiles patching criterion. We follow a version which is due to Rubin.

Lemma 91. *Let k be a field, $n \geq 1$. Suppose we are given k-algebra homomorphisms*

$$k[[S_1,...,S_n]] \to k[[X_1,...,X_n]] \overset{f}{\twoheadrightarrow} A$$

with f surjective, write $J = \ker f$, and suppose that $\dim_k A/(S_1,...,S_n)A = d < \infty$. Assume that for some $N > n^{n-1}d^n$ the induced map

$$k[[S_1,...,S_n]]/(S_1^N,...,S_n^N) \xrightarrow{g} A/(S_1^N,...,S_n^N)A$$

is injective. Then $J \subset (S_1,...,S_n)$, so that f induces an isomorphism

$$k[[X_1,...,X_n]]/(S_1,...,S_n) \simeq A/(S_1,...,S_n)A$$

and $A/(S_1,...,S_n)A$ is a l.c.i..

Proof. Let $I = (X_1,...,X_n) \subset k[[X_1,...,X_n]]$. Since $A/(S_1,...,S_n)A$ has a finite length d as a $k[[X_1,...,X_n]]$-module, it is killed by I^d, so

$$I^d + (S_1,...,S_n) \subset J + (S_1,...,S_n).$$

Claim: $J \subset I^{d+1}$.

The claim will prove tha lemma, because we shall have

$$I^d + (S_1,...,S_n) \subset J + (S_1,...,S_n) \subset I^{d+1} + (S_1,...,S_n),$$

so by Nakayama's Lemma $I^d \subset (S_1,...,S_n)$, hence $J \subset (S_1,...,S_n)$.

To prove the Claim we suppose that there exists an $\alpha \in J$, $\alpha \notin I^{d+1}$, and reach a contradiction. Consider the exact sequence of finite dimensional k vector spaces

$$0 \to \ker \to k[[X_1,...,X_n]]/I^{ndN} \xrightarrow{\alpha} k[[X_1,...,X_n]]/I^{ndN} \to \mathrm{coker} \to 0.$$

We shall compute

$$\dim_k \ker = \dim_k \mathrm{coker}$$

in two ways.

On the one hand,

$$I^{ndN} \subset (J + (S_1, ..., S_n))^{nN} \subset J + (S_1^N, ..., S_n^N),$$

and $\alpha \in J$, so coker $= k[[X_1, ..., X_n]]/(I^{ndN} + (\alpha))$ maps surjectively onto

$$k[[X_1, ..., X_n]]/(J + (S_1^N, ..., S_n^N)) = A/(S_1^N, ..., S_n^N)A.$$

From this we get

$$\dim_k \operatorname{coker} \geq \dim_k A/(S_1^N, ..., S_n^N)A \geq N^n,$$

by the injectivity of g.

On the other hand, since $\alpha \notin I^{d+1}$, we have $\ker \subset I^{ndN-d}/I^{ndN}$. For this note that if the lowest degree of a monomial in α is m, and the lowest degree of a monomial in β is ℓ, then the lowest degree of a monomial in $\alpha\beta$ is $m\ell$. The dimension of I^{ndN-d}/I^{ndN} is

$$\sum_{\ell=ndN-d}^{ndN-1} \binom{\ell+n-1}{n-1} \leq d(ndN)^{n-1}.$$

Combining the two calculations we get

$$N^n \leq \dim_k \operatorname{coker} = \dim_k \ker \leq d(ndN)^{n-1},$$

contradicting $N > n^{n-1}d^n$. □

We now use Nakayama's Lemma to get an analogous statement for $\mathcal{O}$-algebras.

Corollary 92. *Suppose we have $\mathcal{O}$-algebra homomorphisms*

$$\mathcal{O}[[S_1, ..., S_n]] \to \mathcal{O}[[X_1, ..., X_n]] \overset{f}{\twoheadrightarrow} A$$

with f surjective, such that $A/(S_1, ..., S_n)A$ is free of rank d over $\mathcal{O}$. Let

$$J_m = ((1 + S_1)^{\ell^m} - 1, ..., (1 + S_n)^{\ell^m} - 1).$$

Suppose for some m with $\ell^m > n^{n-1}d^n$ the quotient ring A/J_mA is free as a module over $\mathcal{O}[[S_1, ..., S_n]]/J_m$. Then the induced map

$$h : \mathcal{O}[[X_1, ..., X_n]]/(S_1, ..., S_n) \to A/(S_1, ..., S_n)A$$

is an isomorphism between l.c.i..

Proof. Let us use to denote reduction modulo λ. Note that $\overline{J}_m = (S_1^{\ell^m}, ..., S_n^{\ell^m})$. Using the Lemma with $N = \ell^m$ we deduce that $\overline{h}$ is an isomorphism. Since $A/(S_1, ..., S_n)A$ is finite flat over $\mathcal{O}$, it follows that h is an isomorphism. But being finite over $\mathcal{O}$, $\mathcal{O}[[X_1, ..., X_n]]/(S_1, ..., S_n)$ is necessarily a l.c.i. (the S_i form a regular sequence). □

We can now conclude the proof of Theorem 68. We start with our local homomorphism of local complete noetherian $\mathcal{O}$-algebras

$$R \twoheadrightarrow T$$

where T is known to be finite and free, say of rank d, over $\mathcal{O}$.

Recall that for a fixed n and arbitrarily large m we had commutative diagrams (dubbed "J_m-structures")

$$\begin{tikzcd}
 & \mathcal{O}[[S_1,\ldots,S_n]] \arrow[d] \arrow[dr] & \\
\mathcal{O}[[X_1,\ldots X_n]] \arrow[r, two heads] & R_m \arrow[r, two heads] \arrow[d] & T_m \arrow[d] \\
 & R \arrow[r, two heads] & T
\end{tikzcd}$$

in $\mathcal{C}_{\mathcal{O}}$ satisfying:

- T_m is finite and free as an $\mathcal{O}$-module,
- $T_m/(S_1,\ldots,S_n)T_m = T$ and $R_m/(S_1,\ldots,S_n)R_m = R$,
- T_m/J_mT_m is finite free over $\mathcal{O}[[S_1,\ldots,S_n]]/J_m$.

(The last bullet is stronger than the assumption that the map

$$\mathcal{O}[[S_1,\ldots,S_n]]/J_m \to T_m/J_mT_m$$

is injective, figuring in the conditions imposed on the J_m-structure in Theorem 68. However, it is satisfied in Wiles' patching contsruction, see Theorem 61, so we may just as well impose it.)

Choose m with $\ell^m > n^{n-1}d^n$ and work with the corresponding J_m-structure. Lift the homomorphism from $\mathcal{O}[[S_1,\ldots,S_n]]$ to R_m to a homomorphism from $\mathcal{O}[[S_1,\ldots,S_n]]$ to $\mathcal{O}[[X_1,\ldots X_n]]$. Let $A = T_m$ and apply the Corollary. We deduce that the composite map

$$\mathcal{O}[[X_1,\ldots X_n]]/(S_1,...,S_n) \twoheadrightarrow R_m/(S_1,...,S_n)R_m \twoheadrightarrow T_m/(S_1,...,S_n)T_m$$

is an isomorphism of l.c.i.'s. So is $R_m/(S_1,...,S_n)R_m \twoheadrightarrow T_m/(S_1,...,S_n)T_m$, which by the second bullet is the original surjection $R \twoheadrightarrow T$.

References

[All] P. Allen: *Modularity lifings.* Lecture notes and recordings available at: https://patrick-allen.github.io/teaching/f20/f20-modularity-lifting.html.

[B-C-D-T] C. Breuil, B. Conrad, F. Diamond, R. Taylor: *On the modularity of elliptic curves over $\mathbb{Q}$: Wild 3-adic exercises,* J. Amer. Math. Soc. **14** (2001), 843-939.

[Cal] F. Calegari: *30 years of modularity,* https://www.youtube.com/watch?v=EDsK-8SBx-g&ab_channel=InternationalMathematicalUnion

[Co] B. Conrad: *The flat deformation functor,* in: *Modular forms and Fermat's Last Theorem,* G.Cornell, J.Silverman, G.Stevens edts., Springer-Verlag 1997, pp. 373-420.

[De71] P. Deligne: *Formes modulaires et représentations ℓ-adiques,* in: Lecture Notes in Math. **179,** Springer-Verlag, New-York, Berlin, Heidelberg, 1971, pp. 139-172,

[De-Se74] P. Deligne, J.-P. Serre: *Formes modulaires de poids 1,* Ann. Sci. ENS **7** (1974), 507-530.

[dS] E. de Shalit: *Hecke rings and universal deformation rings,* in: *Modular forms and Fermat's Last Theorem,* G.Cornell, J.Silverman, G.Stevens edts., Springer-Verlag 1997, pp. 421-445.

[Di93] F. Diamond: *The refined conjecture of Serre*, in: *Elliptic Curves, Modular Forms, and Fermat's Last Theorem*, J. Coates, S.T. Yau edts., International Press 1995, pp. 22-37.

[Di96] F. Diamond: *On deformation rings and Hecke rings,* Ann. Math. **144** (1996), 137-166.

[Di-Ri] F. Diamond, K. Ribet: *ℓ-adic modular deformations and Wiles' "main conjecture",* in *Elliptic Curves, Modular Forms, and Fermat's Last Theorem*, J. Coates, S.T. Yau edts., International Press 1995, pp. 357-371.

[D-D-T] H. Darmon, F. Diamond, R. Taylor: *Fermat's Last Theorem,* in: Current developments in Mathematics, 1995, International Press, pp. 1-107. A more complete version is on Darmon's web-page.

[FGA] B. Fantechi et al.: *Fundamental Algebraic Geometry, Grothendieck's FGA Explained,* Math. Surveys and Monographs **123**, AMS, 2005.

[F-L] J.-M. Fontaine, G. Laffaille: *Construction de représentations p-adiques,* Ann. Sci. ENS **15** (1982), 547-608.

[Gr] B. Gross: *A tameness criterion for Galois representations associated to modular forms mod p,* Duke Math. J. **61** (1990), 445-517.

[Maz] B. Mazur: *Deforming Galois representations,* in: Galois groups over $\mathbb{Q}$, Y. Ihara, K. Ribet, J.-P. Serre eds., MSRI Publ. **16**, Springer-Verlag, 1989, pp. 385-437.

[Mi] J. Milne: *Arithmetic Duality Theorems,* Perspectives in Mathematics **1**, Academic Press, 1986.

[N-S-W] J. Neukirch, A. Schmidt, K. Wingberg: *Cohomology of Number Fields*, Grundlehren der mathematischen Wissenscahften **323**, Springer-Verlag, 2000.

[Ra] R. Ramakrishna: *On a variation of Mazur's deformation functor,* Comp. Math. **87** (1993), 269-286.

[Ray] M. Raynaud: *Schémas en groupes de type (p,p,...,p),* Bull. Soc. Math. France **102** (1974), 241-280.

[Ri90] K. Ribet: *On modular representations of* $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ *arising from modular forms,* Invent. Math. **100** (1990), 431-476.

[Sh58] G. Shimura: *Correspondances modulaires et les fonctions* ζ *de courbes algébriques*, Journal of the Mathematical Society of Japan **10** (1958), 1–28.

[Sh71] G. Shimura: *Introduction to the arithmetic theory of automorphic functions*, Publ. of Math. Soc. of Japan, **11**, 1971.

[St97] G. Stevens: *An overview of the proof of Fermat's Last Theorem,* in: *Modular forms and Fermat's Last Theorem,* G.Cornell, J.Silverman, G.Stevens edts., Springer-Verlag 1997, pp. 1-15.

[T-W95] R. Taylor, A. Wiles: *Ring-theoretic properties of certain Hecke algebras*, Ann. Math. **141** (1995), 553-572.

[Wa] L. Washington: *Galois cohomology,* in: *Modular forms and Fermat's Last Theorem,* G.Cornell, J.Silverman, G.Stevens edts., Springer-Verlag 1997, pp.101-120.

[We67] A. Weil: *Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen,* Math. Ann. **168** (1967), 149-156.

[W95] A. Wiles: *Modular elliptic curves and Fermat's Last Theorem*, Ann. Math. **141** (1995), 443-551.