RATIONAL ELLIPTIC CURVES ARE MODULAR after Breuil, Conrad, Diamond and Taylor

by Bas EDIXHOVEN

1. INTRODUCTION

In 1994, Wiles and Taylor-Wiles proved that every semistable elliptic curve over \mathbb{Q} is modular, in the sense that it is a quotient of the jacobian of some modular curve (see [64], [60]). This work has been reported upon in this seminar in [50] and [41]; see especially [50, §1.2] for a historical account. As a consequence, Fermat's Last Theorem, known to be a consequence of this modularity result since work of Ribet based on a conjecture of Serre (see [40]), was finally proved. For a more detailed account of all this, see the book [15], and also [17]. Since 1994, this modularity result has been generalized by an increasing sequence of groups of authors: [24], [14], and [4].

Theorem 1.1 (Diamond). — Every elliptic curve over \mathbb{Q} that is semistable at 3 and 5 is modular.

THEOREM 1.2 (Conrad, Diamond, Taylor). — Every elliptic curve over \mathbb{Q} that acquires semistable reduction over a tame extension of \mathbb{Q}_3 is modular.

Theorem 1.3 (Breuil, Conrad, Diamond, Taylor). — Every elliptic curve over \mathbb{Q} is modular.

The method of the proofs is basically that of Wiles, i.e., for a given elliptic curve E over \mathbb{Q} one tries to prove that the mod l Galois representation $\overline{\rho}_{E,l}$ on $E(\overline{\mathbb{Q}})[l]$ is modular for some prime number l, and then that all lifts of $\overline{\rho}_{E,l}$ to l-adic representations of a suitable type are modular. The second step involves studying deformations of Galois representations, the systematic theory of which was initiated by Mazur, triggered by work of Hida. The key result for the first step is the celebrated theorem of Langlands [36] and Tunnell [61] that says that $\overline{\rho}_{E,3}$ is modular, as $\mathrm{GL}_2(\mathbb{F}_3)$ is solvable and has a faithful two-dimensional complex representation. The complications that arise in the proofs of the theorems above simply come from having to prove results as in Wiles and Taylor—Wiles, but with fewer hypotheses. In particular, choosing the right deformations of the

restriction of $\overline{\rho}_{E,l}$ to $G_l := \operatorname{Gal}(\overline{\mathbb{Q}}_l/\mathbb{Q}_l)$ becomes much more complicated if E does not have semistable reduction at l.

The aim of this report is to give a reasonable sketch of the proofs of the theorems above, to describe the relation to some conjectures by Fontaine and Mazur and by Langlands, and to mention some related results. For some applications of the modularity results above, we refer to [16]. The author of this report does not claim to have checked the computations in [4], but he has studied [4] quite seriously and has not encountered any real problem. Let us also state the following theorem (Theorem B of [4]), whose proof is intricately linked to that of Theorem 1.3 above.

THEOREM 1.4 (Breuil, Conrad, Diamond, Taylor). — Every irreducible continuous representation $\overline{\rho}: G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_5)$ with cyclotomic determinant is modular.

2. RELATION WITH CONJECTURES BY LANGLANDS, FONTAINE AND MAZUR

The Langlands program predicts, among many other things, that all L-functions coming from algebraic geometry are in fact automorphic, i.e., arise from automorphic representations. More precisely, every absolutely irreducible motive of rank n over a number field F and with coefficients in a subfield E of $\overline{\mathbb{Q}}$ should correspond to a cuspidal algebraic automorphic representation of $\mathrm{GL}_n(\mathbb{A}_F)$ with coefficients in E: see [10, Question 4.16], and the paragraph after that.

In that paragraph, Clozel explains how this conjecture relates to the conjecture of Hasse–Weil type that says that the L-function of such a motive extends meromorphically to all of \mathbb{C} and satisfies a certain functional equation. He finishes by remarking that the only cases for which the Hasse–Weil conjecture has been proved are cases where one actually proves the stronger conjecture, i.e., the existence of an automorphic representation; this remains true after the work of Wiles and its generalizations.

Of course, if E is an elliptic curve over \mathbb{Q} , the Langlands program predicts that E is modular. Hence the modularity theorem for elliptic curves over \mathbb{Q} is just a tiny part of the Langlands program.

Fontaine and Mazur stated the following conjecture (Conjecture 1 of [32]).

Conjecture 2.1. — Let l be prime, $n \geq 0$, and let $\rho: \operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \operatorname{GL}_n(\mathbb{Q}_l)$ be an irreducible continuous representation. Then ρ is isomorphic to a subquotient of some etale cohomology group $\operatorname{H}^i(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_l(r))$ with X a smooth projective variety over \mathbb{Q} , if and only if ρ satisfies the following two conditions:

- (1) ρ is ramified at only finitely many primes;
- (2) the restriction $\rho|_{G_l}$ to a decomposition group at l is potentially semistable (see [31] for this notion).

In one direction, this conjecture has been proved: the $H^i(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_l(r))$ are known to be unramified at almost all primes, and the restriction to G_l is known to be potentially semistable by work of Tsuji and de Jong (see [1, §6.3.3]). It is the other direction that is even more spectacular: it is amazing that just these two conditions should imply, for example, that the Frobenius elements at almost all primes have eigenvalues that are algebraic numbers, and even Weil numbers, and that ρ should be part of a compatible system of l-adic representations. The evidence that one has today for this direction of the conjecture consists of the potentially abelian cases (treated in [32, §6]; ρ occurs in the tensor category generated by representations with finite image and representations which arise from potentially CM abelian varieties), and the cases treated by Wiles' method. However, see [39] for a representation that does satisfy the two conditions above, but for which one does not know if it satisfies Conjecture 2.1.

Combined with the Langlands program, Conjecture 2.1 implies (Conjecture 3c of [32]) that every 2-dimensional ρ satisfying the two conditions, up to Tate twist, either has a finite image, or arises from a modular form of weight at least two.

Since the space of modular forms of a given weight and level is finite dimensional, one also expects certain finiteness results concerning ρ as in Conjecture 2.1, which become semistable over a given extension of \mathbb{Q} , and are of fixed Hodge-Tate type: see [32, §3]. Most of [32] is in fact concerned with a deformation theoretic study of these finiteness conjectures.

Suppose now that l > 2. For a given absolutely irreducible continuous $\overline{\rho} \colon G_{\mathbb{Q}} \to \operatorname{GL}_2(\mathbb{F}_l)$ one considers all lifts $\rho \colon G_{\mathbb{Q}} \to \operatorname{GL}_2(\mathbb{Z}_l)$ that are unramified outside a fixed set of primes. The l-adic variety (over \mathbb{Q}_l) of such lifts is conjecturally three dimensional. Now suppose moreover that $\overline{\rho}|_{G_l}$ is absolutely irreducible. Then the variety of lifts of $\overline{\rho}|_{G_l}$ is smooth and of dimension five by [42, Thm 4.1]. Since one expects the locus of global lifts that are potentially semistable and of a given type (i.e., Hodge-Tate type at l, and semistable over a fixed extension of \mathbb{Q}) to be zero dimensional, one expects that the locus of such local lifts is of codimension three in the five dimensional variety. Indeed, in the crystalline case with Hodge-Tate weights in the interval [0, l-1], this was proved in [42] (moreover, the two-dimensional space is smooth). We note that, by [3], "potentially Barsotti-Tate" is equivalent to "potentially crystalline with Hodge-Tate weights in [0, 1]" (we recall that l > 2).

Of course, the computations done by Ramakrishna and by Fontaine and Mazur are not directly in terms of representations of G_l . Ramakrishna uses the results of Fontaine and Laffaille, and Fontaine and Mazur work with filtered (ϕ, N) -modules. We note that by recent work of Colmez and Fontaine, [11], one actually has an equivalence of tensor categories between semistable l-adic representations of G_l and weakly admissible filtered (ϕ, N) -modules, which makes it possible to translate problems on the Galois side into

problems in linear algebra, even more than before the equivalence between "weakly admissible" and "admissible" was known. On the other hand, what is still not available in this generality is a theory that works for \mathbb{Z}_l -lattices instead of \mathbb{Q}_l -vector spaces.

3. REVIEW OF WILES' METHOD

Before turning to the work of Breuil, Conrad, Diamond and Taylor, let us review Wiles' method. Good references for this part are [17], [15], [50], [41], and of course [64] (the introduction of which gives the story of the proof) and [60]. For simplicity, we only discuss this method in a case that suffices for modularity of semistable elliptic curves.

Let E be a semistable elliptic curve over \mathbb{Q} . The first observation is that there are many elliptic curves E' over \mathbb{Q} such that E[5] and E'[5] are symplectically isomorphic; this is due to the fact that the modular curve that parameterizes such E' (over \mathbb{Q} -schemes) is a non-empty open part of $\mathbb{P}^1_{\mathbb{Q}}$. One proves that there is such an E', semistable, and such that the representation $\overline{\rho}_{E',3} \colon G_{\mathbb{Q}} \to \operatorname{Aut}(E'[3](\overline{\mathbb{Q}})) \cong \operatorname{GL}_2(\mathbb{F}_3)$ is surjective (see [50, §3]). By Langlands and Tunnell, the representation $\overline{\rho}_{E',3}$ is modular. The aim is now to show that $\rho_{E',3} \colon G_{\mathbb{Q}} \to \operatorname{Aut}(E'(\overline{\mathbb{Q}})[3^{\infty}]) \cong \operatorname{GL}_2(\mathbb{Z}_3)$ is modular, by showing that all 3-adic lifts of $\overline{\rho}_{E',3}$ with reasonable properties are modular, and hence so is E'. Before we discuss how that works, let us see how one then establishes the modularity of E itself.

Of course, if $\overline{\rho}_{E,3}$ is surjective, then we could have taken E'=E, so let us assume that $\overline{\rho}_{E,3}$ is not surjective. Then E[3] is in fact reducible (this uses the semistability at all primes; see [50, Proposition 1]). But then $\overline{\rho}_{E,5}$ is irreducible, or $E_{\overline{\mathbb{Q}}}$ is isogeneous to the elliptic curve E_1 over $\overline{\mathbb{Q}}$ that has j-invariant $-5 \cdot 29^3/2^5$, as one sees by looking at the modular curve $X_0(15)$, which has genus one and exactly eight rational points, four of which are cusps (see [46, §2.1]). The elliptic curve E_1 has a model over \mathbb{Q} with conductor 50, which can be checked to be modular. Since modularity is invariant under isogeny and twisting, we may now assume that $\overline{\rho}_{E,5}$ is irreducible, and hence surjective ([50, Proposition 1]). In this case, we already know that $\overline{\rho}_{E,5}$ is modular, because E' is, and one proves the same type of result for modularity of 5-adic liftings as in the 3-adic case.

Let us now give a precise statement of these lifting results. We need some terminology and notation, adapted to the type of representations that we are interested in, i.e., those coming from modular forms of weight two. For each prime p, we choose an embedding $\overline{\mathbb{Q}} \to \overline{\mathbb{Q}}_p$, and we let G_p and I_p denote the corresponding decomposition and inertia subgroups of $G_{\mathbb{Q}}$. We let $\varepsilon \colon G_{\mathbb{Q}} \to \mathbb{Z}_l^*$ denote the l-cyclotomic character, given by the action on the elements of l-power order in $\overline{\mathbb{Q}}^*$.

DEFINITION 3.1. — Let l be a prime number, and k a finite field of characteristic l. Let R be a complete local noetherian ring with residue field k, and let M be a free R-module of rank 2 with a continuous action by $G_{\mathbb{Q}}$; a choice of basis then gives a continuous

representation $\rho \colon G_{\mathbb{Q}} \to \operatorname{GL}_2(R)$. For p prime and different from l, M is called semistable at p if, with respect to a suitable basis, $\rho|_{I_p}$ is of the form $\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix}$. The representation M is called Barsotti-Tate (at l) if for each finite quotient \overline{M} of M there exists a finite group scheme \overline{M} over \mathbb{Z}_l such that M and $\overline{M}(\overline{\mathbb{Q}}_l)$ are isomorphic as $\mathbb{Z}_l[G_l]$ -modules. The representation M is called semistable at l if it is Barsotti-Tate or if, with respect to a suitable basis, $\rho|_{I_p}$ is of the form $\begin{pmatrix} \varepsilon & * \\ 0 & 1 \end{pmatrix}$.

THEOREM 3.2 (Wiles, Taylor-Wiles). — Let $l \neq 2$ be a prime number. Let K be a finite extension of \mathbb{Q}_l , O its ring of integers, and k its residue field. Let $\rho: G_{\mathbb{Q}} \to \mathrm{GL}_2(O)$ be an odd continuous representation such that:

- (1) its reduction $\overline{\rho}$: $G_{\mathbb{Q}} \to \operatorname{GL}_2(k)$ is modular and its restriction to the quadratic subfield of $\mathbb{Q}(\mu_l)$ is absolutely irreducible;
- (2) $\rho|_{G_l}$ is semistable;
- (3) ρ is ramified at only finitely many primes;
- (4) $\det(\rho) = \varepsilon$;
- (5) for every $p \equiv -1 \mod l$ such that $\overline{\rho}|_{I_p}$ is reducible, $\overline{\rho}|_{G_p}$ is reducible too.

Then ρ is modular.

In view of what has been said above, this result implies that all semistable elliptic curves over \mathbb{Q} are modular. Wiles' strategy to prove Theorem 3.2 is to compare systematically all deformations of $\overline{\rho}$ with certain properties when restricted to decomposition groups to those coming from modular forms of a given level. For simplicity, we will now assume that ρ is semistable at all primes, and follow the exposition in [41], with some modifications, anticipating our discussion of [24], [14] and [4].

So suppose that $\overline{\rho}$ is as in Theorem 3.2, and moreover that ρ is semistable at all primes. We will now forget about ρ , for the moment, but keep $\overline{\rho}$. So $\overline{\rho}$ is a continuous representation of $G_{\mathbb{Q}}$ on a 2-dimensional k-vector space, with k a finite extension of \mathbb{F}_l with $l \neq 2$, and has the following properties: it is modular, absolutely irreducible after restriction to the quadratic subfield of $\mathbb{Q}(\mu_l)$, semistable at all primes, and $\det(\overline{\rho}) = \overline{\varepsilon}$. As nothing about these hypotheses changes if we replace k by a finite extension of it, we may suppose, for example, that the characteristic polynomials of the $\overline{\rho}(\sigma)$, σ in $G_{\mathbb{Q}}$, are all split. Let O be the ring of integers in a finite extension K of \mathbb{Q}_l , with residue field k. (Later in the proof, we need a modular form of "minimal level" giving rise to $\overline{\rho}$, and with coefficients in O.) For any finite set Σ of primes we define two O-algebras $R_{O,\Sigma}$ and $\mathbb{T}_{O,\Sigma}$, as follows.

DEFINITION 3.3. — Let R be a complete local noetherian O-algebra with residue field k. A deformation of $\overline{\rho}$ to R is a free R-module M of rank two, with a continuous $G_{\mathbb{Q}}$ action, such that $k \otimes_R M$ is isomorphic to $\overline{\rho}$. A deformation ρ is said to be of type Σ if $\det(\rho) = \varepsilon$, and ρ is semistable at l and "minimally ramified" outside Σ :

(1) if $l \notin \Sigma$ and $\overline{\rho}$ is Barsotti-Tate, then ρ is Barsotti-Tate;

- (2) if $p \notin \Sigma \cup \{l\}$ and $\overline{\rho}$ is unramified at p, then ρ is unramified at p;
- (3) if $p \notin \Sigma \cup \{l\}$ and $\overline{\rho}$ is ramified at p (and hence semistable, with our hypotheses), then ρ is semistable at p.

With these definitions, there is, for each Σ , a universal deformation ring $R_{O,\Sigma}$ that represents the functor that sends R to the set of isomorphism classes of deformations of type Σ over R. A very good reference for this is [20]. If $K \to K'$ is a finite extension, then $R_{O',\Sigma} = O' \otimes_O R_{O,\Sigma}$.

Let us now turn to the definition of $\mathbb{T}_{O,\Sigma}$. The reader is referred to Appendix A for certain properties of the Galois representation ρ_f associated to a modular form f of weight two with coefficients in $\overline{\mathbb{Q}}_l$. We define \mathcal{N}_{Σ} to be the set of weight two newforms f with coefficients in $\overline{\mathbb{Q}}_l$ such that $\rho_f \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(O_f)$ is of type Σ , where O_f is the sub-O-algebra of $\overline{\mathbb{Q}}_l$ generated by the coefficients of f. The results in Appendix A imply that there is an integer N_{Σ} , such that for each f in \mathcal{N}_{Σ} , the level of f divides N_{Σ} . This implies that \mathcal{N}_{Σ} is a finite set. One can take N_{Σ} as follows:

$$N_\Sigma := l^\delta \cdot \prod_{p \mid N(\overline{
ho})} p \cdot \prod_{p \in \Sigma - \{l\}} p^2,$$

where δ is 0 if $\overline{\rho}$ is Barsotti-Tate and l not in Σ , and δ is 1 otherwise, and where $N(\overline{\rho})$ is the level associated to $\overline{\rho}$ by Serre in [49] (i.e., $N(\overline{\rho})$ is given by the usual formula for the Artin conductor of a representation, in terms of the ramification subgroups at all $p \neq l$). The reason that such a N_{Σ} suffices is that the wild parts of the conductors of ρ_f and $\overline{\rho}_f$ are equal. For simplicity, we will now only consider Σ that do not contain primes dividing $N(\overline{\rho})$ (this suffices for the application to semistable lifts of $\overline{\rho}$). For each f in \mathcal{N}_{Σ} , we have a morphism $R_{O,\Sigma} \to O_f$, and we define $\mathbb{T}_{O,\Sigma}$ to be the image of $R_{O,\Sigma}$ in the product of the O_f . Since R_{Σ} is generated, as O-algebra, by the traces of elements in the universal representation, $\mathbb{T}_{O,\Sigma}$ is generated by the elements $a_p \colon f \mapsto a_p(f)$ for p not dividing lN_{Σ} .

The method of Wiles is now to show that the surjections $R_{O,\Sigma} \to \mathbb{T}_{O,\Sigma}$ are isomorphisms, by studying how they change as Σ varies. The first step in this is to use what has been proved about Serre's conjectures on modularity of mod l representations in [49]: \mathcal{N}_{\emptyset} is not empty (see [45] and [22]). This implies that we can suppose (and we will) that we have a section $\pi = \pi_{\emptyset} \colon \mathbb{T}_{\emptyset} \to O$. We let P_{Σ} denote the corresponding O-valued points of $\operatorname{Spec}(\mathbb{T}_{O,\Sigma})$ and $\operatorname{Spec}(R_{O,\Sigma})$, for each Σ . Wiles introduced the following O-modules associated to each Σ : on the one hand the cotangent spaces at the P_{Σ} , i.e., $P_{\Sigma}^*\Omega^1_{R_{O,\Sigma}/O}$ and $P_{\Sigma}^*\Omega^1_{\mathbb{T}_{O,\Sigma}/O}$, and on the other hand the "module of congruences" $P_{\Sigma}^*\mathcal{O}_{Z_{\Sigma}}$, defined as follows. Since $\mathbb{T}_{O,\Sigma}$ is finite free as O-module, $\operatorname{Spec}(\mathbb{Q}\otimes\mathbb{T}_{O,\Sigma})$ is the disjoint union of two open and closed subschemes $P_{\Sigma,K}$ and $Z_{\Sigma,K}$, with $P_{\Sigma,K}$ consisting of the point $P_{\Sigma}(\operatorname{Spec}(K))$. We let Z_{Σ} be the scheme theoretic closure of $Z_{\Sigma,K}$ in $\operatorname{Spec}(\mathbb{T}_{O,\Sigma})$ (note that the $\mathbb{T}_{O,\Sigma}$ are reduced by construction). These modules, that will intervene only via their lengths, are usually introduced as $\ker(P_{\Sigma}^*)/\ker(P_{\Sigma}^*)^2$ and $P_{\Sigma}^*\operatorname{Ann}_{\mathbb{T}_{O,\Sigma}}(\ker(P_{\Sigma}^*))$. (This last module has finite length if and only if $\operatorname{Spec}(\mathbb{Q}\otimes\mathbb{T}_{O,\Sigma})$ is reduced at $P_{\Sigma,K}$.)

The fact that $R_{O,\Sigma}$ represents the functor of isomorphism classes of deformations of type Σ implies the following Galois cohomological description of $P^*\Omega^1_{R_{O,\Sigma}/O}$:

$$\operatorname{Hom}_{O}(P^{*}\Omega^{1}_{R_{O,\Sigma}/O}, K/O) = \operatorname{H}^{1}_{\Sigma}(G_{\mathbb{Q}}, \operatorname{ad}^{0}(\rho) \otimes K/O),$$

where ρ is the representation corresponding to $P = P_{\Sigma}$, where $\operatorname{ad}^{0}(\rho)$ is the representation of $G_{\mathbb{Q}}$ on the sub-O-module of trace zero elements of $\operatorname{End}_{O}(M_{\rho})$, and where $\operatorname{H}^{1}_{\Sigma}(G_{\mathbb{Q}},\operatorname{ad}^{0}(\rho)\otimes K/O)$ denotes the subgroup of $\operatorname{H}^{1}(G_{\mathbb{Q}},\operatorname{ad}^{0}(\rho)\otimes K/O)$ of classes that map, at all p, to the subgroups $L_{\Sigma,p}$ of the $\operatorname{H}^{1}(G_{p},\operatorname{ad}^{0}(\rho)\otimes K/O)$ that reflect the conditions for deformations to be of type Σ . To be explicit, these $L_{\Sigma,p}$ are:

- $L_{\Sigma,p} = \mathrm{H}^1(G_p/I_p, (\mathrm{ad}^0(\rho) \otimes K/O)^{I_p})$ if $p \notin \Sigma \cup \{l\}$;
- $L_{\Sigma,p} = H^1(G_p, ad^0(\rho) \otimes K/O)$ if $p \in \Sigma$ and $p \neq l$;
- $L_{\Sigma,l}$ is the subspace of $H^1(G_l, ad^0(\rho) \otimes K/O)$ that corresponds to deformations that are Barsotti-Tate, if $l \notin \Sigma$;
- $L_{\Sigma,l}$ is the subspace of $H^1(G_l, ad^0(\rho) \otimes K/O)$ that corresponds to deformations that are semistable at l, if $l \in \Sigma$.

The results of Poitou-Tate on local duality and global Euler characteristic show that, for M a finite discrete $G_{\mathbb{Q}}$ -module, with a Selmer datum $L_v \subset H^1(G_v, M)$ at all places v of \mathbb{Q} , one has:

$$\frac{\#\mathrm{H}^{1}_{L}(G_{\mathbb{Q}}, M)}{\#\mathrm{H}^{1}_{L^{\perp}}(G_{\mathbb{Q}}, M^{*})} = \frac{\#\mathrm{H}^{0}(G_{\mathbb{Q}}, M)}{\#\mathrm{H}^{0}(G_{\mathbb{Q}}, M^{*})} \cdot \prod_{v} \frac{\#L_{v}}{\#\mathrm{H}^{0}(G_{v}, M)},$$

where M^* is the Cartier dual $\operatorname{Hom}(M, \overline{\mathbb{Q}}^*)$ of M, and where, for each v, L_v^{\perp} is the orthogonal of L_v . Moreover, if $L \subset L'$ are two Selmer data for M, then one has an exact sequence:

$$0 \to \mathrm{H}^1_L(G_{\mathbb{Q}}, M) \to \mathrm{H}^1_{L'}(G_{\mathbb{Q}}, M) \to \prod_v L'_v/L_v \to \mathrm{H}^1_{L^{\perp}}(G_{\mathbb{Q}}, M^*)^{\vee} \to \mathrm{H}^1_{L'^{\perp}}(G_{\mathbb{Q}}, M^*)^{\vee} \to 0.$$

Having established this, Wiles first proves that $R_{O,\emptyset} \to \mathbb{T}_{O,\emptyset}$ is an isomorphism, and then that this remains so if one enlarges Σ . The argument for the first step is really amazing, he somehow manages to "patch", for a suitable sequence of Σ_n , the R_{O,Σ_n} and the \mathbb{T}_{O,Σ_n} into power series rings, with the same number of generators, and deduce from that that $R_{O,\emptyset} \to \mathbb{T}_{O,\emptyset}$ is an isomorphism. (This patching argument was introduced in [60], and used to show that $\mathbb{T}_{O,\emptyset}$ is a complete intersection, but Faltings pointed out that one could also use the argument directly in proving $R_{O,\emptyset} \to \mathbb{T}_{O,\emptyset}$ to be an isomorphism.) We will now take a closer look at this argument, in order to see which conditions have to be satisfied by the type of deformations that one considers for it to work.

So suppose that one wants to do this argument for $R_{O,\Sigma}$. The primes p that one wants to add to Σ are congruent to 1 modulo a high power of l, and such that $\overline{\rho}$ is unramified at p with distinct Frobenius eigenvalues in k^* . For such a p, and for Σ' containing p, $\rho_{O,\Sigma'}^{\text{univ}}|_{G_p}$ is a direct sum of two characters, whose restrictions to I_p factor through the l-part Δ_p of $\text{Gal}(\mathbb{Q}_p^{\text{unr}}(\mu_p)/\mathbb{Q}_p^{\text{unr}}) = (\mathbb{Z}/p\mathbb{Z})^*$. Choosing one of the two Frobenius eigenvalues gives

 $R_{O,\Sigma'}$ the structure of an $O[\Delta_p]$ -algebra. The Σ' that one wants to consider are of the form $\Sigma' = \Sigma \cup Q$, with Q a set of r elements, for some fixed r, and such that $R_{O,\Sigma'}$ can also be topologically generated by r elements. (Note that $R_{O,\Sigma'}$ is an algebra over $O[\Delta_Q]$ with $\Delta_Q = \prod_{p \in Q} \Delta_p$, and that $O[\Delta_Q]$ looks more and more as a power series ring in r variables, as the primes p are closer to 1, l-adically.) Let L and L' be the Selmer data corresponding to Σ and Σ' . Then $\dim_k \prod_v L'_v/L_v = r$, so one finds, by the exact sequence above, that $\dim_k \mathrm{H}^1_{L^\perp}(G_\mathbb{Q}, \mathrm{ad}^0(\overline{\rho})^*) \geq \dim_k(\mathrm{H}^1_L(G_\mathbb{Q}, \mathrm{ad}^0(\overline{\rho})))$. But in the displayed formula above, one has $\dim_k L_p \geq \dim_k \mathrm{H}^1(G_p/I_p, \mathrm{ad}^0(\overline{\rho})^{I_p}) = \dim_k(\mathrm{H}^0(G_p, \mathrm{ad}^0(\overline{\rho}))$, for all $p \neq l$, whereas $\dim_k L_\infty = 0$ and $\dim_k \mathrm{H}^0(G_\infty, \mathrm{ad}^0(\overline{\rho})) = 1$. Moreover, in that formula one has $\#\mathrm{H}^0(G_\mathbb{Q}, M) = 1$ (since $\overline{\rho}$ is absolutely irreducible), and $\#\mathrm{H}^0(G_\mathbb{Q}, M^*) = 1$ (since the restriction of $\overline{\rho}$ to the quadratic subfield of $\mathbb{Q}(\mu_l)$ is absolutely irreducible). It follows that this setup can only work if $L_p = \mathrm{H}^1(G_p/I_p, \mathrm{ad}^0(\overline{\rho})^{I_p})$ for all $p \neq l$, and $\dim_k(L_l) \leq 1 + \dim_k \mathrm{ad}^0(\overline{\rho})^{G_l}$. This means that Σ must be \emptyset , and that L_l must be of dimension 1, unless $\overline{\rho}|_{I_l} \cong \overline{\varepsilon} \oplus 1$. This last condition puts a very strong restriction on the type of local deformations at l that one can use.

In order to prove that $R_{O,\emptyset} \to \mathbb{T}_{O,\emptyset}$ is an isomorphism, Taylor and Wiles use that, in their situation, the localization at $\overline{\rho}$ of $H^1(X_0(N_\emptyset)(\mathbb{C}), O)$ is a free $\mathbb{T}_{O,\emptyset}$ -module, and similarly for the Σ 's that they choose. Such results are quite delicate to prove. In the next section we will discuss how Diamond and Fujiwara have gotten around this, and actually obtain such freeness results as a consequence of the method. In [17], the freeness assumption is not used, but the given proof still relies on q-expansions (see [17, Remark 4.15]).

Let us now briefly discuss how Wiles proved that the $R_{O,\Sigma} \to \mathbb{T}_{O,\Sigma}$ are isomorphisms. This is done by induction on the number of elements of Σ , but, in order to carry out this induction, one actually proves more, namely, that these O-algebras are complete intersections. Indeed, Wiles found a criterion for doing the induction, in terms of the changes of the lengths of $P^*\Omega^1_{R_{O,\Sigma}/O}$ and $P^*\mathcal{O}_{Z_{\Sigma}}$ when comparing between Σ and $\Sigma' := \Sigma \cup \{p\}$. On the Galois side, the exact sequence above gives an upper bound for the length of $P^*\Omega^1_{R_{O,\Sigma'}/O}$. On the Hecke side, [17, §4.4] gives a new proof of the lower bound for the length of $P^*\mathcal{O}_{Z_{\Sigma}'}$ that was proved by Wiles. This proof does not use freeness, and it nicely relates this change of length to the residue at 2 of the L-function of the symmetric square of the system of representations associated to P, giving a relation to the Bloch-Kato conjectures. Wiles' argument, which is to compute the composite $J_0(N_{\Sigma}) \to J_0(N_{\Sigma'}) \to J_0(N_{\Sigma})$, is also sketched in [17, §4.4].

4. IMPROVEMENTS OF THE COMMUTATIVE ALGEBRA PART

The results in commutative algebra that are used in [14] and [4] are improvements of those in [64] and [60]. These improvements were found independently by Diamond [25] and

Fujiwara [33], motivated by Fujiwara's work on modularity over totally real number fields. We also note that Lenstra, de Smit, Rubin, and Schoof have established isomorphism and complete intersection criteria as in Wiles, without the Gorenstein hypothesis, and without the limiting process, see [21]. Let us now state the criteria as in [25, Theorems 2.1 and 2.4].

THEOREM 4.1. — Let k be a finite field, and $r \ge 0$ an integer. Let $A := k[[S_1, \ldots, S_r]]$, $B := k[[X_1, \ldots, X_r]]$, let R be a k-algebra, and let H be a non-zero R-module that has finite k-dimension. Suppose that for every $n \ge 1$ one has a commutative diagram:

$$\begin{array}{ccc}
A & \xrightarrow{\phi_n} & B \\
\downarrow & & \downarrow \psi_n \\
k & \longrightarrow & R
\end{array}$$

and a B-module H_n with a morphism $\pi_n \colon H_n \to H$ such that as an A-module, H_n is free over A/m_A^n , and such that the morphism $k \otimes_A H_n \to H$ induced by π_n is an isomorphism. Then H is free over R, and R is a (zero dimensional) complete intersection.

In the application of this result, k is as above, A is a projective limit of k-algebras of the form $k[\Delta_Q]$, with Q a set of r distinct primes $p \equiv 1 \mod l^n$ and Δ_Q the product of the $(\mathbb{Z}/p\mathbb{Z})_l^*$, B is a projective limit of $k \otimes_O R_{O,Q}$'s, $R = k \otimes_O R_{O,\emptyset}$, and H and H_n come from (co)homology groups of modular curves. The freeness of H_n over A/m_A^n basically comes from standard facts about cohomology of locally constant sheaves and unramified covers of affine Riemann surfaces. The Hecke algebra $k \otimes_O \mathbb{T}_{O,\emptyset}$ is the image T of R in $\operatorname{End}_k(M)$, so the conclusion that H is free over R implies that R = T. The freeness version of Wiles' numerical criterion is as follows.

THEOREM 4.2. — Let O be a complete discrete valuation ring with finite residue field k, and let R be a complete noetherian local O-algebra. Let H be an R-module, finite free over O, let $\phi: R \to T$ be the quotient by $\operatorname{Ann}_R(H)$, and suppose that T has a section $\pi_T \colon T \to O$. Put $\pi_R := \phi \pi_T$. Define $\Omega := H/(H[\ker(\pi_T)] + H[\operatorname{Ann}_T(\ker(\pi_T))])$. Let d be the O-rank of $H[\ker(\pi_R)]$. If Ω has finite length over O, then the following are equivalent:

- (1) $\operatorname{rank}_O H \leq d \cdot \operatorname{rank}_O T$ and $\operatorname{length}_O \Omega \geq d \cdot \operatorname{length}_O(\ker(\pi_R) / \ker(\pi_R)^2);$
- (2) $\operatorname{rank}_O H = d \cdot \operatorname{rank}_O T$ and $\Omega \cong (O/\operatorname{Fitt}_O(\ker(\pi_R)/\ker(\pi_R)^2))^d$;
- (3) R is a complete intersection and H is free (of rank d) over R.

5. THE WORK OF BREUIL, CONRAD, DIAMOND, AND TAYLOR

We are now ready to discuss the work of the four authors mentioned above in [24], [14], and [4]. Before getting into any details, let us see what problems were solved in each of these three articles. In [24, Theorem 5.3], Diamond gets rid of condition (5) in Theorem 3.2. To be precise, let us state his result.

THEOREM 5.1 (Diamond). — Let l > 2 be prime, K a finite extension of \mathbb{Q}_l , O its ring of integers, k its residue field, and $\rho: G_{\mathbb{Q}} \to \mathrm{GL}_2(O)$ an odd continuous representation such that:

- (1) its reduction $\overline{\rho} \colon G_{\mathbb{Q}} \to \operatorname{GL}_2(k)$ is modular, and its restriction to the quadratic subfield of $\mathbb{Q}(\mu_l)$ is absolutely irreducible;
- (2) $\rho|_{G_l}$ is Barsotti-Tate and $\det(\rho)|_{I_l} = \varepsilon|_{I_l}$, or, with respect to a suitable basis, $\rho|_{G_l}$ is of the form $\begin{pmatrix} \phi & * \\ 0 & \psi \end{pmatrix}$, with ψ unramified, $\overline{\psi} \neq \overline{\phi}$, and $\phi|_{I_l} = \chi \varepsilon^{k-1}|_{I_l}$ for some integer $k \geq 2$ and χ of finite order;
- (3) ρ is ramified at only finitely many primes.

Then ρ is modular.

We should note here that Theorem 3.2 is weaker than the result that was proved by Wiles. What is proved in [64] is the theorem above, with the extra condition (5) of Theorem 3.2. Let us now explain what the problem is in a case that does not satisfy this condition (5).

So suppose that $\overline{\rho}$ satisfies the conditions of Theorem 5.1, that $p \neq l$, that $\overline{\rho}_p = \overline{\rho}|_{G_p}$ is irreducible, but $\overline{\rho}|_{I_p}$ is reducible. Then $\overline{\rho}_p$ is of the form $\operatorname{Ind}_{\mathbb{Q}_{p^2}}^{\mathbb{Q}_p} \psi$, with \mathbb{Q}_{p^2} the unramified extension of degree two of \mathbb{Q}_p , and $\psi \colon G_{\mathbb{Q}_{p^2}} \to k^*$ a continuous character such that $\psi^\sigma \neq \psi$. (To prove this, note that $\overline{\rho}(I_p)$ must have exactly two fixed points on $\mathbb{P}^1(\mathbb{F}_p)$, that are interchanged by Frob_p , since otherwise $\overline{\rho}_p$ would be reducible.) But then, if l divides p+1, there are nontrivial deformations of $\overline{\rho}_p$ of the form $\rho_p = \operatorname{Ind}_{\mathbb{Q}_p}^{\mathbb{Q}_p} (\psi \mu)$ from G_p to $\operatorname{GL}_2(O')$, with O' a finite extension of O, $\psi \colon G_p \to O'^*$ the Teichmüller lift of $\psi \colon G_p \to k^*$, and with $\mu \colon G_{\mathbb{Q}_p^2} \to O'^*$ of order l. One checks that $\det(\rho)|_{I_p}$ is the Teichmüller lift of $\det(\overline{\rho})|_{I_p}$ (use that Frob_p acts on the tame inertia group $I_p^{\operatorname{tame}} = \hat{\mathbb{Z}}^{(p)}(1)$ by multiplication by p). The whole problem arises from the fact that, on the one hand, $\overline{\rho}_p$ and ρ_p have the same Artin conductor, namely, the square of the conductor of ψ , but that, on the other hand, $\overline{\rho}_p$ admits different lifts with this conductor. This means that if we consider lifts of $\overline{\rho}$ to be minimally ramified at p if they have Artin conductor $\operatorname{cond}(\psi)^2$ at p, then we get an $L_p \subset \operatorname{H}^1(G_p, \operatorname{ad}^0(\overline{\rho}))$, in the notation of Section 3, that is nonzero, making already Wiles' method at the minimal level impossible.

The conclusion is that, on the automorphic side, levels of newforms are not fine enough invariants to work with; one should impose finer conditions on the restrictions to the G_p (at the Galois side), and corresponding conditions on the irreducible admissible representations on the other side. Wiles already notes this in the second remark following Conjecture 2.16 in [64].

The finer conditions that will be imposed are in terms of what are called "types" and "extended types" in the articles that we discuss here. An extended type, at a prime p (p = l is allowed) is simply an isomorphism class of two-dimensional representations over $\overline{\mathbb{Q}}_l$ of the Weil-Deligne group W'_p of \mathbb{Q}_p (see Appendix A), and then types are isomorphism

classes of restrictions to I_p of extended types. The local Langlands correspondence makes extended types correspond to isomorphism classes of infinite dimensional irreducible admissible representations of $\mathrm{GL}_2(\mathbb{Q}_p)$, over $\overline{\mathbb{Q}}_l$. We will not discuss the proof of Theorem 5.1 here, as it is repeated in [14] and [4], with some changes, however. Diamond used, in order to simplify the representation theory at the automorphic side, the Jacquet-Langlands correspondence to work with a quaternion algebra over \mathbb{Q}_p instead of the matrix algebra. In the two subsequent articles, one works directly with modular curves. We recommend [23] for an overview of [24], that does not become too technical. But, as the reader can already guess, the rest of this section will get more technical, especially on the automorphic side, because of these finer restrictions.

With Theorem 5.1 above, it is not hard to prove that all elliptic curves over \mathbb{Q} that are semistable at 3 and 5 are modular; we refer to [24, §5] for details. So the only remaining problem to get modularity for all elliptic curves over \mathbb{Q} is to get rid of the semistability conditions at 3 and 5. Since modularity is invariant under twisting, Diamond's result actually implies that the only elliptic curves E that remain to be dealt with are those that have potentially good reduction at 3 and 5, but that do not have a twist with good reduction at 3 and 5. Since one knows that the two j-invariants that correspond to E with more than two automorphisms are modular, the only twists to consider are quadratic twists.

The first step in the direction of relaxing the conditions at 3 and 5 was made in [14], where it is proved that E is modular if it acquires good reduction over a tame extension of \mathbb{Q}_3 . The main new ingredient of this paper, compared to [24], is a new type of deformation problem, for a mod l representation of G_l . Roughly speaking, one considers deformations ρ over R of $\overline{\rho}$ over k such that the restriction of ρ to G_F , or a twist of it by a fixed quadratic character, is Barsotti-Tate, for F a fixed finite extension of \mathbb{Q}_l with ramification index $e \leq l-1$. We have seen, in Section 3, that it is crucial that the tangent space over k of the universal deformation ring of the type of deformations of $\overline{\rho}|_{G_l}$ that one considers be of dimension at most one. This crucial result for [14] was proved by Conrad in [13], using his description of finite free group schemes over the rings of integers of such F obtained in [12], generalizing earlier work of Fontaine in [30]. Using Conrad's result, it was then proved ([14, Theorem 7.1.2]) that any elliptic curve E over \mathbb{Q} that acquires good reduction over a tame extension of \mathbb{Q}_3 is modular.

The final step in relaxing the conditions at 3 is done in [4]. It is the work of Breuil [3], summarized in [2], that gives a workable enough description of certain finite free group schemes over rings of integers of arbitrary finite extensions F of \mathbb{Q}_l , that makes this possible. With this tool available, it is then proved that the remaining E, i.e., those that acquire good reduction only after a wild extension of \mathbb{Q}_3 , are modular. The article [4] consists for about 70% (of 77 pages) of the proof that, in the cases that are needed, the local deformation space at l has dimension at most one.

In order to keep this section of reasonable length, we postpone the discussion of these results at l to the next one, and in this one we focus more on the global aspects of the proof, and on Conjecture 1.3.1 of [4], which says for which $\rho: G_{\mathbb{Q}} \to \mathrm{GL}_2(O)$ one hopes to be able to prove modularity.

We will now follow [4, §1], in order to introduce the necessary terminology, and to state the conjecture just mentioned. We suppose l > 2. An extended l-type is defined as an isomorphism class of two-dimensional representations of W'_l over $\overline{\mathbb{Q}}_l$ (with open kernel), and l-types are isomorphism classes of restrictions to I_l of extended l-types.

Suppose that τ' is an extended l-type, with restriction τ to I_l .

DEFINITION 5.2. — A continuous representation $\rho: G_l \to \operatorname{GL}_2(O)$ (with O the ring of integers in a finite extension K of \mathbb{Q}_l) is said to be of extended type τ' (resp. of type τ) if:

- (1) ρ is potentially Barsotti-Tate;
- (2) $WD(\rho)$ (as in Appendix B) is in τ' (resp. $WD(\rho)|_{I_l}$ is in τ);
- (3) the character $\varepsilon^{-1} \det(\rho)$ has finite order prime to 1.

Now fix a finite extension K of \mathbb{Q}_l , let O be its ring of integers and k be its residue field. Let $\overline{\rho}$ be a two-dimensional continuous representation of G_l over k, say on a vector space V, such that $\operatorname{End}_{G_l}(V) = k$ (i.e., either $\overline{\rho}$ is absolutely irreducible, or it is a non-split extension of a character by another character). Under this last hypothesis, we have a universal deformation ring R_O^l representing the functor of deformations of $\overline{\rho}$ to complete noetherian local O-algebras with residue field k. (The superscript l is there to indicate that we are considering representations of G_l .) We remark that extended types will only be considered if their restriction to I_l is irreducible. Now consider $\operatorname{Spec}(R_O^l)$. In it, we have a minimal closed subset that contains all deformations of $\overline{\rho}$ to finite extensions O' of O that are of type τ , and similarly for extended type τ' . These minimal closed subsets correspond to (reduced) quotients $R_{O,\tau}^l$ and $R_{O,\tau'}^l$. A deformation ρ over R of $\overline{\rho}$ is said to be weakly of type τ (resp. weakly of extended type τ'), if the corresponding morphism $R_O^l \to R$ factors through $R_{O,\tau}^l$ (resp. through $R_{O,\tau'}^l$).

Definition 5.3. — A type τ (resp. an extended type τ') is weakly acceptable for $\overline{\rho}$ if there exists a surjection of O-algebras $O[[X]] \to R^l_{O,\tau}$ (resp. $O[[X]] \to R^l_{O,\tau'}$). A type τ (resp. an extended type τ') is acceptable for $\overline{\rho}$ if moreover $R^l_{O,\tau} \neq 0$ (resp. $R^l_{O,\tau'} \neq 0$), i.e., if there exists at least one l-adic deformation of type τ (resp. of extended type τ'). We will also speak of $\overline{\rho}$ accepting τ (resp. τ').

Of course, with these definitions, it is very hard to check whether a given $\overline{\rho}$ accepts a given τ or τ' . It is precisely this kind of verifications that occupy the most of [4], and it is there that crucial use is made of Conrad's and Breuil's results on finite group schemes. We note that [4] conjectures that an l-adic lift of $\overline{\rho}$ is of type τ (resp. extended

type τ') if and only if it is weakly of that kind (Conjecture 1.1.1 of [4]), but this has no importance for what follows. What is much more important, is that [4, Conjecture 1.3.1] tries to predict acceptability purely in computable, representation theoretic terms. In order to state this conjecture, [4] needs about 4 pages of preparation, consisting mostly of definitions. Instead of trying to state all these definitions, we will try to see where they come from.

The question one should ask oneself is: if f is a newform over $\overline{\mathbb{Q}}_l$, then what can one say about $\overline{\rho}_{f,l} \colon G_l \to \mathrm{GL}_2(\overline{\mathbb{F}}_l)$ in terms of $\pi_{f,l}$, assuming $\overline{\rho}_f$ irreducible? In particular, for a given $\overline{\rho}$, what are the irreducible admissible representations that occur as $\pi_{f,l}$, for f with $\overline{\rho}_{f,l} \cong \overline{\rho}$?

An answer to this question will then say for which τ and τ' there does exist an l-adic lift of $\overline{\rho}$ of that type. Moreover, from the mechanism that is used to find this, one may guess under what conditions one expects $R_{O,\tau}^l$ or $R_{O,\tau'}^l$ to be topologically generated by one element.

To find the answer to the question (and for other reasons as well), [4] constructs certain l-adic sheaves on certain modular curves, that pick out a non-zero part of exactly those π_f such that $\pi_{f,l}$ has a prescribed type (or extended type). For each τ and for each τ' , with $\tau'|_{I_l}$ irreducible, one defines open subgroups U_{τ} of $\mathrm{GL}_2(\mathbb{Z}_l)$ and $U_{\tau'}$ of $\mathrm{GL}_2(\mathbb{Q}_l)$, and irreducible representations σ_{τ} and $\sigma_{\tau'}$ on finite dimensional $\overline{\mathbb{Q}}_l$ -vector spaces, with open kernel. The choice of these subgroups and representations is justified by [4, Lemma 1.2.1]: for every irreducible admissible representation π of $\mathrm{GL}_2(\mathbb{Q}_l)$ over $\overline{\mathbb{Q}}_l$ one has

- $\operatorname{Hom}_{U_{\tau}}(\sigma_{\tau}, \pi) = \overline{\mathbb{Q}}_{l} \text{ if } \operatorname{WD}(\pi)|_{I_{l}} \cong \tau, \text{ and } \operatorname{Hom}_{U_{\tau}}(\sigma_{\tau}, \pi) = 0 \text{ if } \operatorname{WD}(\pi)|_{I_{l}} \ncong \tau;$
- $\operatorname{Hom}_{U_{\tau'}}(\sigma_{\tau'}, \pi) = \overline{\mathbb{Q}}_l$ if $\operatorname{WD}(\pi) \cong \tau'$, and $\operatorname{Hom}_{U_{\tau'}}(\sigma_{\tau'}, \pi) = 0$ if $\operatorname{WD}(\pi) \not\cong \tau'$.

The fact that such subgroups and representations exist is not particular to our situation. There is a general theory, called (no surprise) type theory, whose goal it is to describe smooth representations of p-adic groups in terms of their restrictions to compact open subgroups; see [6]. Before we go on, let us mention that Khare has also asked and answered the question above, at least in the case of types, in [34].

With these $(U_{\tau}, \sigma_{\tau})$ and $(U_{\tau'}, \sigma_{\tau'})$ one constructs sheaves on modular curves as follows. One defines U_l to be U_{τ} if one considers a type, and $U_{\tau'} \cap \operatorname{GL}_2(\mathbb{Z}_l)$ if one considers an extended type. In each case, one has a representation σ_l of U_l , namely, σ_{τ} and the restriction of $\sigma_{\tau'}$. Let $U^{(l)}$ be a sufficiently small open subgroup of $\operatorname{GL}_2(\hat{\mathbb{Z}}^{(l)})$, and let σ be the representation of $U:=U_lU^{(l)}$ given by σ_l . Then the modular curve and the sheaf are:

$$Y_U := \operatorname{GL}_2(\mathbb{Q}) \backslash \operatorname{GL}_2(\mathbb{A}) / U\mathbb{C}^*, \quad \mathcal{F}_{\sigma} := \operatorname{GL}_2(\mathbb{Q}) \backslash \left(\operatorname{GL}_2(\mathbb{A}) \times M_{\sigma}^{\vee} \right) / U\mathbb{C}^*,$$

with M_{σ} the $\overline{\mathbb{Q}}_{l}[U]$ -module given by σ . In the case where one considers an extended type, one also gets an automorphism w_{l} of the pair $(Y_{U}, \mathcal{F}_{\sigma})$. By construction, the duals of the two-dimensional Galois representations ρ that occur in $H_{!}^{1}(Y_{U}, \mathcal{F}_{\sigma})$ if one considers a type (resp. in $H_{!}^{1}(Y_{U}, \mathcal{F}_{\sigma})^{w_{l}=1}$ if one considers an extended type) for some $U^{(l)}$ are precisely

those that correspond to newforms f such that $\pi_{f,l}$ are of the prescribed kind that is described by τ or τ' . Let $U' := \mathrm{GL}_2(\mathbb{Z}_l)U^{(l)}$, and consider the morphism $\pi \colon Y_U \to Y_{U'}$. Then one has $H_!^1(Y_U, \mathcal{F}_{\sigma}) = H_!^1(Y_{U'}, \pi_* \mathcal{F}_{\sigma})$. In order to get information on the corresponding $\overline{\rho}_l$'s, one now reduces the sheaf modulo l, i.e, one chooses a \mathbb{Z}_l -lattice for σ , and one reduces modulo the maximal ideal. By construction, the Jordan-Hölder constituents of this reduction are of the form $\mathcal{F}_{n,m} := \operatorname{Sym}^n(\overline{\mathcal{F}}) \otimes \det(\overline{\mathcal{F}})^{\otimes m}$, with $0 \leq n < l$ and m in $\mathbb{Z}/(l-1)\mathbb{Z}$, and with $\overline{\mathcal{F}}$ the locally constant sheaf of two-dimensional $\overline{\mathbb{F}}_l$ -vector spaces given by the standard representation of $GL_2(\mathbb{F}_l)$, or, if one wishes, by the dual of the l-torsion of the universal elliptic curve. Moreover, the (n, m) that occur, and their multiplicities, can be computed by representation theory (Brauer characters). But now one can use the results of Deligne and Fontaine stated in [29, Theorems 2.5 and 2.6] that describe the $\overline{\rho}_{f,l}$ for newforms of prime to l level, and weight between 2 and l+1, in order to see which $\overline{\rho}$ occur in the $H^1_!(Y_{U'}, \mathcal{F}_{n,0})$. Since $\det(\overline{\mathcal{F}})$ is simply $\overline{\mathbb{F}}_l(\varepsilon)$, $H^1_!(Y_{U'}, \mathcal{F}_{n,m})$ is just $H^1(Y_{U'}, \mathcal{F}_{n,0})(\varepsilon^{-m})$. Let us note that in the case of an extended type, one also has to deal with the automorphism w_l ; this is done in [4, §1.4]. The role played by the $\mathcal{F}_{n,m}$ explains that the dependence upon $\overline{\rho}$ of " τ (resp. τ') admits $\overline{\rho}$ " is via its properties that determine the weight that Serre has attached to $\overline{\rho}|_{I_l}$ (see [29, Sections 2-4]).

Having seen this, we can understand what goes on behind the definition of the notion of " τ (resp. τ') admits $\overline{\rho}$ " in [4, §1.3]: this means that there exist newforms f of that type such that ρ_f is a lift of $\overline{\rho}$ of the required type. What is harder to understand, is what is behind the corresponding two notions "simply admits", because [4] defines this by simply listing all elements of this relation. A reasonable guess seems that this condition means that $\overline{\pi}_* \mathcal{F}_{\sigma}$ has exactly one Jordan-Hölder constituent that can give rise to $\overline{\rho}^{\vee}$. (In fact, the freeness results in Theorems 4.1 and 4.2 and the definition of the Hecke modules that are used imply that in the situations where these theorems can be applied, $\overline{\pi}_* \mathcal{F}_{\sigma}$ has exactly one Jordan-Hölder constituent that can give rise to $\overline{\rho}^{\vee}$.) It would be interesting to know how much of the relation between $\pi_{f,l}$ and $\overline{\rho}_{f,l}$ can be computed using Fontaine's functors. Let us now state this Conjecture 1.3.1, and the two main theorems (1.4.1 and 1.4.2) of [4].

Conjecture 5.4. — Let k be a finite subfield of $\overline{\mathbb{F}}_l$, $\overline{\rho} \colon G_l \to \operatorname{GL}_2(k)$ a continuous representation, τ an l-type and τ' an extended l-type with irreducible restriction to I_l . Suppose that the centraliser of the image of $\overline{\rho}$ is k and that the image of τ is not contained in the center of $\operatorname{GL}_2(\overline{\mathbb{Q}}_l)$.

- (1) τ (resp. τ') admits $\overline{\rho}$ if and only if $R_{O,\tau}^l \neq \{0\}$ (resp. $R_{O,\tau'}^l \neq \{0\}$), i.e., if and only if there is a finite extension K' of \mathbb{Q}_l in $\overline{\mathbb{Q}}_l$ and a continuous representation $\rho \colon G_l \to \mathrm{GL}_2(O_{K'})$ which reduces to $\overline{\rho}$ and has type τ (resp. has extended type τ').
- (2) τ (resp. τ') simply admits $\overline{\rho}$ if and only if τ (resp. τ') is acceptable for $\overline{\rho}$.

THEOREM 5.5. — Let l > 2 be prime, K a finite extension of \mathbb{Q}_l in $\overline{\mathbb{Q}}_l$ and k its residue field. Let $\rho \colon G_{\mathbb{Q}} \to \operatorname{GL}_2(K)$ be an odd continuous representation, ramified at only finitely many primes. Assume that its reduction $\overline{\rho} \colon G_{\mathbb{Q}} \to \operatorname{GL}_2(k)$ is absolutely irreducible after restriction to the quadratic subfield of $\mathbb{Q}(\mu_l)$, and is modular. Further, suppose that:

- $\overline{\rho}|_{G_l}$ has centraliser k;
- $\rho|_{G_l}$ is potentially Barsotti-Tate with l-type τ (resp. with extended l-type τ');
- τ (resp. τ') admits $\overline{\rho}$;
- τ (resp. τ') is weakly acceptable for $\overline{\rho}$.

Then ρ is modular.

The proof of this theorem is very parallel to the proof of [14, Theorem 7.1.1], and is just written in terms of the changes to make. The strategy is of course the same as Wiles', especially in the way that we have described it, but now one imposes, at all primes where this is required (i.e., l and the so-called vexing primes of [23]), these finer restrictions to define the right notion of minimally ramified deformations. The commutative algebra that is used consists of the results that we have described in Section 4. The required Hecke modules are constructed as cohomology groups of sheaves on modular curves just as the \mathcal{F}_{σ} above, with the difference that one will also have types τ_p at some primes $p \neq l$.

Of course, in order to apply this theorem, one has to prove that the last condition holds, i.e., that there exists a surjection $O[[X]] \to R^l_{O,\overline{\rho}}$. This condition has indeed been proved in sufficiently many cases in order to prove Theorem 1.4, by Conrad in [13] for tamely ramified types with small image, and in [4] for some more types, using Breuil's work. We will come back to this question of proving weak acceptability in the next section.

Let us now see what is still required in order to prove the theorem that all elliptic curves E over \mathbb{Q} are modular. The proof of this is divided into three cases:

- (1) $\overline{\rho}_{E,5}|_{\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\sqrt{5}))}$ is irreducible;
- (2) $\overline{\rho}_{E,5}|_{\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\sqrt{5}))}$ is reducible, but $\overline{\rho}_{E,3}|_{\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\sqrt{-3}))}$ is absolutely irreducible;
- (3) the remaining cases.

First of all, the last case corresponds to rational points on a few modular curves of small level. It is proved in [14, Lemma 7.2.3], with help of Elkies, that the set of all such elliptic curves has, up to isogeny and twist, just three elements, which are known to be modular by calculations. Let us consider the second case. Then E acquires semistable reduction over a tame extension of \mathbb{Q}_3 because $\overline{\rho}_{E,5}(I_3)$ has order dividing $(5-1)^25$. If a quadratic twist E' of E is semistable at 3, one switches to E', and one is in the situation considered in [24]. If not, then any ramified quadratic twist of E_K with K a ramified quadratic extension of \mathbb{Q}_3 has good reduction, so that one can use [13, Theorem 4.2.1]. Let us finally consider the first case. In this case, Theorem 1.4 says that $\overline{\rho}_{E,5}$ is modular. Moreover, since 5 > 3, E acquires semi-stable reduction over a tame extension of \mathbb{Q}_5 of degree dividing 4 or dividing 6; in the first case, where E is potentially ordinary at 5,

Theorem 5.3 of [24] applies, in the second case, there is a ramified extension K of \mathbb{Q}_5 of degree 3, such that a ramified quadratic twist of E_K has good, supersingular reduction, and [13, Theorem 4.2.1] applies.

So it remains to explain how Theorem 1.4 is proved. Let $\overline{\rho}$ be as in that theorem. One first twists $\overline{\rho}$ by a suitable quadratic character, such that $\overline{\rho}|_{G_3}$ falls into one of the 6 cases of [4], page 3, whose Artin conductors at 3 are 3^i , $0 \le i \le 5$. Then one considers elliptic curves E over \mathbb{Q} such that $\overline{\rho}_{E,5}$ is isomorphic to $\overline{\rho}$. The moduli space of these is the union of two non-empty open subschemes of $\mathbb{P}^1_{\mathbb{Q}}$, hence there are plenty of such E. Using Hilbert irreducibility, and some computations by Manoharmayum [37], one can show that there exists such an E such that $\overline{\rho}_{E,3}$ is surjective on $\mathrm{GL}_2(\mathbb{F}_3)$, and such that, in the cases of conductor 3^i with $i \ge 3$, $\rho_{E,3}$ is such that Theorem 5.5 above can be applied to it, i.e., such that the type, or extended type, can be proved to be weakly acceptable for $\overline{\rho}_{E,3}$. These results are proved in [4, §2.1]. The use of an extended type is required only in the case of conductor 3^5 .

6. DEFORMATION PROBLEMS AT l

Let us now discuss the results concerning weak acceptability, obtained in [13], [14], and in Sections 4-9 of [4]. We recall what this means. Let l > 2 be prime, K a finite extension of \mathbb{Q}_l , with ring of integers O and residue field k. Let $\overline{\rho}$ be a two-dimensional representation of G_l over k, with centraliser k. Let τ be an l-type, and τ' an extended l-type with irreducible restriction to I_l . Then the quotients $R_{O,\tau}^l$ and $R_{O,\tau'}^l$ of the universal deformation ring R_O^l of $\overline{\rho}$ have been defined in the previous section, by taking the Zariski closures in $\operatorname{Spec}(R_O^l)$ of the sets of l-adic lifts of type τ (resp. extended type τ'). Weak admissibility of τ (resp. τ') just means that there exists a surjection $O[[X]] \to R^l_{O,\tau}$ (resp. $O[[X]] \to R^l_{O,\tau'}$). So what one wants to compute is the dimension over k of the space of deformations over $k[t]/(t^2)$ of $\overline{\rho}$ that are weakly of type τ (or extended type τ'). But the way that these kind of deformations have been defined makes this impossible. So, in order to deal with this problem, one defines other deformation problems, whose universal deformation rings surject to the ones above, and for which one then proves that they admit a surjection from O[X]. The aim of this section is just to describe these new deformation problems, and to sketch the tools that are used in their study. Let us mention that there seems to be some hope that one can deal directly with the rings $R_{O,\tau}^l$ and $R_{O,\tau'}^l$ (see the beginning of [4, §4]). In what follows, we drop the superscript l from the notation, as we are only considering representations of G_l .

We follow [4, §4]. We first discuss some generalities, and then what happpens in the worst case, i.e., the conductor 3^5 case. Let F be a finite Galois extension of \mathbb{Q}_l , with group Γ . We let R be the ring of integers in F, and k its residue field (if we need to refer to the residue field of O, we will call it k_O). For \mathcal{G} a finite free group scheme over R,

of l-power order, we let $\mathbb{D}(\mathcal{G})$ denote the contravariant Dieudonné module of \mathcal{G}_k ; it is a W(k)-module, equipped with operators \mathbf{F} and \mathbf{V} , such that $\mathbf{F}\mathbf{V} = \mathbf{V}\mathbf{F} = l$ and, for all x in W(k): $\mathbf{F}x = \operatorname{Frob}_l(x)\mathbf{F}$. A descent datum for a finite free group scheme \mathcal{G} over R is a right action of Γ on \mathcal{G} , compatible with its action on $\operatorname{Spec}(R)$, i.e., for each γ in Γ , one has a commutative diagram:

$$\begin{array}{ccc}
\mathcal{G} & \xrightarrow{[\gamma]} & \mathcal{G} \\
\downarrow & & \downarrow \\
\operatorname{Spec}(R) & \xrightarrow{\sim} & \operatorname{Spec}(R)
\end{array}$$

such that $[\gamma_1 \gamma_2] = [\gamma_2][\gamma_1]$ for all γ_1 and γ_2 in Γ . Note that, since $\mathbb{Z}_l \to R$ may be ramified, this is not what one should call a descent datum; however, it is a descent datum after restriction to F. In particular, we can associate in this way, to a pair $(\mathcal{G}, [\cdot])$, a group scheme over \mathbb{Q}_l . A descent datum as above gives an action of Γ on $\mathbb{D}(\mathcal{G})$, compatible with the action of Γ on W(k), and commuting with \mathbf{F} and \mathbf{V} , i.e., it becomes a module over the ring $W(k)[\mathbf{F}, \mathbf{V}][\Gamma]$ (with suitable commutation relations). The idea is now that to τ and τ' , one can associate ideals I and I' of this ring, that will impose the right conditions on l-adic lifts of $\overline{\rho}$ to be of type τ or of extended type τ' .

Let us now describe the kind of deformation problems that are considered. The extension F of \mathbb{Q}_l should be taken such that the type τ (or the extended type τ') becomes unramified over it. Then one fixes a model over R for $\overline{\rho}$, i.e., a pair $(\mathcal{G}_0, [\cdot])$ as above, giving $\overline{\rho}$ as the module of its $\overline{\mathbb{Q}}_l$ -points, and such that I (or I') annihilates $\mathbb{D}(\mathcal{G}_0)$. Once a model is chosen, one can consider all deformations ρ of $\overline{\rho}$, say to artinian rings, that admit a model $(\mathcal{G}, [\cdot])$ with $\mathbb{D}(\mathcal{G})$ killed by I (or by I'), with a filtration in which each successive quotient is isomorphic to $(\mathcal{G}_0, [\cdot])$. A nice condition to impose is then that such models are unique, and indeed, this can be realized in the situations that are needed (this is what $[4, \S 4.2]$ is about). Let us denote by $R_{O,I}$ and $R_{O,I'}$ the universal deformation rings thus obtained.

This is where the generalities end, and where one considers each of the 3 cases (conductor 3^3 , 3^4 and 3^5) separately. Let us just describe what happens in the worst case: conductor 3^5 . In that case, Conjecture 1.3.1 of [4] suggests 3 extended types τ_i' , $i \in \mathbb{Z}/3\mathbb{Z}$. One can take $O = \mathbb{Z}_3$. The restrictions of the τ_i' to $W_{\mathbb{Q}_3(\sqrt{-3})}$ are given by a morphism

$$\mathbb{Q}_{3}(\sqrt{-3})^{*} \longrightarrow \mathbb{Q}_{3}(\zeta)^{*}, \qquad \begin{cases} \sqrt{-3} & \mapsto & \zeta - \zeta^{-1} \\ -1 & \mapsto & -1 \\ 4 & \mapsto & 1 \\ 1 + 3\sqrt{-3} & \mapsto & \zeta \\ 1 + \sqrt{-3} & \mapsto & \zeta^{i} \end{cases}$$

with ζ of order 3. This defines abelian extensions F_i of $\mathbb{Q}_3(\sqrt{-3})$ of degree 12, that are Galois over \mathbb{Q}_3 . The ideals I_i that one uses are all three generated by: $\mathbf{F} + \mathbf{V}$, $[\gamma_4^2] + 1$, and $([\gamma_3] - [\gamma_3^{-1}])[\gamma_2] - \mathbf{F}$, with γ_2 , γ_3 and γ_4 certain elements of $\operatorname{Gal}(F_i/\mathbb{Q}_3)$. In particular, γ_2 is not in the inertia subgroup of $\operatorname{Gal}(F_i/\mathbb{Q}_3)$, hence the last generator of the I_i reflects that one works with an extended type. Without this condition, or even with a similar condition coming from other possible extended types, one wouldn't expect the tangent space of the deformation problem to be of dimension one.

Theorem 4.6.1 of [4] says that, for each i, there exist four models of $\overline{\rho}$ with the property that models for deformations as above are unique, and, moreover, that each deformation of $\overline{\rho}$ as above has a filtration with all successive quotients isomorphic to one of these four. Theorem 4.6.2 of [4] then says that all four have the property that the universal deformation ring is topologically generated by one element, but three of them have universal deformation ring $\mathbb{F}_3[[X]]$ ([4, Theorem 4.6.3]). Using this, it is finally proved that every 3-adic lift of $\overline{\rho}$ that is of extended type τ'_i , say over a finite extension O of \mathbb{Z}_3 , comes from a morphism $R_O \to R_{O,\tau'_i}$ that factors through the universal deformation ring R_{O,I'_i} associated to this last model. Hence, finally, this proves that each of the τ'_i weakly accepts $\overline{\rho}$.

To finish this section, let us point out that proving the theorems in the preceding paragraph takes about 45 pages in [4], with about 30 of them filled with computations with Breuil's ϕ_1 -modules. It is my hope that these notes will encourage readers to take a look at those pages, and understand what is going on. In order to say at least something about these modules, let us give the definition of the l-torsion ϕ_1 -modules over R, the category of which is anti-equivalent to that of finite free group schemes over R that are killed by l.

Let R etc. be as in the second paragraph of this section. Let π be a uniformizer of R, and $E_{\pi}(u) = u^e - lG_{\pi}(u)$ its minimal polynomial over W(k). Let ϕ denote the l-th power map on $k[u]/(u^{el})$. An l-torsion ϕ_1 -module is then a triple (M, M_1, ϕ_1) with M a finite free $k[u]/(u^{el})$ -module, with M_1 a $k[u]/(u^{el})$ -submodule containing u^eM , and with $\phi_1 \colon M_1 \to M$ ϕ -semilinear, such that $\phi_1(M_1)$ generates M as $k[u]/(u^{el})$ -module.

We note that the category just described does not depend on the choice of the uniformizer π , but that the functors giving the anti-equivalence mentioned above do depend on that choice. This fact causes a lot of trouble in [4], as one has to study the action of $\operatorname{Gal}(F/\mathbb{Q})$ on models \mathcal{G} over $\operatorname{Spec}(R)$, and, of course, $\operatorname{Gal}(F/\mathbb{Q})$ will not fix the choice of π .

7. TWO RELATED RESULTS

The aim of this section is to briefly state two modularity results on two-dimensional Galois representations that are obtained by others than those mentioned in the title. They can be found in [51] and [7].

THEOREM 7.1 (Skinner, Wiles). — Let l > 2 be prime, and let $\rho: G_{\mathbb{Q}} \to \operatorname{GL}_2(\overline{\mathbb{Q}}_l)$ be an odd continuous representation, ramified at only finitely many primes, with $\det(\rho) = \psi \varepsilon^{k-1}$ with ψ of finite order, $\varepsilon: G_{\mathbb{Q}} \to \mathbb{Z}_l^*$ the cyclotomic character and $k \geq 2$ an integer. Suppose that $\rho|_{G_l}$ is the extension of a character ψ_2 by a character ψ_1 with $\psi_2|_{I_l}$ of finite order and such that $\overline{\psi_1}|_{G_l} \neq \overline{\psi_2}|_{G_l}$. If $\overline{\rho}: G_{\mathbb{Q}} \to \operatorname{GL}_2(\overline{\mathbb{F}}_l)$ is irreducible, then suppose that $\overline{\rho}$ is modular. Then ρ is modular.

In fact, they even prove a stronger result, with \mathbb{Q} replaced by an arbitrary totally real field. This generality is actually even necessary for the theorem above, since the proof involves changing the field \mathbb{Q} . Let us note that the representations ρ considered by Skinner and Wiles are very different from those considered by Breuil, Conrad, Diamond, and Taylor, which mostly have irreducible restrictions to any open subgroup of G_l .

THEOREM 7.2 (Buzzard, Dickinson, Shepherd-Barron, Taylor). — Let ρ be a continuous, irreducible, odd representation from $G_{\mathbb{Q}}$ to $GL_2(\mathbb{C})$ with unsolvable image. Suppose that ρ is unramified at 2 and at 5, and that the image of $\rho(\operatorname{Frob}_2)$ in $\operatorname{PGL}_2(\mathbb{C})$ has order 3. Then ρ is modular.

Let us note that all continuous representations $\rho: G_F \to \operatorname{GL}_2(\mathbb{C})$ with solvable image and with F a number field are known to be associated to automorphic representations, by Hecke (in the dihedral case) and Langlands and Tunnell ([36], [61]). The strategy of the proof is explained in [57], and carried out in [8], [54], and [28]. The paper [7] mainly pulls everything together, and provides slight technical but needed improvements of previously obtained results.

In a nutshell, the strategy consists in realizing a suitable twist of ρ over a number field in \mathbb{C} , such that it has a reduction $\overline{\rho}$ mod 2 with values in $\operatorname{GL}_2(\mathbb{F}_4)$. Then one shows that $\overline{\rho}$ is modular, that ρ arises from an overconvergent 2-adic modular form, and finally that ρ arises from a weight one form.

8. LATEST NEWS

This section has been added at the time the final version of this text was written (June 2000). Its aim is just to direct the reader to some developments that took place after the lecture (March).

Taylor has released two preprints [58] and [59]. In the first one, he proves, using work of Skinner and Wiles ([51], [52], [53]), and of many other people, the following result.

THEOREM 8.1 (Taylor). — Let l be an odd prime, and $\rho: G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_l)$ a continuous irreducible representation such that:

- (1) ρ is unramified at all but finitely many primes;
- (2) ρ is odd (i.e., $\det \rho(c) = -1$);
- (3) $\rho|_{G_l}$ is an extension of χ_2 by $\varepsilon^n \chi_1$, with $\chi_1|_{I_l}$ and $\chi_2|_{I_l}$ of finite order and n a non-zero positive integer, such that $\varepsilon^n \chi_1 \chi_2^{-1}(I_l)$ is not pro-l.

Then there is a totally real number field E, a regular algebraic cuspidal automorphic representation π of $GL_2(\mathbb{A}_E)$ and a place λ of the field of coefficients of π above l such that $\rho_{\pi,\lambda}$ (the λ -adic representation associated to π) is equivalent to $\rho|_{G_E}$.

As a consequence, such a ρ has an L-function, and this L-function is meromorphic and satisfies the expected functional equation. Also, under some mild hypothesis, it follows that ρ occurs in some $H^i(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_l(r))$, as in Conjecture 2.1. The idea that allows one to use the results of Skinner and Wiles, and others, concerning ρ such that $\overline{\rho}$ has a soluble image, is to use abelian varieties with real multiplications such that the $\overline{\rho}$ of the Theorem above is related to the l-torsion, and such that the p-torsion (for some other prime p) gives a suitable soluble image. The existence of such abelian varieties is proved by applying Skolem type results to Hilbert-Blumenthal modular varieties (see for example [38]).

Ramakrishna proved in [43] that, under mild hypotheses, a continuous mod l representation $\overline{\rho} \colon G_{\mathbb{Q}} \to \operatorname{GL}_2(k)$ (not supposed to be modular) can be lifted to a representation $\rho \colon G_{\mathbb{Q}} \to \operatorname{GL}_2(W(k))$ over the Witt vectors of k, with ρ unramified at almost all primes. More recently, in [44], he has proved that one can even obtain that ρ is semi-stable or crystalline at l. His main innovation is to consider deformation problems that lead to a universal deformation ring W(k). He requires $\rho|_{G_p}$ to be of the form $\begin{pmatrix} \varepsilon & * \\ 0 & 1 \end{pmatrix}$ for suitable p. Note that this is stronger than a condition on $\rho|_{I_p}$ (until now, only conditions on $\rho|_{I_p}$ were imposed), which makes it reasonable that the deformation ring will be small. One finds a slight generalization of Ramakrishna's results in [59], where they are used to prove some more cases of the Artin conjecture. Combined with Theorem 8.1, one obtains that $\overline{\rho}$ as above becomes modular after restriction to G_E with E a suitable totally real extension of \mathbb{Q} (see [58]). This can be seen as a "potential" version of Serre's conjecture.

Khare has used Ramakrishna's work ([35]) to give another proof of certain "R=T" theorems. Assuming $\overline{\rho}$ to be modular, one gets an "R=T" theorem for Ramakrishna's deformation problem (the main result of [27] implies that Ramakrishna's lift ρ is modular). Starting with this result, Khare proves that for large enough Σ , $R_{O,\Sigma} \to \mathbb{T}_{O,\Sigma}$ is an isomorphism; his proof avoids the special arguments of Wiles and Taylor-Wiles in the minimal case. Of course, this last result suffices for proving modularity results. (It seems that from this one also easily obtains the result for all Σ .)

Breuil and Mézard have released a preprint ([5]) in which they give a conjectural description of the Samuel multiplicity of local deformation rings $R_{O,\tau}^l$ in automorphic terms. They also prove this conjecture in many cases.

Appendix A. GALOIS REPRESENTATIONS ASSOCIATED TO MODULAR FORMS

The aim of this section is to recall what we need about the Galois representations associated to modular forms. For simplicity, we only discuss the case of forms of weight two, so that we only need to deal with the cohomology of the constant sheaf on modular curves. We use the now standard point of view that was initiated by Deligne in [18]. As a general reference for this section, we recommend [26].

The object from which everything originates here is the Shimura datum (GL_2, \mathbb{H}^{\pm}) , with $GL_2(\mathbb{R})$ acting on $\mathbb{H}^{\pm} = \mathbb{P}^1(\mathbb{C}) - \mathbb{P}^1(\mathbb{R})$ in the usual way. Let \mathbb{A} denote the ring of adèles of \mathbb{Q} , and \mathbb{A}_f its subring of the finite adèles. For every compact open subgroup U of $GL_2(\mathbb{A}_f)$, let $X_U^0(\mathbb{C})$ denote the complex analytic variety $GL_2(\mathbb{Q}) \setminus (\mathbb{H}^{\pm} \times GL_2(\mathbb{A}_f)/U)$; it can be compactified, by adding a finite number of points, to a smooth compact Riemann surface (usually not connected) $X_U(\mathbb{C})$. We denote the associated complex algebraic curve by $X_{U,\mathbb{C}}$. The interpretation of these curves as moduli spaces of elliptic curves with level structures give models $X_{U,\mathbb{Q}}$ over \mathbb{Q} . The inverse limit $X_{\mathbb{Q}}$ of the $X_{U,\mathbb{Q}}$ has an action, from the right, by $GL_2(\mathbb{A}_f)$. For l prime, the \mathbb{Q}_l -vector space:

$$H_l:=\lim_U \mathrm{H}^1(X_{U,\overline{\mathbb{Q}},\mathrm{et}},\overline{\mathbb{Q}}_l)$$

has an action by $G_{\mathbb{Q}} \times GL_2(\mathbb{A}_f)$. In order to understand the decomposition of H_l as a representation of $GL_2(\mathbb{A}_f)$, one uses the Hodge decomposition:

$$\mathrm{H}^1(X_U(\mathbb{C}),\mathbb{C}) = \Omega^1(X_U(\mathbb{C})) \oplus \overline{\Omega^1(X_U(\mathbb{C}))}.$$

On $\Omega^1(X_U(\mathbb{C}))$ one no longer has an action of $GL_2(\mathbb{A}_f)$, but is still is a module over the Hecke algebra associated to U: the convolution algebra of compactly supported bi-U-invariant functions on $GL_2(\mathbb{A}_f)$ (say that $GL_2(\hat{\mathbb{Z}})$ has measure one). The q-expansion principle and some theory of smooth irreducible representations of the $GL_2(\mathbb{Q}_p)$ show that $\Omega^1(X(\mathbb{C}))$ decomposes into a direct sum of irreducible ones, each one occurring only once:

$$\Omega^1(X(\mathbb{C})) = \lim_U \Omega^1(X_U(\mathbb{C})) = \bigoplus_f \pi_f,$$

where f runs through the set of newforms of weight two with coefficients in \mathbb{C} . We recall that for all p we have a chosen embedding $\overline{\mathbb{Q}} \to \overline{\mathbb{Q}}_p$. It follows that H_l decomposes as a direct sum:

$$H_l \cong \bigoplus_f \rho_f^{\vee} \otimes \pi_f,$$

with f running through the set of weight two newforms with coefficients in $\overline{\mathbb{Q}}_l$, and with $\rho_f: G_{\mathbb{Q}} \to \operatorname{GL}_2(\overline{\mathbb{Q}}_l)$ a continuous representation. We note that ρ_f is realized over any finite extension over which π_f is defined. The representation π_f is a restricted tensor product $\otimes_p' \pi_{f,p}$ over all primes, with each π_p an infinite dimensional irreducible admissible representation of $\operatorname{GL}_2(\mathbb{Q}_p)$. On the Galois side, we define, for each prime p, $\rho_{f,p} := \rho_f|_{G_p}$.

With these definitions, one knows that, for $p \neq l$, $\pi_{f,p}$ and $\rho_{f,p}$ determine each other via a suitably normalized local Langlands correspondence. (This was first proved at the unramified places by Eichler and Shimura, then for $\pi_{f,p}$ principal series or special by Langlands, then for $p \neq 2$ by Deligne, and finally for all p by Carayol, and simplified by Nyssen.) The representation $\rho_{f,l}$ is usually not determined by $\pi_{f,l}$ (just think of the case where f corresponds to an elliptic curve with split multiplicative reduction at l, where $\rho_{f,l}$ almost determines the elliptic curve), but Saito has shown in [47] that the (ϕ, N, G_l) -module obtained by forgetting the filtration of the filtered (ϕ, N, G_l) -module corresponding to $\rho_{f,l}$ via Fontaine's functor ([32, §10]) corresponds to $\pi_{f,l}$. (Actually, in the weight two case that we consider this is in fact easily deduced from the results for $p \neq l$; see [14, Appendix B].)

In order to fix notation, let us give a precise description of this local correspondence, so that there is no ambiguity about the normalization. To do this, we first recall that the best way to formulate the local Langlands correspondence is in terms of the Weil-Deligne group (see [56, §4] and [19]). For p prime, the Weil group W_p of \mathbb{Q}_p is the subgroup of G_p consisting of elements whose image in $G_{\mathbb{F}_p}$ is in $\operatorname{Frob}_p^{\mathbb{Z}}$ ($\operatorname{Frob}_p \colon x \mapsto x^p$ is the arithmetic Frobenius). The Weil-Deligne group is an object W'_p that is defined so that a representation of W_p' on a finite dimensional E-vector space $(E \supset \mathbb{Q})$ is a pair (V, N) with V a continuous representation of W_p (with the discrete topology on V), and a nilpotent endomorphism N of V such that $wNw^{-1}v = pNv$ for all v in V and w in W_p mapping to Frob_p. Such a pair is called F-semisimple if V is semisimple as a representation of W_p . With these definitions, there are canonical bijections between the set of (isomorphism classes of) infinite dimensional irreducible admissible representations of $GL_2(\mathbb{Q}_p)$ over \mathbb{Q} , and the set of 2-dimensional F-semisimple representations of W'_n over $\overline{\mathbb{Q}}$. These bijections are such, that local L and ε -factors on both sides (suitably normalized) correspond, and they are compatible with the action of $G_{\mathbb{O}}$ on both sides. For $p \neq 2$, this is easy, since one can easily write down the elements on both sides (on the Galois side, one uses that the wild inertia acts reducibly). For p=2, this is harder; the general case was worked out by Kutzko. If V is a finite dimensional K-vector space, with K a finite extension of \mathbb{Q}_l , and $p \neq l$, there is an equivalence between representations of W_p' on V as above, and continuous representations of W_n on V.

Following [14], we normalize the local Langlands correspondence WD in such a way that $\rho_{f,p}|_{W_p}$, viewed as a representation of W'_p over $\overline{\mathbb{Q}}_l$, is isomorphic to $\overline{\mathbb{Q}}_l \otimes_{\overline{\mathbb{Q}}} \mathrm{WD}(\pi_{f,p})$, for each newform f with coefficients in $\overline{\mathbb{Q}}$. If $\sigma(\pi_{f,p})$ is as in [9], then we have $\mathrm{WD}(\pi_{f,p}) = \sigma(\pi_{f,p}) \otimes \chi$, where χ is the unramified character that sends Frob_p to p. If $\pi_{f,p}$ is unramified, i.e., if p does not divide the level of f, then $\rho_{f,p}$ is unramified and $\rho_{f,p}(\mathrm{Frob}_p)$ is semisimple (remember that we are in weight two) and has characteristic polynomial $X^2 - t_p X + p s_p$, where t_p and s_p are the eigenvalues of f for the Hecke and diamond operators T_p and S_p that are defined by the double cosets $U(\frac{p}{0}, 0)U$ and $U(\frac{p}{0}, 0)U$,

with $U = \operatorname{GL}_2(\mathbb{Z}_p)$. The determinant of $\rho_{f,p}$ is $\varepsilon \chi_{\pi_{f,p}}$, where $\chi_{\pi_{f,p}}$ is the central character of $\pi_{f,p}$, viewed as a character of W_p^{ab} via the isomorphism of class field theory under which the image of p in $\mathbb{Q}_p^*/\mathbb{Z}_p^*$ corresponds to Frob_p . If χ is a continuous character $\mathbb{Q}_p^* \to \overline{\mathbb{Q}}^*$, then $\operatorname{WD}(\pi_{f,p} \otimes (\chi \operatorname{det})) \cong \operatorname{WD}(\pi_{f,p}) \otimes \chi$.

To finish this section, let us recall some facts about the classification of the two-dimensional semisimple representations of W_p , over \mathbb{C} , say. Let ρ be such a representation, on a \mathbb{C} -vector space V, say. If ρ is reducible, it is a sum of two characters. Suppose ρ irreducible. Since the wild inertia subgroup I_p^{wild} of I_p acts on V via a finite p-group, it acts via two characters (possibly equal), unless p=2 (recall that the dimension of an irreducible complex representation of a finite group divides the order of the group). Suppose that $\rho | I_p^{\text{wild}}$ splits as a sum of two distinct characters. Then W_p acts on the set of the two corresponding lines in V, and non-trivially because ρ is irreducible. It follows that ρ becomes reducible over a quadratic extension of \mathbb{Q}_p . Suppose now that I_p^{wild} acts via scalars on V. Then, considering the action of W_p/I_p (which is the semi-direct product of \mathbb{Z} by I_p^{tame}) on $\mathbb{P}(V)$, one sees that, again, ρ becomes reducible over a quadratic extension of \mathbb{Q}_p . So the conclusion is this: the two-dimensional complex semisimple representations of W_p are sums of two characters, or induced from a character of W_K with K quadratic over \mathbb{Q}_p , or such that p=2 and I_p^{wild} acts irreducibly (these latter ones were first classified by Weil [63], in his "exercices dyadiques"; clearly, the title of [4] is inspired by this).

Appendix B. TYPES ASSOCIATED TO *l*-ADIC REPRESENTATIONS AND TO ELLIPTIC CURVES

Let l be any prime. We recall that an extended l-type (over $\overline{\mathbb{Q}}_l$) is an isomorphism class of two-dimensional representations over $\overline{\mathbb{Q}}_l$ of the Weil-Deligne group W'_l of \mathbb{Q}_l (see Appendix A), and that types are isomorphism classes of restrictions to I_l of extended l-types. We want to describe how one attaches an extended type to a continuous representation $\rho \colon G_l \to \mathrm{GL}_2(O)$, with O the ring of integers of a finite extension K of \mathbb{Q}_l contained in $\overline{\mathbb{Q}}_l$, under the assumption that ρ is potentially Barsotti-Tate. So let ρ be such a representation, and let F be a finite extension of \mathbb{Q}_l over which ρ becomes Barsotti-Tate, i.e., such that $\rho|_{G_F}$ is isomorphic to $\mathcal{G}(\overline{\mathbb{Q}}_l)$ for some l-divisible group with O-action over the ring of integers O_F of F. Of course, one solution to this is simply to apply Fontaine's $D_{\mathrm{st},F}$ functor as in [32, §10(b)], but in this simple case of p-divisible group schemes one can be more explicit. Another reason for doing this more explicitly is that one wants to do computations in the case of elliptic curves. For more details we refer to [14, Appendix B].

The representation ρ we have corresponds to an l-divisible group $\mathcal{G}_{\mathbb{Q}_l}$ over \mathbb{Q}_l , with O-action. Let F be a finite Galois extension of \mathbb{Q}_l such that \mathcal{G}_F extends (uniquely, by [55, Theorem 4]) to an l-divisible group \mathcal{G}_{O_F} over O_F , with O-action. Let Γ denote $\operatorname{Gal}(F/\mathbb{Q}_l)$.

For every σ in Γ , we have commutative diagrams:

with k_F the residue field of O_F . The last diagram gives a right action of G_l on \mathcal{G}_{k_F}/k_F . Let $d\colon W_l\to\mathbb{Z}$ be the morphism such that σ in W_l induces the $d(\sigma)$ th power of the absolute Frobenius on $\overline{\mathbb{F}}_l$. Then we get a morphism from W_l to $(\mathbb{Q}\otimes\operatorname{End}_{k_F}(\mathcal{G}_{k_F}))^*$ by sending σ to $[\sigma]^{-1}\operatorname{Frob}_{\operatorname{abs}}^{d(\sigma)}$. Now let \mathbb{D} denote the contravariant Dieudonné module functor. Then $\mathbb{Q}\otimes\mathbb{D}(\mathcal{G}_{k_F})^\vee$, with $^\vee$ denoting $W(k_F)$ -dual, is a free $K\otimes_{\mathbb{Z}_l}W(k_F)$ -module of rank two, with a left action by W_l . The extended l-type $\operatorname{WD}(\rho)$ associated to ρ is then the two-dimensional $\overline{\mathbb{Q}}_l$ vector space obtained by base change via $K\otimes_{\mathbb{Z}_l}W(k_F)\to\overline{\mathbb{Q}}_l$ (note that both K and F are subfields of $\overline{\mathbb{Q}}_l$); the monodromy operator is defined to be zero, as we have good reduction.

We can repeat the construction above, with $\mathcal{G}_{\mathbb{Q}_l}$ replaced by an elliptic curve E over \mathbb{Q}_l , with good reduction E_{O_F} over O_F . Then one gets morphisms:

$$W_l \longrightarrow (\mathbb{Z}[1/l] \otimes \operatorname{End}_{k_F}(E_{k_F}))^* \longrightarrow (\overline{\mathbb{Q}} \otimes \operatorname{End}_{k_F}(E_{k_F}))^* \cong \operatorname{GL}_2(\overline{\mathbb{Q}}).$$

More generally, one can start with a newform f with coefficients in $\overline{\mathbb{Q}}$, and then one has:

$$WD(\rho_{f,l}) = \overline{\mathbb{Q}}_l \otimes_{\overline{\mathbb{Q}}} WD(\pi_{f,l}),$$

by the results of Appendix A.

In [62] one can find a complete description of all ρ_l that arise from elliptic curves over \mathbb{Q}_l , in terms of their associated filtered (ϕ, N, G_l) -modules.

Acknowledgements

I would like to thank Christophe Breuil, Fred Diamond, Reinie Erné, Eyal Goren, Chandrashekhar Khare, Rutger Noot and Richard Taylor for their comments on an earlier version of this text, and Michel Gros for the reference to [6]. Attending the conference "Modularity of Elliptic Curves and Beyond" at the MSRI in Berkeley, December 6–10, 1999, was very useful, as well as the series of talks given by Brian Conrad on this subject in Rome, July 1999.

REFERENCES

[1] P. BERTHELOT – Altérations de variétés algébriques. Séminaire Bourbaki, 1995–96, exposé 815, Astérisque **241**, S.M.F. (1997), 273–311.

- [2] C. BREUIL Schémas en groupes et modules filtrés. C. R. Acad. Sci. Paris 328 (1999), 93–97.
- [3] C. BREUIL Groupes p-divisibles, groupes finis et modules filtrés. Preprint, 1999. To appear in Annals of Mathematics.
- [4] C. BREUIL, B. CONRAD, F. DIAMOND, R. TAYLOR On the modularity of elliptic curves over Q: wild 3-adic exercises.

 Preprint, http://www.math.harvard.edu/
- [5] C. BREUIL, A. MEZARD Multiplicités modulaires et représentations de $\operatorname{GL}_2(\mathbb{Z}_p)$ et de $\operatorname{Gal}(\overline{\mathbb{Q}}_n/\mathbb{Q}_p)$ en l=p. Preprint, June 2000, Orsay.
- [6] C.J. BUSHNELL Smooth representations of p-adic groups: the role of compact open subgroups. Proceedings of the international congress of mathematicians, Vol. 1, 2 (Zürich, 1994), 770–779, Birkhaüser, Basel, 1995.
- [7] K. BUZZARD, M. DICKINSON, N. SHEPHERD-BARRON, R. TAYLOR On icosahedral Artin representations. Preprint, http://www.math.harvard.edu/
- [8] K. BUZZARD, R. TAYLOR Companion forms and weight one forms. Ann. of Math. (2) 149 (1999), no. 3, 905-91
- [9] H. CARAYOL Sur les représentations l-adiques associées aux formes modulaires de Hilbert. Ann. Sci. Éc. Norm. Sup. 19 (1986), 409–468.
- [10] L. CLOZEL Motifs et formes automorphes. Automorphic forms, Shimura varieties, and L-functions, Vol. I, edited by L. Clozel and J.S. Milne, Perspectives in Mathematics 10, Academic Press (1990), 77–159.
- [11] P. COLMEZ, J-M. FONTAINE Construction des représentations p-adiques semistables. Preprint Ecole Normale Supérieure, LMENS-99-18.
- [12] B. CONRAD Finite group schemes over bases with low ramification. Compositio Math. 119 (1999), 239–320.
- [13] B. CONRAD Ramified deformation problems. Duke Math. J. 97 (1999), 439–513.
- [14] B. CONRAD, F. DIAMOND, R. TAYLOR Modularity of certain potentially Barsotti-Tate Galois representations. J.A.M.S. 12 (1999), 521–567.
- [15] G. CORNELL, J. SILVERMAN, G. STEVENS, editors Modular Forms and Fermat's Last Theorem Springer-Verlag, 1997.
- [16] H. DARMON A proof of the full Shimura-Taniyama-Weil conjecture is announced. Notices of the A.M.S., December 1999.
- [17] H. DARMON, F. DIAMOND, R. TAYLOR Fermat's Last Theorem. Elliptic curves, modular forms and Fermat's last theorem (Hong Kong, 1993). Second edition. International Press, Cambridge, MA, 1997, 2–140.
- [18] P. DELIGNE Formes modulaires et représentations l-adiques. Séminaire Bourbaki, 1968–69, exposé 355, Lecture Notes in Mathematics 179, Springer (1969), 139–172.

- [19] P. DELIGNE Formes modulaires et représentations de GL(2). Antwerp II, Lecture Notes in Math. **349**, Springer (1973), 55–106.
- [20] B. DE SMIT, H. LENSTRA Explicit construction of universal deformation rings. In [15], pp. 313–326.
- [21] B. DE SMIT, K. RUBIN, R. SCHOOF Criteria for complete intersections. In [15], 343–355.
- [22] F. DIAMOND The refined conjecture of Serre. In Elliptic curves, modular forms and Fermat's last theorem (Hong Kong, 1993). International Press, 1995, 22–37.
- [23] F. DIAMOND An extension of Wiles' results. In [15], 475–498.
- [24] F. DIAMOND On deformation rings and Hecke rings. Ann. Math. **144** (1996), 137–166.
- [25] F. DIAMOND The Taylor-Wiles construction and multiplicity one. Invent. math. 128 (1997), 379–391.
- [26] F. DIAMOND, J. IM *Modular forms and modular curves*. In "Seminar on Fermat's last theorem", Canadian Mathematical Society Conference Proceedings 17, 1995 (V. Kumar Murty, editor).
- [27] F. DIAMOND, R. TAYLOR Lifting modular mod l representations. Duke Math. J. **74** (1994), no. 2, 253–269.
- [28] M. DICKINSON On the modularity of certain 2-adic Galois representations. Preprint.
- [29] S.J. EDIXHOVEN The weight in Serre's conjectures on modular forms. Invent. math. 109 (1992), 563–594.
- [30] J-M. FONTAINE Groupes p-divisibles sur les corps locaux. Astérique 47–48, S.M.F., 1977.
- [31] J-M. FONTAINE Représentations p-adiques semi-stables. In "Périodes p-adiques," Astérisque 223, S.M.F., 1994, p. 113–184.
- [32] J-M. FONTAINE, B. MAZUR Geometric Galois representations. In "Elliptic curves, modular forms and Fermat's Last Theorem" (Hong Kong, 1993), International Press, 1995, pp. 41–78.
- [33] K. FUJIWARA Deformation rings and Hecke rings in the totally real case. Preprint.
- [34] C. KHARE A local analysis of congruences in the (p, p)-case: Part II. To appear in Invent. math.
- [35] C. KHARE On isomorphisms between deformation and Hecke rings. Work in progress.
- [36] R.P. LANGLANDS Base change for GL(2). Ann. of Math. Studies **96**, Princeton University Press, Princeton, 1980.
- [37] J. MANOHARMAYUM Pairs of mod 3 and mod 5 representations arising from elliptic curves. Thesis, Cambridge University, 1998.

- [38] L. MORET-BAILLY Groupes de Picard et problèmes de Skolem II. Ann. Sci. ENS **22** (1989), 181–194.
- [39] R. NOOT Abelian varieties arising from Mumford's Shimura curves. Preprint, Rennes, 1997.
- [40] J. OESTERLÉ Nouvelles approches du "théorème" de Fermat. Séminaire Bourbaki, 1987–88, exposé 694, Astérisque **161–162**, S.M.F. (1988), 175–186.
- [41] J. OESTERLÉ Travaux de Wiles (et Taylor, . . .), partie II. Séminaire Bourbaki, 1994–95, exposé 804, Astérisque 237, S.M.F. (1996), 333–355.
- [42] R. RAMAKRISHNA On a variation of Mazur's deformation functor. Compositio Math. 87 (1993), 269–286.
- [43] R. RAMAKRISHNA Lifting Galois representations. Invent. math. 138 (1999), 537–562.
- [44] R. RAMAKRISHNA Deforming Galois representations and the conjectures of Serre and Fontaine-Mazur. Preprint.
- [45] K.A. RIBET, On modular representations of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms. Invent. math. **100** (1990), 431–476.
- [46] K. RUBIN Modularity of mod 5 representations. In [15], 463-474.
- [47] T. SAITO Modular forms and p-adic Hodge theory. Invent. math. 129 (1997), 607–620.
- [48] J-P. SERRE Représentations l-adiques. Kyoto Int. Symposium on Algebraic Number Theory, Japan Soc. for the Promotion of Science (1977), pp. 177–193, (Oeuvres, t. III, pp. 384–400).
- [49] J-P. SERRE Sur les représentations modulaires de degré 2 de $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$. Duke Math. J. **54** (1987), 179–230.
- [50] J-P. SERRE Travaux de Wiles (et Taylor, ...), partie I. Séminaire Bourbaki, 1994–95, exposé 803, Astérisque 237, S.M.F. (1996), 319–332.
- [51] C. SKINNER, A. WILES Residually reducible representations and modular forms. Preprint. To appear in Pub. Math. IHES.
- [52] C. SKINNER, A. WILES Base change and a problem of Serre. Preprint.
- [53] C. SKINNER, A. WILES Nearly ordinary deformations of irreducible residual representations. Preprint.
- [54] N. SHEPHERD-BARRON, R. TAYLOR Mod 2 and mod 5 icosahedral representations. J. Amer. Math. Soc. 10 (1997), 281–332.
- [55] J. TATE p-divisible groups. Proceedings of a conference on local fields, (Driebergen, 1966), Springer, 1967, 158–183.
- [56] J. TATE Number theoretic background. Proceedings of Symposia in Pure Mathematics **33** (1979), part 2, 3–26.
- [57] R. TAYLOR *Icosahedral Galois representations*. Pacific J. Math.: Special issue: Olga Taussky-Todd: in memoriam (1997), 337–347.

- [58] R. TAYLOR Remarks on a conjecture of Fontaine and Mazur. Preprint, http://www.math.harvard.edu/
- [59] R. TAYLOR On icosahedral Artin representations II.

 Preprint, http://www.math.harvard.edu/
- [60] R. TAYLOR, A. WILES Ring-theoretic properties of certain Hecke algebras. Ann. Math. **141** (1995), 553–572.
- [61] J. TUNNELL Artin's conjecture for representations of octahedral type. Bull. A.M.S. 5 (1981), 173–175.
- [62] M. VOLKOV Les représentations l-adiques associées aux courbes elliptiques sur \mathbb{Q}_p . Preprint, http://www.math.uiuc.edu/Algebraic-Number-Theory/
- [63] A. WEIL Exercices dyadiques. Invent. math. 27 (1974), 1–22.
- [64] A. WILES Modular elliptic curves and Fermat's Last Theorem. Ann. Math. 141 (1995), 443–551.

Bas EDIXHOVEN¹

I.R.M.A.R., U.M.R. 6625 du CNRS Université de Rennes I Campus de Beaulieu F-35042 RENNES Cedex

 $E ext{-}mail: ext{edix@maths.univ-rennes1.fr}$

¹partially supported by the Institut Universitaire de France, and by the European TMR Network Contract ERB FMRX 960006 "arithmetic algebraic geometry".