IP Current Developments in Mathematics, 2006

The Sato-Tate conjecture

L. Clozel

0. Introduction

In April 2006, Richard Taylor has completed the proof of the Sato-Tate conjecture for elliptic curves over totally real fields, under a mild assumption (see below). This is the completion of a project started several years ago by Taylor and Michael Harris, aiming at developing the higher-dimensional automorphic deformation theory of Galois representation, with its applications to cases of Langlands functoriality – here, "Sato-Tate". I will report specifically on 3 papers:

Clozel, Harris, Taylor, *Automorphy for some ℓ-adic lifts of automorphic* mod ℓ *representations* ([**CHT**]);

Harris, Shepherd-Barron, Taylor, *Ihara's lemma and potential automorphy* ([**HSBT**]);

Taylor, *Automorphy for some ℓ-adic lifts of automorphic* mod ℓ *representations* II ([**T**]).

Together the three papers give a proof of "Sato-Tate" (with the restriction alluded to). The proof is based on the continuous development, since '95, of Wiles' deformation method. After the initial impetus by Wiles and Taylor-Wiles, I see two strands of development, characterized by:

- The systematic study, and use, of forms on **higher** groups and their associated Galois representations (Harris, Taylor).

- Deep technical progress (in degree 2, but this adapts to the higher, dimensional situation) in the Taylor-Wiles method: here fundamental contributions are due to Taylor, Diamond, Fujiwara, Skinner-Wiles, Kisin.

The purpose of these notes is to explain the result, even for a reader who is not familiar with the modern theory of automorphic forms and

Wiles' deformation theory. I have aimed at displaying, as clearly as possible, the logical structure of the proof.

It is, in fact, not easy to envision the whole proof with some clarity: first, because it systematically involves forms on higher groups – here, unitary groups of arbitrary rank – which may not be, even now, familiar to all number theorists. Secondly, because the deformation theory – essentially used here to give an **approximate** proof [1] of a case of Langlands functoriality – has to be combined, in an intricate fashion, with two other ingredients. The first (see § 5) is a very powerful elaboration of the "change of primes" or "3-5" trick of Taylor and Wiles. (This has been developed by Taylor, in the case of GL(2), in two very important papers [**32, 33**]. They are also basic for the proof of Serre's conjecture [**17**] on which Ribet reported at the conference.) The second, which is closely connected, is the idea of proving **potential** modularity rather than modularity. This is already introduced in § 1: see the section on "Brauer-Taylor".

These notes, then, endeavour to explain as clearly as possible the fundamental ideas of the proof. I have tried to give statements which involve as little technique as possible; however, at least in the five first paragraphs, I have also attempted to give essentially correct statements of the main theorems. (A few terms, however, will not be defined, such as "crystalline representations" or "Steinberg condition".) Also, I have always stated the results with the minimal content required for the proof of "Sato-Tate", and even that **over** $\mathbb{Q}$. (The general results apply to totally real fields.)

In § 1, we give the statements of the Sato-Tate conjecture, and of the (slightly weaker) theorem now proven (Theorem A). I also explain three fondamental "reductions": one, due to Serre and Tate, to an analycity property of "high L-functions"; Langlands' expectation that these higher L-functions are "automorphic"; and the arguments, introduced by Taylor and relying on Brauer induction, reducing the automorphic problem to a potential one (Thm. B).

In § 2 I review "known" material on Galois representations associated to automorphic forms ($\equiv$ automorphic representations). The results are due to a number of people, culminating with the work of Harris and Taylor.

§ 3 states one of the main results of [**CHT**] on "conditional modularity", the analogue of Wiles' theorem in higher degree. (Some indication of the proof will wait until § 6.)

§ 4-5 are devoted to certain fibered Calabi-Yau varieties, whose usefulness in Arithmetic was revealed in [**HSBT**]. They play the role that modular curves (Wiles) or Hilbert-Blumenthal varieties (Taylor) played earlier as a medium for the "3-5" trick. Their basic properties are

[1] The correct term is **potential** proof.

described in § 4, and in § 5, using them, we complete the proof of Theorem A (Sato-Tate), following Harris, Shepherd-Barron and Taylor. At this point, however, the proof is **conditional**. (This was the situation when [**HSBT**] was – electronically – published, i.e., September 2005.) It depends on the generalization (to forms on higher unitary groups) of a classical lemma of Ihara.

The object of § 6 is to give some indications on the arithmetic theory of automorphic forms on "compact" unitary groups, used for the proof of Theorem C in § 3. Here we have not been able to give precise statements, but the reader familiar with the Wiles/Taylor-Wiles proof will recognize the usual objects, as well as the difficult problem named "non-minimal level". One way to obviate this would be to prove Conjecture I (end of § 6), the generalized Ihara lemma. A correct statement is given. This conjecture seems very plausible (the natural analog in characteristic 0 is true) but difficult. Even with Taylor's new proof it remains a very attractive problem.

Finally, § 7 is devoted to a mere outline of Taylor's new method. This relies (1) on proving "potential modularity" rather than "modularity" results, and (2) on a new method of Kisin.

In conclusion, the reader can expect here, of course, only an outline of the 225 pp. of difficult mathematics contained in the 3 papers. He should also consult Harris's excellent exposition [**11**]. I would like to thank Professor Yau, and the Harvard Mathematics Department, for inviting me to give these lectures. I thank Peter Sarnak for forcing me to read [**HSBT**] in one night. I thank Michael Harris and Richard Taylor for their patience in explaining their work (and the common work with Shepherd-Barron); Michel Raynaud for reading with me the commutative algebra in Taylor's paper; Jean-Pierre Serre and John Tate for historical comments.

1. The conjecture. Reductions

Assume E is an elliptic curve over $\mathbb{Q}$. There is then an integer $N \geqslant 1$ such that the equation of E (as a smooth cubic in $\mathbb{P}^2$) has coefficients in $\mathbb{Z}[\frac{1}{N}]$. We may further (multiplicatively) increase N so, if p is a prime and $p \nmid N$, the reduced curve E over $\mathbb{F}_p$ is still smooth, i.e., still an elliptic curve. (There is, of course, an optimal choice of N, the conductor, but this will not concern us. See [**28**].)

Denote by E_p the reduced curve, over $\mathbb{F}_p$. It has been known since Hasse (about 1930) that the number of points of $E_p(\mathbb{F}_p)$ is

$$N_p = p + 1 - (\beta_p + \bar{\beta}_p) \ , \ |\beta_p| = \sqrt{p}.$$

This uniquely determines β_p, up to complex conjugation. We **normalize** (irrationally) by setting

$$\alpha_p = \beta_p / \sqrt{p}.$$

Thus $|\alpha_p| = 1$; only the couple $\{\alpha_p, \alpha_p^{-1}\}$ is, of course, determined.

Now let $E_{\mathbb{C}}$ denote the complex points of E, and let $M = \mathrm{End}(E_{\mathbb{C}})$ be the ring of complex multiplications of E (over $\mathbb{C}$), i.e., its endomorphisms as a complex Lie group. If $M \neq \mathbb{Z}$ one says that E has complex multiplication, and the behaviour of the $\alpha's$ (or $\beta's$) is severely constrained and (essentially) understood. See [**10**]. Therefore, assume now:

• E **has no complex multiplication** (over $\mathbb{C}$).

• **Question**: What is the repartition of the $\alpha_p's$ ($p \nmid N$) on the unit circle?

We may think of the diagonal matrix $\begin{pmatrix} \alpha_p & \\ & \alpha_p^{-1} \end{pmatrix}$ as a conjugacy class in SU(2) - for which we have taken a diagonal representative. Consider the Haar measure on SU(2). By Weyl's formula for invariant integration, it determines a mesure μ on the set of conjugacy classes - our set of $\alpha's$, modulo the identification ($\alpha \mapsto \alpha^{-1}$). If

$$\alpha = e^{i\theta} \qquad (\theta \in [0, \pi])$$

then $d\mu = \frac{2}{\pi} \sin^2 \theta \; d\theta$.

As usual in Number theory, we analyze the distribution of the $\alpha_p's$ as a **density**. For $X > 0$, write

$$\mu_X = \frac{\sum\limits_{p \leqslant X} \delta(\alpha_p)}{\sum\limits_{p \leqslant X} 1},$$

a probability measure on the set of conjugacy classes, where $\delta(\alpha_p)$ is the Dirac measure of the point $(\alpha_p, \alpha_p^{-1})$.

Conjecture (Sato, Tate 1963). *For $X \to +\infty$, $\mu_X \to \mu$.*

(The convergence is the weak convergence of probability measures.)

The present theorem is a little weaker. The considerations at the beginning of this §, about smoothness, can be made more precise (cf. [**28**]). Given $E/\mathbb{Q}$, one can define intrinsically the primes "of good reduction" where E has a smooth projective model over $\mathbb{Z}_p$. This is actually defined by E seen as a curve over $\mathbb{Q}_p$. One says that E has potential good reduction at p if $E \times_{\mathbb{Q}} L$ has good reduction for a finite extension L of $\mathbb{Q}_p$. Here and further, for X an algebraic variety over a field k, $X \times_k K$ is the variety over K (an extension of k) obtained by extension of scalars.

The following theorem is one of the consequences of the 3 papers:

Theorem A. *Assume*

(i) *E has no complex multiplication* (*over* $\mathbb{C}$).

(ii) *At some prime p_0, E does* **not** *have potential good reduction.*

Then the Sato-Tate conjecture is true for E.

From now on, we will abreviate the "Sato-Tate Conjecture" to "ST". Note that (i) and (ii) are (negative) **potential** conditions, i.e., can be formulated in terms of the curves $E \times_{\mathbb{Q}} F$ for $F/\mathbb{Q}$ a number field. (Complex multiplication over $\mathbb{C}$ is "acquired" over a finite extension F; if E does not satisfy (ii), $E \times_{\mathbb{Q}} F$ will have good reduction for a suitable number field.)

1st reduction: Serre

Fix N such that E has good reduction for $p \nmid N$. We will simply write $L(s,E)$ for the Hasse-Weil zêta function of E with the bad primes removed:

$$L(s,E) = \prod_{p \nmid N} \frac{1}{(1-\alpha_p p^{-s})(1-\alpha_p^{-1} p^{-s})}.$$

By the work of Wiles **et al** [**37, 35, 3**], one knows that E is associated to a modular form of might 2 on $\Gamma_0(N)$. Thus there exists a modular form

$$f(z) = \sum_1^\infty a_n\, q^n \qquad (Im\ z > 0\ ,\ q = e^{2i\pi z}),$$

eigenform of the Hecke operators, such that

$$L(s,E) = L^N(s,f)$$

where the right-hand side is the L-function of f (an Euler product) with the "bad" primes removed. In particular $L(s,E)$, a priori convergent (with our normalization) for Re $s > 1$, is

$$\begin{cases} \text{holomorphic in the } s\text{-plane} \\ \neq 0 \text{ on the line Re } s = 1. \end{cases}$$

(We will return to modular forms later in this §.) We will say that a function $L(s)$ is **good** if it is holomorphic and $\neq 0$ in a neigbourhood of the line (Re $s = 1$). We note once and for all that the complete L-functions of E (with suitable factors at the bad primes) can be rigourously defined, but that a finite number of factors in the Euler products do not change the **analytic** arguments which follow. (In particular, they have no effect on being "good".)

We now form, for each $n \geqslant 1$, the **n-th symmetric power** L-function:

$$L^{(n)}(s,E) = \prod_{p \nmid N} \frac{1}{(1-\alpha_p^n p^{-s})(1-\alpha_p^{n-2} p^{-s}) \cdots (1-\alpha_p^{-n} p^{-s})}.$$

Note that, if there was a matrix $t_p \in \mathrm{SU}(2)$ naturally associated to $(\alpha_p, \alpha_p^{-1})$, the p-th factor of $L(s, E)$ would be

$$L_p(s, E) = \det(1 - t_p p^{-s})^{-1}$$

and the p-th factor of $L_p^{(n)}$ would be

$$L_p^{(n)}(s, E) = \det(1 - S^n(t_p)p^{-s})^{-1}$$

where S^n is the representation of SU(2) on the homogeneous polynomials of degree n in (X, Y).

Theorem (Serre, Tate). *Assume $L^{(n)}(s, E)$ is good for all $n > 0$. Then* (ST) *is true for E.*

The proof essentially reduces to the Hadamard-de la Vallée Poussin argument.

The measures μ_X are invariant; to compute their limit it suffices, according to Weyl, to test them against (finite linear combinations of) characters of SU(2), which approximate continuous functions. Assume χ_n $(n = 0, 1, \ldots)$ is the character of the irreducible representation of degree $n + 1$. In order to prove (ST) we must show that

$$(\mu_X, \chi_n) \xrightarrow[(X \to +\infty)]{} 0\ ,\ n = 1, 2, \ldots; \tag{1.1}$$

of course, (μ_X, χ_0) is identically 1. However, writing $L(s) = L^{(n)}(s)$, $\chi = \chi_n$; the expression of $L^{(n)}$ yields:

$$-\frac{L'(s)}{L(s)} = \sum_{p,\alpha} \frac{\chi(t_p^\alpha) \log\ p}{p^{\alpha s}} \quad (\mathrm{Re}\ s > 1),$$

where p runs over primes $(\nmid N)$, $\alpha \geqslant 1$.

Under our assumptions, the left-hand side extends holomorphically to $(\mathrm{Re}\ s > 1)$. The Wiener-Ikehara theorem then yields:

$$\sum_{p \leqslant X} \chi(t_p) = o(X/\log X) \quad (X \to +\infty)$$

Since

$$\sum_{p \leqslant X} 1 = X/\log X + o(X/\log X)$$

this implies (1.1). For more details see Serre [**25**, Ch. I]. (This argument was already given, albeit heuristically, by Tate [**31**, § 4]. However Tate, motivated by his conjectures on cycles, considered the tensor products, rather than the symmetric products, of the 2-dimensional representation.)

2nd reduction: Langlands

Our problem is now to show that, for all n, $L^{(n)}$ is good. The only known way to prove "good" properties of L-functions is to associate automorphic forms to them. For instance, in the case of our elliptic curve over $\mathbb{Q}$, Wiles and his successors have proved the following. Let

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) : c \equiv 0 \ \ [N] \right\}.$$

A cusp form of weight 2 for $\Gamma_0(N)$ is a holomorphic function f on the upper half plane $\{z \in \mathbb{C} : \mathrm{Im}\ z > 0\}$ satisfying the functional equation

$$f\Big(\frac{az+b}{cz+d}\Big) = (cz+d)^2 f(z)$$

for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$, and vanishing at infinity:

$$f(z) = \sum_1^\infty a_n q^n \ , \ q = e^{2i\pi z}.$$

Given E, there exists a cusp form f of weight 2 on $\Gamma_0(N)$ - which is, morever, an eigenform of the Hecke operators - such that

$$L(s + \frac{1}{2}, f) = L^{\mathrm{full}}(s, E)$$

where $L(s,f) = \sum_1^\infty a_n n^{-s}$, and $L^{\mathrm{full}}(s,E)$ is the full L-function (completed by the correct factors for $p|N$). (The shift of $\frac{1}{2}$ cannot be avoided, because of our "bad" normalization for $L(s,E)$, necessary for a simple description of the higher L-functions.) Then Hecke's theory implies the holomorphy of $L(s,E)$ - in the whole s-plane.

In higher degrees, the classical definition of automorphic forms is not adequate, so we use the adelic theory. Let

$$\mathbb{A} = \mathbb{R} \times \prod_p{}' \ \mathbb{Q}_p$$

(restricted product) be the full ring of adèles. If G is a reductive group over $\mathbb{Q}$, $G(\mathbb{Q})$ embeds into $G(\mathbb{A})$ as a discrete subgroup. In particular, take $G = \mathrm{GL}(2)$; $\mathbb{R}^\times$ embeds centrally into $G(\mathbb{R})$. We consider the space

$$\mathcal{A}_G = L^2(G(\mathbb{Q})\mathbb{R}_+^\times \backslash G(\mathbb{A}))$$

where $G(\mathbb{A})$ acts by right translations. We can define an automorphic representation of $G(\mathbb{A})$ as an irreducible submodule π of $\mathcal{A}_G$ [2]. Simple

[2] This notion is slightly incorrect, i.e., not sufficiently general. Here it is enough for our purposes.

facts about the topology of $G(\mathbb{A})$ then imply that π decomposes as a tensor product:

$$\pi = \bigotimes_v \pi_v \tag{1.2}$$

over all primes v of $\mathbb{Q}$ (including ∞), π_v being an irreducible representation of $G(\mathbb{Q}_v)$ ($= \mathrm{GL}(2, \mathbb{R})$ or $\mathrm{GL}(2, \mathbb{Q}_p)$).

A standard construction associates to f an automorphic repersentation π, which is then uniquely associated to E. In fact, π is a **cuspidal** representation - the notion of cuspidal representations is a translation of the notion of cusp forms in the classical language.

Replacing GL(2) by GL(n), then, we have a notion of automorphic or cuspidal representation of $\mathrm{GL}(n, \mathbb{A})$, again with a decomposition (1.2).

There is a notion of ramification for these objects: given π, the factor π_p will be unramified for almost all p (= all p but a finite number), and the adélic theory of automorphic forms associates to π_p a **Hecke matrix** $t(\pi_p)$: this is a diagonal matrix of degree n, modulo conjugation (i.e., the action of $\mathfrak{S}_n$ on the entries). Then one can form the (partial) L-function:

$$L(s, \pi) = \prod_p \det(1 - t(\pi_p)p^{-s})^{-1} \tag{1.3}$$

(p unramified). The work of Godement-Jacquet show that $L(s, \pi)$ (and also the "completed" L-function, with suitable factors at the bad primes) extends to the complex s-plane, with the usual properties familiar from Hecke. (In particular $L(s, \pi)$ will be holomorphic in the whole plane; a functional equation relates $L(s, \pi)$ and $L(1 - s, \tilde{\pi})$ where $\tilde{\pi}$ is the dual representation.)

Conjecture (Langlands, 1970). *For any n, there exists a cuspidal representation Π of $GL(n + 1, \mathbb{A})$ such that*

$$L(s, \Pi) = L^{(n)}(s, E) := L^{(n)}(s, \pi)$$

where π is the representation of $GL(2, \mathbb{A})$ obtained from E.

We note that Langlands's conjecture (even in this context) is more general and complicated: in particular, given a cuspidal π, Π will be "automorphic" - in a more general sense, see the previous footnote - but not cuspidal in general. The fact that Π should be **cuspidal** follows from our assumption on E (no potential good reduction).

Now the **non-vanishing** of $L(s, \Pi)$ on (Re $s = 1$), for cuspidal Π, has been proved by Jacquet-Shalika [**16**]. By the Theorem of Serre and Tate, then, we see that

Langlands's Conjecture imples Thm. A.

3rd reduction: Brauer-Taylor

Note that if we could prove Langlands' conjecture we would show that the higher L-functions are better than "good" in our narrow sense, since we would know their behaviour in the whole s-plane. In the case of $L(s,E)$, let us show how we could obtain the "good" behaviour by proving only a result much weaker than the existence of f. This idea has been introduced by Richard Taylor, who uses it very efficiently in [**32, 33**]. See also, his lecture in Beijing [**34**].

We must recall that $L(s,E)$ can be defined by a representation of $G_{\mathbb{Q}} = \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$. We fix ℓ, and consider

$$H^1 = H^1_{\text{ét}}(E \times_{\mathbb{Q}} \bar{\mathbb{Q}}, \mathbb{Q}_\ell).$$

This is, of course, the "dual Tate module" of the curve E; however we will have to consider similar representations in degrees higher than 1, so it is better to think already in terms of étale cohomology. The Galois group $G_{\mathbb{Q}}$ acts on this space, and we can define, for $p \nmid N$, the action of the geometric Frobenius Frob_p; we then form

$$L(s, r^1) = \prod_{p \nmid N} \det(1 - \mathrm{Frob}_p p^{-s} | H^1);$$

with our analytic normalization we then have

$$L(s, r^1) = L(s - \frac{1}{2}, E).$$

Now for $F/\mathbb{Q}$ finite, we have natural **operations** on Galois representations. We identify $G_F = \mathrm{Gal}(\bar{\mathbb{Q}}/F)$ with a subgroup of $G_{\mathbb{Q}}$. Then:

(i) Given a representation $r_{\mathbb{Q}}$ of $G_{\mathbb{Q}}$, we can define $r_F = \mathrm{res}^F_{\mathbb{Q}}\, r_{\mathbb{Q}}$, its restriction to G_F.

(ii) Given r_F, we can form the induced representation

$$r_{\mathbb{Q}} = \mathrm{ind}^{\mathbb{Q}}_F r_F = \mathrm{ind}^{G_{\mathbb{Q}}}_{G_F} r_F.$$

Our representations are finite-dimensional and continuous, on ℓ-adic vector spaces. Assume moreover that we can associate L-functions to them, satisfying for a finite extension $F/\mathbb{Q}$:

(iii) $L_F(s, r_F) = L_{\mathbb{Q}}(s, \mathrm{ind}^{\mathbb{Q}}_F r_F)$.

(So these L-functions are "inductive".) This is true for the L-functions we consider, either on removing a sufficiently large number of bad primes, or after defining the good Euler factors at all primes, which is possible for the automorphic L-functions, and for those associated to Galois representations in the cohomology of algebraic varieties.

Now assume $F/\mathbb{Q}$ is a Galois extension, so we have an exact sequence

$$1 \to G_F \to G_{\mathbb{Q}} \to G_{F/\mathbb{Q}} \to 1$$

with $G_{F/\mathbb{Q}} = \mathrm{Gal}(F/\mathbb{Q})$. By Brauer's theorem, we have in the Grothendieck group of finite-dimensional representations of $G_{F/\mathbb{Q}}$:

$$1_{G_{F/\mathbb{Q}}} = \sum_i \; n_i \, \mathrm{ind}_{G_F/F_i}^{G_{F/\mathbb{Q}}} \; \chi_i$$

where 1 denotes the trivial representation, $F_i \subset F$ is a subfield such that $\mathrm{Gal}(F/F_i)$ is nilpotent, χ_i is an Abelian character of this group, and the n_i are relative integers. Now the maps $G_\mathbb{Q} \to G_{F/\mathbb{Q}}$, $G_{F_i} \to G_{F/F_i}$ allow us to view characters of the finite quotients as characters of the full Galois groups. Using obvious equalities, and (i-iii) we find, for our 2-dimensional representation $r = r^1$ of $G_\mathbb{Q}$ – or any representation having the properties recalled after (iii):

$$\begin{aligned} L_\mathbb{Q}(s,r) &= L_\mathbb{Q}(s, r \otimes 1_{G_{F/\mathbb{Q}}}) \\ &= \prod_i \; L_\mathbb{Q}(s, r \otimes \mathrm{ind}_{G_{F_i}}^{G_\mathbb{Q}} \; \chi_i)^{n_i}. \end{aligned}$$

Now

$$r \otimes \mathrm{ind}_{G_{F_i}}^{G_\mathbb{Q}} \; \chi_i = \mathrm{ind}_{G_{F_i}}^{G_\mathbb{Q}} \, (r|G_{F_i} \otimes \chi_i)$$

just as in the case of representations of finite groups, so:

$$L_\mathbb{Q}(s,r) = \prod_i L_{F_i}(r|_{G_{F_i}} \otimes \chi_i)^{n_i}. \tag{1.4}$$

Now assume we know that $r|_{G_F}$ is not only good but **automorphic** - i.e., associated (with F now playing the rôle of $\mathbb{Q}$) to an automorphic representation π_F of $\mathrm{GL}(2, \mathbb{A}_F)$.

We have the following diagram of extensions:

$$\begin{tikzcd} & F \arrow[dl, dash] \arrow[dd, dash] \\ F_i & \\ & \mathbb{Q} \end{tikzcd}$$

with F/F_i nilpotent ($F_i/\mathbb{Q}$ is not necessarily Galois). The Galois group $G_i = \mathrm{Gal}(F/F_i)$ acts on F_i; since $r_{F_i} = r|_{G_{F_i}}$ comes, by restriction, from G_F, it is isomorphic to all its conjugates by G_i. By well-known facts ("strong multiplicity one"), the same is true of π_F:

$$\pi_F \cong \pi_F \circ \sigma \quad (\sigma \in \mathrm{Gal}(F/F_i)). \tag{1.5}$$

We are now in a situation where we can use Langlands' theory of **automorphic** base change. This implies that - assuming (1.5) - there exists an automorphic representation π_{F_i} of $\mathrm{GL}(2, \mathbb{A}_{F_i})$ such that r_{F_i} is associated to π_{F_i}.

This means that

$$L(s, \pi_{F_i}) = L_{F_i}(s - \frac{1}{2}, r_{F_i}) \tag{1.6}$$

where the left-hand term is defined by (1.3) - for $n = 2$, but with a natural analogue when $\mathbb{Q}$ is replaced by F_i; and the right-hand term is defined naturally by the (restricted) Galois representation. It suffices for us to have such an identity at the good primes.

Now Galois representations, as well as automorphic representations, can be twisted by characters of finite order of the Galois group ($\equiv$ of the idèle class group by class field theory) and (1.6) **implies**:

$$L(s, \pi_{F_i} \otimes \chi_i) = L_{F_i}(s - \frac{1}{2}, r_{F_i} \otimes \chi_i).$$

Under our assumptions on E and r, π_{F_i} is actually a cuspidal representation of $\mathrm{GL}(2, \mathbb{A}_{F_i})$. Thus also $\pi_{F_i} \otimes \chi_i$. The L-function on the left, therefore, is "good", in particular holomorphic and non-vanishing on (Re $s = 1$). Since this property is conserved by products and **quotients**, (1.4) shows the same for $L_{\mathbb{Q}}(s, r)$.

Of course, giving this argument for the standard L-function $L(s, E)$ of E is absurd: while we know that this L-function "is automorphic" we do not know it in general when $\mathbb{Q}$ is replaced by F and E by $E \times_{\mathbb{Q}} F$ (recall that non-soluble base change is not available!). However we see that we could prove $L(s, E)$ **good** by showing only that $E \times_{\mathbb{Q}} F$ is associated to a cuspidal representation **for a suitable Galois extension** F. Following Taylor we say that E (or r) is **potentially automorphic** if this is the case.

Before we proceed we insert a remark concerning the argument of descent yielding π_{F_i} from π_F (verifying (1.5)). The group $\mathrm{Gal}(F/F_i)$ is nilpotent (in fact, "p-elementary", cf Serre [**24**]). If it is cyclic of prime order the descent argument is just Langlands'. Otherwise we take a tower of extensions, cyclic of prime order, between F_i and F and descend from step to step. Assume for instance there are two steps:

$$\begin{tikzcd} F \arrow[d, dash] \\ F' \arrow[d, dash] \\ F_i \end{tikzcd}$$

Using (1.5) we descend π_F to $\pi_{F'}$. If $\mathrm{Gal}(F'/F_i) =< \tau >$ (with $\tau^\ell = 1$ for some prime ℓ) then it follows a priori from the first descent that $(\pi_{F'})^\tau$ is π_F **up to an Abelian twist**. To proceed we need $\pi_{F'} \cong \pi_F^\tau$ exactly, which need not be the case [3]. It is a crucial remark of Harris that in our situation we can choose $\pi_{F'}$ associated to $r|_{G_{F'}}$ and therefore isomorphic to its τ-conjugate.

Since base change (Abelian or solvable – taking the previous remark into account) is available for $\mathrm{GL}(n)$ by the work of Arthur and this

[3] This is a notorious mistake in [**1**, § 3.6].

author [**1**], the foregoing arguments extend to representations of higher degree.

There is another trick in the same vein. Translating Langlands's conjecture in obvious terminology, we want to prove that "$S^n r$ is automorphic". Write R_{n+1} for $S^n r$ (we index by the degree). By the Clebsch-Gordan formula,

$$R_n \otimes R_2 = R_{n+1} \oplus R_{n-1}. \tag{1.7}$$

(To be correct, this is true for the "normalized" representations occurring in Langlands's conjecture, artificially "placed in weight 0". Otherwise (1.7) includes Tate twists....) Now – and this is correct for the Langlands-normalized L-functions –

$$L(s, R_{n+1}) = \frac{L(s, R_n \otimes R_2)}{L(s, R_{n-1})}.$$

If we know that R_n and R_2 are automorphic, we know that the Rankin L-function $L(s, R_n \otimes R_2)$ is good (this is due to Shahidi [**27**]). Therefore if $L(s, R_{n-1})$ is good so is $L(s, R_{n+1})$.

To show that all the $L(s, R_n)$ are good, we see by induction that it suffices to show that $L(s, R_n)$ for n **even** is automorphic. Further, by the previous argument, "potentially automorphic" is enough. Finally, we see that $ST(E)$ is implied by the following

THEOREM B. *Let r be the 2-dimensional Galois representation associated to $H^1(E)$.*

*For all **odd** n, there exists a totally real Galois extension F of $\mathbb{Q}$ such that*

$$\mathrm{res}^F_{\mathbb{Q}}(S^n r)$$

is automorphic, in fact, associated to a cuspidal representation Π of $GL(n+1, \mathbb{A}_F)$.

Note that we have not adopted Langlands' normalization here: π is in "weight 1" and $S^n r$ in "weight n". The cuspidal representation, then, will not be unitary. A twist by $|\det|^{\frac{n}{2}}$, where $|\ |$ is the idèle norm, reduces us to our previous notion of cuspidal representation.

2. Galois representations from cuspidal representations

The arguments in the 3rd reduction of § 1 were motivated by the Galois representation associated to r. In fact, in order to **prove** the existence of the representation Π in Thm. B, we will resort to the associated Galois representation, and an extension of Wiles' deformation theory. For this we need a large supply of Galois representations associated to cuspidal representations.

Denote by F a totally real number field or a CM-field (a quadratic, totally imaginary extension of a totally real number field F_+). The reader who prefers so may, again, consider only the case of $F = \mathbb{Q}$; this will not suffice, however, for the proof of Thm. B.

Consider a cuspidal (unitary) representation Π of $\mathrm{GL}(n, \mathbb{A}_F)$.To Π we wish to associate Galois representations of G_F; as in § 1 they will in fact be ℓ-adic representations. Denote by c complex conjugation (so c is trivial if F is real). Then Π^c is defined; we write $\tilde{\Pi}$ for the dual of Π.

We will need to make stringent assumptions on Π:

(a) Π_∞ **is cohomological**.

If for instance, $F = \mathbb{Q}$, so $\Pi = \bigotimes\limits_{v=p,\infty} \Pi_v$, let $\mathcal{H}$ be the Hilbert space of Π_∞. Then (a) means that

$$H^\bullet(\mathrm{GL}(n, \mathbb{R}), \Pi_\infty) \neq 0$$

for a suitable cohomology theory (cf. [**2**]). More generally, we may assume

$$H^\bullet(\mathrm{GL}(n, \mathbb{R}), \Pi_\infty \otimes V) \neq 0,$$

V being an algebraic, finite-dimensional complex representation of $\mathrm{GL}(n)$. A similar definition applies to any F, by considering all Archimedean primes.

(b) $\Pi \cong \tilde{\Pi}^c$.

Note that if F is real this means simply that Π is self-dual ($\Pi \cong \tilde{\Pi}$).

(c) For some finite prime v of F, Π_v belongs to the discrete series.

For this notion see, e.g., [**4**].

We now have the following theorem. If ℓ is prime, a λ-adic representation of G_F is a representation on a vector space over L_λ, a finite extension of $\mathbb{Q}_\ell$. In the statement v denotes a finite prime of F, and q_v the cardinality of the residue field.

Theorem (Kottwitz, Clozel, Harris, Taylor, Yoshida). *Assume* Π *is a cuspidal representation of GL*$(n, \mathbb{A}_F)$ *verifying* (a, b, c).

For any ℓ*, there exists a* λ*-adic representation* $R_\lambda(\pi)$ *of* G_F*, of degree* n*, such that*

(i) *If* Π *is unramified at* v *and* $v \nmid \ell$, $R_\lambda|_{G_v}$ *is unramified and* $R_\lambda(\mathrm{Frob}_v)$ *and* $t(\Pi_v)q_v^{\frac{n-1}{2}}$ *have the same characteristic polynomial.*

(ii) *For any* $v \nmid \ell$, $\Pi_v \ |\det|^{\frac{1-n}{2}}$ *and* $R_\lambda|_{G_v}$ *are associated by the Langlands local conjecture.*

A few comments:

• $t(\Pi_v)$ is the Hecke matrix of Π at an unramified prime v (for Π), which already occurred in (1.3) – for $F = \mathbb{Q}$.

- For a finite prime, $G_v \subset G_F$ is a decomposition group, isomorphic to $\mathrm{Gal}(\bar{F}_v/F_v)$.

- Note that (i) determines $R_\lambda(\mathrm{Frob}_v)$ only up to semi-simplification.

- For the p-adic field F_v, the **local conjecture** (proved by Harris and Taylor) associates representations of degree n of G_v and irreducible representations of $\mathrm{GL}(n, F_v)$. The Galois representations must be taken up to **Frobenius semi-simplification**: see [**30**], and the Introduction to [**13**].

  The local correspondence is then bijective.

- In many cases the family (L_λ) of ℓ-adic fields is given by the completions (at all finite primes) of a field of coefficients L associated to Π. It is not clear to me that this is always true – cf. the proof of [**13**, Thm. VII.1.9].

- There is a further property of R_λ, concerning the restriction to G_v for $v|\ell$. It is potentially semi-stable (a property of the Galois representations occurring in the cohomology of algebraic varieties). If Π_v is unramified it is crystalline. In both cases it has Hodge-Tate numbers $h^{i,d-i} (d = n-1)$. When Π_∞ is cohomological with trivial coefficients,

$$h^{0,n-1} = h^{1,n-2} = \cdots = h^{n-1,0} = 1 \ . \tag{2.1}$$

Only this case occurs in the proof of ST.

- Unlike (a,b), condition (c) is not a priori necessary: it is not needed in the case of GL(2) or GL(3). Further work on the Arthur trace formula is expected to remove it.

In order to justify the assumptions, we sketch the basic construction. If F is totally real, denote it by F_+ and choose a CM-extension F of it. Consider the diagram of groups

$$\begin{tikzcd} \mathrm{GL}(n,F) \arrow[r, phantom, "="] \arrow[d, dash] & \mathrm{GL}(n,F) \arrow[d, dash] \\ U(n,F_+) & \mathrm{GL}(n,F_+) \ . \end{tikzcd} \tag{2.2}$$

Here $U(n, F_+)$ is a unitary group of rank n over F_+. The vertical lines associate a group over F_+ and the group obtained by extension of scalars to F (say, G and $G_F = G \times_{F_+} F$).

As already seen in § 1, there is a theory of base change associating cuspidal representations π_+ of $\mathrm{GL}(n, \mathbb{A}_{F_+})$ and cuspidal representations [4] π of $\mathrm{GL}(n, \mathbb{A}_F)$ stable the action of c. Since the rational F-structure on $U(n, F_+)$ is twisted (from the F-structure of $\mathrm{GL}(n)$) by the outer automorphism of $\mathrm{GL}(n)$, we have likewise a base change relation:

[4] cuspidal in most cases, see [**1**]

$$\left\{\begin{matrix}\tau = \text{automorphic representation} \\ \text{of } U(n, \mathbb{A}_{F_+})\end{matrix}\right\} \longleftrightarrow$$
$$\left\{\begin{matrix}\pi = \text{cuspidal (?) representation} \\ \text{of } \mathrm{GL}(n, \mathbb{A}_F), \pi \cong \tilde{\pi}^c\end{matrix}\right\}.$$

If π_+ was self-dual, π will be self-conjugate **and** self dual, so $\pi \cong \tilde{\pi}^c$. Thus, either in the real and or in the complex case, we can hope by descent in the left-hand side of (2.2) to associate to our given π (or π_+) a representation of the unitary group.

The point is that, unlike $\mathrm{GL}(n)$ for $n > 2$, unitary groups define arithmetic quotients of the associated (Hermitian) symmetric spaces which are algebraic varieties. The sought representations are then obtained in the (étale) ℓ-adic cohomology of these varieties. Condition (a) ensures that τ will indeed be detected by the cohomology of the quotient variety, as in [**2**]; (b) was needed for descent; and (c) allows us to work with specific unitary groups for which all this can be proved. See [**5, 13**].

3. Conditional modularity

We fix a CM-field F (maximal totally real subfield F_+, complex conjugation c) and an n-dimensional representation R of G_F over an ℓ-adic field L_λ. We seek a cuspidal representation Π of $\mathrm{GL}(n, \mathbb{A}_F)$ such that $R = R_\lambda(\Pi)$ as in § 2.

Conditions on R:

(a) R **is crystalline at** ℓ.

This means that, for all primes v of F dividing ℓ, $R|_{G_v}$ is crystalline. (For a survey of the properties of ℓ-adic representations of ℓ-adic Galois groups see Illusie [**15**].) In particular it has a Hodge-Tate decomposition. Write

$$h^i = \dim_L((\mathbb{C}_\ell \otimes_L R)(i))^{G_p}, \ L = L_\lambda,$$

where i is the twist by the i-th power of the cyclotomic character. We assume moreover

(a1) $$h^0 = h^1 = \cdots = h^{n-1} = 1$$

which translates (2.1). (This suffices for ST. In general it suffices to assume that for all v, "the Hodge-Tate structure is regular", i.e., $h^i \leqslant 1$ for all i.)

(b) R **is conjugate self-dual**.

This means that $R^c \cong \tilde{R}\ \varepsilon^{1-n}$ where ε is the cyclotomic character $G_F \to \mathbb{Z}_\ell^\times$. (This twist is imposed by the usual translation between Π and R, which has already occurred repeatedly.)

(c) R **is Steinberg at** v **for some (finite)** v.

This is a condition of indecomposability on $R_v = R|_{G_v}$. By the local Langlands correspondence, it means that R_v is associated to a discrete series representation of $\mathrm{GL}(n, F_v)$. (The correct term is "generalized Steinberg", which includes supercuspidal representations.)

Note that these 3 conditions mirror those in § 2, except that (a) is not always true for (the Galois representations coming from) cuspidal representations: it is an assumption of good reduction.

We can choose a lattice in the space L_λ^n of R, fixed by R, and this defines a modular representation:

$$\bar{R} : G_F \to \mathrm{GL}(n, \mathcal{O}_\lambda) \to \mathrm{GL}(n, k_\lambda)$$

where $\mathcal{O}_\lambda$, k_λ are the integers and the residue field.

Up to semisimplification, $\bar{R}$ does not depend on the lattice. Assume in fact:

(d) $\bar{R}$ is absolutely irreducible and **big**.

We will explain this notion presently. We will also need the condition ("f" for "finite"):

(f) $\bar{R}$ is unramified except at a finite number of primes.

The notion of big.

In Wiles's proof had already appeared the condition (on a modular Galois representation) of having sufficiently large image that, by Cebotarev density, the images of suitable Frobenius elements are needed matrices in $\mathrm{GL}(2, k)$. See the proof of Thm. 2.49 in [**6**].

Assume $\bar{R}$ is an absolutely irreducible representation of G_F, of degree n, over a finite field k. Let $H \subset \mathrm{GL}(n, k)$ be its image. For $h \in H$, $\alpha \in k$, let $V_{h,\alpha}$ be the corresponding generalized eigenspace. Then we have unique, h-equivariant projections $\pi : k^n \to V_{h,\alpha}$ and injections $\iota : V_{h,\alpha} \to k^n$. Write $\mathfrak{g}\ell(n)^0$ for the matrices of zero trace in $M_n(k)$.

We say $\bar{R}$ is big if

(i) $H^1(H, \mathfrak{g}\ell(n)^0) = \{0\}$

(ii) For any irreducible $k[H]$-submodule W in $\mathfrak{g}\ell(n)$ we can find $h \in H$, $\alpha \in k$ such that $V_{h,\alpha}$ is one-dimensional, and $\pi \circ W \circ \iota \neq \{0\}$.

This definition makes the proofs work but is obviously unilluminating. To motivate it return to our elliptic curve E (without complex multiplication). We have the associated representations:

$$r_\ell : G_\mathbb{Q} \to \mathrm{GL}(2, \mathbb{Z}_\ell),\ \bar{r}_\ell : G_\mathbb{Q} \to \mathrm{GL}(2, \mathbb{F}_\ell).$$

Serre has shown [**26**] that r_ℓ is surjective for almost all ℓ. For given n, this easily implies [**HSBT**, Lemma 3.2] that $\mathrm{Sym}^n \bar{r}_\ell$ has big image for $\ell \gg 0$; in fact, the image of $\mathrm{SL}(2, \mathbb{F}_\ell)$ is already big. In fact, I expect the following:

(2.3.1) *Assume G is a semi-simple Chevalley group over $\mathbb{Z}$, and $R : G \to GL(n, \mathbb{Z})$ is irreducible* (over $\mathbb{Q}$ or $\mathbb{C}$). *Then, for almost all ℓ, $R(G(\mathbb{F}_\ell) \subset GL(n, \mathbb{F}_\ell)$ is big.*

This expresses the essence of the notion in applications; but compare with Step A in the proof of Thm. D (§ 4). The proof is an exercise for the reader.

I can now state, in a form sufficient for our purposes, the main result of [**CHT**]. This is conditional, in general, on a conjecture relative to automorphic forms on unitary groups, and which will be stated in § 7. It plays the rôle, in this context, of a famous lemma of Ihara brilliantly used by Ribet [**23**], Wiles and others. I call it Conjecture (I).

Start with a cuspidal representation Π_0 of $\mathrm{GL}(n, \mathbb{A}_F)$ verifying (a-c) of § 2. Let R_λ be the associated Galois representation and $\bar{R}$ its reduction, on $k = k_\lambda$. *We assume $\ell > n$, and ℓ unramified in F.*

Let K/F be the finite Galois extension of F associated to the subgroup $\ker(ad\ \bar{R})$ of G_F, where $ad\ \bar{R}$ is the adjoint representation. We introduce a last, technical condition:

(e) $F(\zeta_\ell)$ is linearly disjoint from K over F.

Theorem C [**CHT**, Thm. 4.3.4].

Assume $\bar{R}$ verifies (a-d) *and* (e).

Let R be a λ-adic representation, of reduction $\bar{R}$, verifying (a-c) *and* (f).

Assume conjecture (I) [5].

Then R is automorphic, i.e., there exists a cuspidal representation Π of $G(\mathbb{A}_F)$, verifying (a-c) *of* § 2, *such that $R = R_\lambda(\Pi)$.*

Remarks:

(i) The last statement is slightly incorrect, because there is an ℓ-adic coefficient field L_λ involved (in the Theorem of § 2). We may have to extend the field of definition of R and $R_\lambda(\Pi)$. (Harris and Taylor in [**13**] state their results over $\bar{\mathbb{Q}}_\ell$).

(ii) (a,b,c,d,f) are clearly true for $R = S^n r$, where r is associated to E, and ℓ is sufficiently large. The problem, of course, is the modularity of $\bar{R}$!

We will say a little more on the proof of this result in § 6.

[5] See § 6.

4. A remarkable family of Calabi-Yau varieties

We parametrize the projective line $\mathbb{P}^1$ by homogeneous coordinates $(s:t)$; we will often write $t=(1:t)$. Consider the following subvariety of $\mathbb{P}^{n-1}\times\mathbb{P}^1$, defined over $\mathbb{Q}$:

$$Y\subset\mathbb{P}^{n-1}\times\mathbb{P}^1$$

$$Y: s(X_0^n+X_1^n+\cdots+X_{n-1}^n)=n\ t\ X_0\cdots X_{n-1} \tag{4.1}$$

where $(X_0,\ldots,X_{n-1})$ are homogeneous coordinates in $\mathbb{P}^{n-1}$. We view Y as fibered over $\mathbb{P}^1$ by the second projection; thus Y defines a pencil of hypersurfaces $(Y_t)_{t\in\mathbb{P}^1}$ in $\mathbb{P}^{n-1}$. Note that for $t=0$ we get the Fermat hypersurface

$$X_0^n+X_1^n+\cdots+X_{n-1}^n=0 \tag{4.2}$$

already considered, after Weil, by Tate [**31**]; the computation of its cohomology can be found in Deligne's paper [**8**, Ch. I, § 7].

Let $T_0=\mathbb{P}^1-(\{\infty\}\cup\boldsymbol{\mu}_n)$.
Then $Y\to T_0$ is smooth, and in fact, defines a smooth projective map $Y\to T_0$ over $\mathbb{Z}[\frac{1}{n}]$.

The group $H=(\boldsymbol{\mu}_n)^n$ acts on Y by

$$X_i\mapsto\zeta_i\ X_i$$
$$t\mapsto\zeta_0\cdots\zeta_{n-1}t$$

for $\zeta=(\zeta_0,\ldots\zeta_{n-1})\in H$. Thus $H_0=\{\zeta:\Pi\zeta_i=1\}$ acts on Y_t $(t\in\mathbb{P}^1)$, with the diagonal subgroup $\{(\zeta,\ldots\zeta):\zeta\in\ \boldsymbol{\mu}_n\}$ acting trivially.

The smooth fibration $Y\to T_0$ yields a locally constant sheaf $V_\mathbb{Z}$ (in the complex topology):

$$(V_\mathbb{Z})_t=H^{n-2}(Y_t,\mathbb{Z})^{H_0}\ ,\ t\in T_0(\mathbb{C}).$$

This has, of course, variants in (algebraic) de Rham cohomology and in étale cohomology, over $\mathbb{Z}_\ell$ or $\mathbb{Q}_\ell$; in particular, if $\pi:Y\to\mathbb{P}^1$ is the projection,

$$(R^{n-2}\Pi_*(\mathbb{Z}_\ell))^{H_0}=V_{\mathbb{Z}_\ell}$$

is a lisse étale sheaf on $T_0/\mathbb{Q}$.

Fact. *$V_\mathbb{Z}$ is (locally constant) free of rank $(n-1)$ on $T_0(\mathbb{C})$.*

Since we know that $V_\mathbb{Z}$ is locally constant, it suffices to compute at $(t=0)$. This is done in [**8**, I, 7.4].

(The computation there is, ostensibly, done neglecting torsion; but it is known that cohomology of complete intersections such as (4.1) has no torsion [**7**].)

Of course, the same is true for $V_{\mathbb{Z}_\ell}$, $V_{\mathbb{Q}_\ell}$. Finally note that if $t\in T(\mathbb{C})$, the monodromy group $\pi_1(T_0(\mathbb{C}),t)$ acts on $(V_\mathbb{Z})_t$.

We now change notations, replacing n by $(n+1)$ in the defining equations. Moreover, **we now assume n even**. Thus:

V **is locally constant, free of even rank** n.

Moreover $(V_{\mathbb{Z}})_t$ is identified with $(H^{n-1}(Y_t, \mathbb{Z}))^{H_0}$; since it is the cohomology in the odd median degree of the smooth variety $Y_t(\mathbb{C})$, $H^{n-1}(Y_t, \mathbb{Z})$ carries a symplectic pairing; because the characters of H_0 occurring in $H^{n-1}(Y_t, \mathbb{C})$ do so with multiplicity 1 [**8**, I.7.4], this restricts to a non-degenerate pairing on $(V_{\mathbb{Z}})_t$ – again, we may compute at $(t = 0)$. Finally, this pairing is invariant by $\pi_1(T(\mathbb{C}), t)$ – par transport de structure.

The first main result in [**HSBT**] shows that the image of $\pi_1(T(\mathbb{C}), t)$ in $Sp(V_t)$ has maximal size:

Theorem ([**HSBT**, Cor. 1.10]). *For $t \in T_0(\mathbb{C})$, the image of $\pi_1(T_0(\mathbb{C}), t)$ in $Sp(V_t \otimes \mathbb{C}$ is Zariski-dense.*

We now combine this with a very useful theorem of Matthews, Vaserstein and Weisfeiler. Assume G is a semi-simple group over $\mathbb{Q}$: thus G is in fact, defined over $\mathbb{Z}[\frac{1}{N}]$ for suitable N, and $G(\mathbb{F}_p)$ is defined for sufficiently large p. If $\Gamma \subset G(\mathbb{Q})$ is a finitely generated subgroup, $\Gamma \subset G(\mathbb{Z}_{(p)})$ for p large, so the image of Γ in $G(\mathbb{F}_p)$ is defined.

Theorem ([**20**]). *Assume $\Gamma \subset G(\mathbb{Q})$ is finitely generated and Zariski-dense in $G(\mathbb{C})$. Then, for p sufficiently large, $\Gamma \to G(\mathbb{F}_p)$ is surjective.*

As the authors point out the proof of this theorem in [**20**] relies on the classification of finite simple groups.

Another proof is given by Nori in [**22**, Thm. 5.1]; it would not depend on the classification, but Nori does not give the details. Yet another proof has been given by Hrushovski and Pillay using definability theory [**14**, Prop. 7.3]. The reader can choose the proof suiting his philosophical preferences.

At any rate, we get from the two theorems:

Corollary. *There exists a constant $C(n)$ such that, if $\ell_1 \neq \ell_2$ are two primes and $\ell_1, \ell_2 > C(n)$,*

$$\pi_1(t, T_0(\mathbb{C})) \to Sp(V_t(\mathbb{Z}) \otimes \mathbb{Z}/\ell_1\ell_2\mathbb{Z})$$

is surjective.

Now let $W_{\ell_1\ell_2}$ denote a fixed $\mathbb{Z}/\ell_1\ell_2\mathbb{Z}$-module, free of rank n, and endowed with a symplectic pairing. Denote by $T_{W_{\ell_1\ell_2}}$ the étale covering of T_0 (over $\mathbb{C}$) defined by

$$T_{W_{\ell_1\ell_2}}(t) = \mathrm{Isom}(W_{\ell_1\ell_2}, V_{\ell_1\ell_2}(t))$$

where $V_{\ell_1\ell_2}(t) = V_t(\mathbb{Z}) \otimes \mathbb{Z}/\ell_1\ell_2\mathbb{Z}$ and the isomorphisms are symplectic. This is a principal bundle over $T_0(\mathbb{C})$ with fiber $Sp(n, \mathbb{Z}/\ell_1\ell_2\mathbb{Z})$; from the previous Corollary we get:

Corollary. *$T_{W_{\ell_1\ell_2}}$ is a connected quasi-projective variety (over $\mathbb{C}$).*

We can endow this variety with a $\mathbb{Q}$-structure, or more generally with an F-structure, if F is a number field. For this fix:

- F = a number field, $G_F = \mathrm{Gal}(F/F)$;

- W = a free $\mathbb{Z}/(\ell_1\ell_2)$-module of rank n, with a representation of G_F and an alternating pairing

$$W \times W \to (\mathbb{Z}/\ell_1\ell_2\mathbb{Z})(1-n),$$

thus:

$$(\sigma w, \sigma w') = \varepsilon(\sigma)^{1-n}(w, w') \tag{4.3}$$

where $\varepsilon : G_F \to (\mathbb{Z}/\ell_1\ell_2\mathbb{Z})^\times$ is the cyclotomic character. Define

$$T_W \to T_0 \quad /\bar{\mathbb{Q}}$$

by $(T_W)_t = \mathrm{Isom}(W, V_{\ell_1\ell_2}(t))$.

Then T_W is a finite étale covering of T_0 (over $/\bar{\mathbb{Q}}$); over $\mathbb{C}$ it is $T_{W_{\ell_1\ell_2}}$; moreover T_W is defined over F.

In particular T_W is geometrically connected. Note that an F-point of T_W is given by $t \in T_0(F)$ and by an isomorphism of G_F-modules between W and $H^{n-1}(Y_t, \mathbb{Z}/\ell_1\ell_2\mathbb{Z})^{H_0}$.

We can view W as the data of two modular representations $\bar{r}_{\ell_1}$, $\bar{r}_{\ell_2}$, which are moreover symplectic, with the natural multiplier given by (4.3). A point in $T_W(F)$, then, yields Y_t such that the natural representation on $H^{n-1}(Y_t)^{H_0}$, reduced mod ℓ_1 **and** ℓ_2 is isomorphic with $\bar{r}_{\ell_1}$ and $\bar{r}_{\ell_2}$. However, this cohomology space yields a strongly compatible system of Galois representations (for varying ℓ). Thus, if v is a place of F where Y_t has good reduction, and dividing neither ℓ_1 nor ℓ_2, we have

$$\begin{aligned} &\mathrm{trace}(\mathrm{Frob}_v \mid H^{n-1}(Y_t, \mathbb{Q}_{\ell_1})^{H_0}) \\ &\qquad = \mathrm{trace}(\mathrm{Frob}_v \mid H^{n-1}(Y_t, \mathbb{Q}_{\ell_2})^{H_0}). \end{aligned} \tag{4.4}$$

In fact, we have the stronger property that the characteristic polynomials of Frob_v in the two representations are equal.

We include in this section a last ingredient which will be crucial in the next paragraph.

A theorem of Moret-Bailly

Assume

- F = a number field

- S = a finite set of primes (Archimedean or not) of F

- T = a smooth, geometrically connected quasi-projective variety over F.

For $v \in S$, let $\Omega_v \subset T(F_v)$ be a non-empty open subset (for the topology given by the local field).

Theorem ([**21**]). *There existe a finite Galois extension K/F such that:*

(i) $v \in S \Rightarrow v$ *splits in* K.

(ii) *There exists a point $P \in T(K)$ such that, for $v \in S$ and w (a place of K) dividing v:*

$$P_w \in \Omega_v \subset T(F_v) = T(K_w).$$

Note that because v splits, F_v is naturally isomorphic with K_w. Also remark that if F is totally real, we also obtain K totally real.

5. Potential modularity

In this paragraph we will sketch the proof, in [**HSBT**] of a result of **potential** modularity in the sense of § 1. This applies to a symplectic representation of $G_{\mathbb{Q}}$ (with $\mathbb{Q}_\ell$-coefficients). The result **does not suppose** residual modularity (modularity of the associated representation mod ℓ). On the other hand, it is **conditional** on Conjecture (I) (§ 7).

We have stated the result, consistent with our general policy, with $\mathbb{Q}$ as the ground field. However, as we will see later, this is not sufficient for the proof of Theorem B (§ 1), so of Sato-Tate. A similar result holds when $\mathbb{Q}$ is replaced by a totally real field [**HSBT**, Theorem 3.1].

Assume ℓ is a prime such that

- $\ell > C(n)$, $\ell \equiv 1 \quad [n+1]$.

Given:

- $r : G_{\mathbb{Q}} \to GSp(n, \mathbb{Z}_\ell)$.

We make the following assumptions. As in § 2 $G_p \subset G_{\mathbb{Q}}$ denotes a decomposition group, $G_p \cong \mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)$. We denote by I_ℓ the inertia subgroup of G_ℓ, and by ε_ℓ the cyclotomic character on G_ℓ.

- **Assumptions on** r.

(a) r is crystalline at ℓ, with Hodge-Tate weights $(0, 1, \dots, n-1)$. Moreover

$$\bar{r}|_{I_\ell} \cong 1 \oplus \bar{\varepsilon}_\ell^{-1} \oplus \cdots \oplus \bar{\varepsilon}_\ell^{1-n}.$$

(b) r has multiplier ε^{1-n}.

(c) At a prime $q \neq \ell$, $q \nmid n+1$ such that $1, q, \dots, q^{n-1}$ are distinct mod ℓ, the semisimplification of $r|_{G_q}$ is unramified, $\bar{r}|_{G_q}$ is unramified, and $r(\mathrm{Frob}_q)$ has eigenvalues $(1, q, \dots, q^{n-1})$.

(d) $\bar{r}$ is irreducible and **big**.

(e) $\mathbb{Q}(\zeta_\ell)$ is not contained in K, where $K/\mathbb{Q}$ is associated to $\ker(ad\ \bar{r})$.

(f) r ramifies at only finitely many primes.

Note that these are the assumptions which occurred in § 3; in particular (c) is an explicit form of "Steinberg at v" there. Note however than (a) is stronger than (a) of § 3, where we just assumed r crystalline. We now have:

Theorem D [**HSBT**, Thm. 3.1]. *Assume Conjecture* (I).
Then, if r verifies $(a-f)$, there exists a Galois, totally real extension $F/\mathbb{Q}$ such that $\bar{r}_F$ has the same image as $\bar{r}$ and r_F is modular.

This is a difficult theorem, so we can only give a **hint** of the proof. We do it is steps.

Step A. Choose a large ("absurdly large") prime ℓ' and construct $r' = r_{\ell'} : G_{\mathbb{Q}} \to \mathrm{GL}(n, L)$, where L is an ℓ'-adic field with residue field $k_L = \mathbb{F}_{\ell'}$. The representation r' must have the following properties:

- r' is automorphic, essentially self-dual;
- $\bar{r}' : G_{\mathbb{Q}} \to GSp(n, \mathbb{F}_{\ell'})$ (and moreover the multiplier is $\varepsilon_{\ell'}^{1-n}$).
- $\bar{r}'$ is **large** and verifies the conditions for conditional modularity (Theorem C).

The representation r' is constructed by induction:

$$r' = \mathrm{ind}_{G_M}^{G_{\mathbb{Q}}}(\chi)$$

where $M/\mathbb{Q}$ is a CM-field (note that the degree is even!) and χ is a suitable algebraic Grössencharakter of M, identified with its ℓ-adic representation of G_M. The field M is in fact, an Abelian extension of $\mathbb{Q}$. Then r' is automorphic, by the base change results [**1**] already mentioned at the end of § 1. However the difficulty is to ensure that $\bar{r}'$ is (essentially) symplectic and verifies the other conditions in § 3. The existence of χ is proved in [**CHT**, § 4.3].

Note that, by the usual properties of induction-restriction, $r'|_{G_F}$ is still automorphic for any extension F of $\mathbb{Q}$ – automorphic in our usual sense, "associated to a cuspidal representation of $\mathrm{GL}(n, \mathbb{A}_F)$" if the restriction to G_F remains irreducible.

Step B. Representations r and r' define (irreducible) reductions $\bar{r}$, $\bar{r}'$. This defines a symplectic Galois module $W_{\ell\ell'} = W$, so a variety T_W over $\mathbb{Q}$ (§ 4).

Step C. We now define a suitable set of local conditions, in order to apply Moret-Bailly's theorem. These pertain to a finite set S of rational primes. For $q = \ell, \ell'$ let $V_{t,q}$ denote $V_t \otimes \mathbb{Q}_q$ ($t \in T_0$) seen as a local Galois representation. For $p \in S$, the local conditions should specify an open subset Ω_p in $T_w(\mathbb{Q}_p)$ such that, for $x \in \Omega_p$, the associated representation of $G_p = \mathrm{Gal}(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)$ on $V_{t,q}$ has certain properties (q may be equal to p); here $t \in T_0(\mathbb{Q}_p)$ is the image of x. For instance:

• For $p = \ell$ the condition insures that $r_{t,\ell}$ is crystalline (with the correct Hodge-Tate weights).

• Similar conditions in ℓ'.

• For some $q \neq \ell, \ell'$ the condition ensures that $r_{t,\ell}$ and $r_{t,\ell'}$ are "Steinberg" ar q.

• At infinity simply take $\Omega_\infty = T_W(\mathbb{R})$; this is not empty because complex conjugation acts on $V_t \otimes \mathbb{Q}_q$, for $t \in T_0(\mathbb{R})$, as the only element of $GSp(\mathbb{Q}_\ell)$ (up to conjugation) of order 2 and multiplier (-1).

We return to the first condition, the discussion for ℓ' being similar. Note that $x \in T_W(\mathbb{Q}_\ell)$, so it does not suffice to exhibit $t \in T_0(\mathbb{Q}_\ell)$. We would want $t \in T_0(\mathbb{Q}_\ell)$ such that the associated representation of G_ℓ on $V_{t,\ell}$ verifies (a). This cannot be ensured on $\mathbb{Q}_\ell$, but it can be realized on an unramified extension by Lemmas 1.13 and 1.14 of [**HSBT**]. The set of points so obtained is then open.

(The fact that we may have to consider (unramified) extensions of $\mathbb{Q}_\ell$ forces us, in fact, to use a strengthening of Moret-Bailly's theorem: for the statement see [**HSBT**, Prop. 2.1]. We neglect this in the rest of the sketch, and other technical details as well.)

Step D. By Moret-Bailly's theorem, we can find a finite Galois extension $F/\mathbf{Q}$, and $x \in T_W(F)$ verifying our local conditions. In fact, F is totally real (since the prime ∞ splits in F). Further (again, see [**HSBT**, Prop. 2.1]) we can take F linearly disjoint from ker $\bar{r}_\ell$ and ker $\bar{r}_{\ell'}$. Now, the representations in the following being restricted to G_F, we see that for t equal to the projection of x:

• $\bar{r}_{t,\ell'}$ is modular (**and verifies the conditions of Theorem** C) $\Rightarrow r_{t,\ell'}$ is modular.

Now recall that $r_{t,\ell}$ and $r_{t,\ell'}$ are two specializations of a compatible system of Galois representations. Therefore:

• $r_{t,\ell'}$ modular $\Rightarrow r_{t,\ell}$ modular which implies, by definition, that $\bar{r}_{t,\ell}$ is modular. But by the definition ot T_W, $\bar{r}_{t,\ell}$ is $\bar{r}_\ell$. Since $\bar{r}_\ell$ verifies the assumptions of Theorem C, we have now:

• $\bar{r}_\ell$ modular $\Rightarrow r_\ell$ modular
which concludes the proof.

Besides the technical details omitted above, we have glossed over two points. First, Theorem C (§ 3) was formulated for Galois representations of G_F where F is a CM-field, while we are working with symplectic representations of $G_{\mathbb{Q}}$. We can choose a quadratic imaginary field E; the arguments leading to step D gives us, by restriction, a representation of G_{EF} which will (trivially for a suitable choice of E) verify all conditions of Theorem C. We then associate to $r_\ell|_{EF}$ a cuspidal representation of $\mathrm{GL}(n, \mathbb{A}_{EF})$. This representation is invariant by

$\mathrm{Gal}(EF/\mathbb{Q})$ so descends to a cuspidal representation of $\mathrm{GL}(n, \mathbb{A}_F)$ by [**1**]. In fact, it descends to two representations π and $\pi \otimes \varepsilon$ where ε is the quadratic character of $\mathbb{A}_F^\times$ defined by $EF/\mathbb{Q}$. But since we know the existence of a representation of G_F associated (for example) to π, an argument of descent already seen in § 1 ("Brauer-Taylor") shows that r is associated to π or $\pi \otimes \varepsilon$. (See [**CHT**, Lemma 4.3.2].)

The next point is more serious. For Theorem C we had to assume ℓ unramified in the CM-field, here EF. We can choose E unramified at ℓ; but the foregoing proof must also produce an unramified F. However, as we have seen, this is implied by the local condition (a), already forced on us by the previous argument. (The splitting property in Moret-Bailly's theorem implies that it is true in F.)

Of course, we would now want to apply Theorem D to the $\mathbb{Z}_\ell$-representation $r = \mathrm{Sym}^{n-1} r_E$ where r_E is the 2-dimensional representation afforded by H^1 of our elliptic curve. We inspect the conditions imposed on r. Conditions (b,d,e,f) are easily checked if r is sufficiently large. So is crystallinity in (a); the first part of (c) will be true for q equal to the prime of bad reduction of E, perhaps after replacing $\mathbb{Q}$ by a quadratic (totally real) extension. In particular, we already see that we must apply Theorem D, not to $\mathbb{Q}$ in general, but to a totally real field. This is, of course, the stronger statement contained in [**HSBT**].

Even then, however, there is no reason why the second part of (a):

$$\text{(a}'\text{)} \qquad \bar{r}|_{I_\ell} = 1 \oplus \bar{\varepsilon}_\ell^{-1} \oplus \cdots \oplus \bar{\varepsilon}_\ell^{1-n}$$

as well as:

$$\text{(c}'\text{)} \qquad \bar{r}|_{G_q} \text{unramified}$$

in (c), should be verified for some ℓ.

However, we can first choose a totally real field F, and an elliptic curve E' over F, such that $S^{n-1}(r_{E'})$ verifies all the conditions imposed on r (. . . for the extension to F of Theorem D). These conditions can in fact, be verified directly on the 2-dimensional representation $r_{E'}$; the verification then relies on a version of Hilbert's irreducibility theorem.

Consider the modular scheme over F of elliptic curves $\mathcal{E}$ endowed with an isomorphism $W \cong H^1_{\text{ét}}(\mathcal{E} \times \bar{F}, \mathbb{Z}/\ell\ell'\,\mathbb{Z})$ where W is now the symplectic G_F-module defined by $H^1(E' \times \bar{F}, \mathbb{Z}/\ell\mathbb{Z}) \times H^1(E' \times \bar{F}, \mathbb{Z}/\ell'\mathbb{Z})$. As in § 4 this defines a (geometrically connected) variety, in fact, a twisted modular curve.

Imposing suitable local conditions, and arguing as in the earlier proof, we find a totally real field F' and a point of our variety over F', ie. a curve E'' with

$$H^1(E'' \times \bar{F}, \mathbb{Z}/\ell\mathbb{Z}) \cong \bar{r}_{E,\ell}$$

$$H^1(E'' \times \bar{F}, \mathbb{Z}/\ell'\mathbb{Z}) \cong \bar{r}_{E',\ell'}$$

the isomorphisms, of course, being under the action of $G_{F'}$. Now by construction, $S^{n-1}(r(\bar{r}_{E',\ell'})$ verifies all the conditions of Theorem D, so (after a further extension F'') $S^{n-1}(r_{E'',\ell'})$ is modular. By strict compatibily so is $S^{n-1}(r_{E'',\ell})$. But then $S^{n-1}(\bar{r}_{E,\ell})$ is modular, and we are reduced again to an application of Theorem C. Finally, as in the earlier case, we have to check in the constructions that the final field F'' is unramified at ℓ. See [**HSBT**, § 3].

The final outcome is that $\mathrm{Sym}^{n-1} r_E$ becomes modular, or automorphic, on some Galois, totally real extension F of $\mathbb{Q}$. This is still not good enough, for the argument at the end of § 1, bases on Clebsch-Gordan, for taking case of odd-dimensional representations, imposes the existence of a uniform F for a finite number of odd powers S^{n-1}. This imposes a further complication on the correct statement of Thm D. (See the final version of Theorems 3.1, 3.3 in [**HSBT**].) This completes our sketch of proof for Theorem B.

6. Automorphic forms on $\mathbb{R}$-compact unitary groups and “Ihara’s lemma”

6.1. In this section we want to give more details on the theory of automorphic forms on $\mathbb{R}$- anisotropic unitary groups which leads to the deformation theory of § 3, and state the conjecture (I) on which depends the proof of Theorem B sketched in § 6. (Recall that this dependence is removed by Taylor’s final argument, for which see § 7.) The proofs rely systematically on the theory of automorphic forms on adèle groups, on Wiles’ deformation theory, and on the Harris-Taylor solution of the local/global Langlands conjecture [**13**]. We will expect from the reader some familiarity with these topics.

Recall the essential content of Theorem C. We will simply take here a quadratic imaginary field, which we denote by F. Thus $F^+ = \mathbb{Q}$. We start with a cuspidal representation Π_0 of $\mathrm{GL}(n, \mathbb{A}_F)$ verifying (a,b,c) of § 2. Then there is an associated λ-adic representation $R_0 = R(\Pi_0)$, n-dimensional and which we assume irreducible. If k is the residue field of L_λ, we assume in fact, $\bar{R}$ irreducible (over k). If now R is another λ-adic representation of G_F, verifying suitable conditions, we want to show that R is similarly an $R(\Pi)$.

The proof uses automorphic forms on unitary groups. Thus let

- G = unitary group of rank n over $\mathbb{Q}$, relative to $F/\mathbb{Q}$.

So far we may assume that G is the unitary group of a Hermitian form on F^n. Thus $G(F) \cong \mathrm{GL}(n, F)$. We further assume:

- $G(\mathbb{R})$ is compact.

Now let $K \subset G(\mathbb{A}_f) = \prod'_p G(\mathbb{Q}_p)$ (restricted product) be a compact-open subgroup. Because $G(\mathbb{R})$ is compact,

$$X_K = G(\mathbb{Q})\backslash G(\mathbb{A})/G(\mathbb{R})K$$

is a finite set. If R is any commutative ring, we can consider the finite, free R-module

$$\mathcal{A}(R) = \mathcal{A}_K(R) = C(X_K, R)$$

of functions on X_K with values in R. Our rings of interest will be $L = \mathbb{Q}_\ell$, $\mathcal{O} = \mathbb{Z}_\ell$, $k = \mathbb{F}_\ell$ (or finite extensions of such) and $\mathbb{C}$. Trivially, $\mathcal{A}(\mathcal{O}) \otimes k = \mathcal{A}(k)$, $\mathcal{A}(\mathcal{O}) \otimes L = \mathcal{A}(L)$. If we choose an embedding $\mathbb{Q}_\ell \subset \mathbb{C}$, $\mathcal{A}(\mathbb{Q}_\ell) \otimes \mathbb{C}$ is identified with the usual space of automorphic forms, $\mathcal{A}(\mathbb{C})$.

By right convolution, the Hecke algebra

$$\mathfrak{H}(\mathbb{Z}) = \mathrm{Hecke}(G(\mathbb{A}_f), K)$$

of $\mathbb{Z}$-valued functions on $G(\mathbb{A}_f)$, bi-invariant by K, acts on $\mathcal{A}(\mathbb{Z})$. Tensoring with R, we get an action of $\mathfrak{H}(R)$ on $\mathcal{A}(R)$. In general, $\mathfrak{H}$ is not commutative; however the local factors

$$\mathfrak{H}_p = \mathrm{Hecke}(G(\mathbb{Q}_p), K_p)$$

will be commutative if K_p is a "good" maximal subgroup of $G(\mathbb{Q}_p)$. (We can take $K = \prod_p K_p$, and this will then be the case at almost all p.) At the ramified primes, one describes an explicit, commutative subalgebra of $\mathfrak{H}_p$, and $\mathfrak{H}$ will be the ensuing "restricted" algebra.

As in the classical theory for GL(2), we now denote by $\mathcal{T}$ the **image** of

$$\mathfrak{H}_{\mathbb{Z}_\ell} \to \mathrm{End}(\mathcal{A}_{\mathbb{Z}_\ell}).$$

It is a commutative algebra, finite and free (as a module) over $\mathbb{Z}_\ell$. Assume $f \in \mathcal{A}(\mathbb{Z}_\ell)$ is an eigenform of $\mathcal{T}$. Extending scalars, we obtain a complex eigenform $f_\mathbb{C}$ for the Hecke algebra $\mathfrak{H}_\mathbb{C}$, i.e., a "classical" automorphic form on $G(\mathbb{A})$ invariant by $G(\mathbb{R})$. (A similar construction can be made by introducing further a finite-dimensional representation V of $G(\mathbb{R})$, but we will reglect this.)

Now $f_\mathbb{C}$ belongs to the space of a finite sum of irreducible representations τ of $G(\mathbb{A}_f)$ on

$$L^2(G(\mathbb{Q})G(\mathbb{R})\backslash G(\mathbb{A}_f)).$$

If we are able to effect the base change explained in § 2, we will be able to associate to (each) τ an automorphic representation Π of $\mathrm{GL}(n, \mathbb{A}_E)$. In fact, Π is uniquely defined by the character of $\mathcal{T}$ associated to f, by strong multiplicity one. (We assume, which will generally be the case, that Π is **cuspidal**.)

In order to start deformation theory we need a Galois representation associated to Π. Conditions (a), (b) of § 2 will be automatically satisfied; however we need the "square-integrability" condition (c). For G

a true unitary group and Π so obtained by base change, this condition will generally not be satisfied.

We must, then return to our choice of unitary groups. Recall that there are "twisted" unitary groups, associated to division algebras. We let D be a division algebra (of rank n, degree n^2) over the quadratic imaginary field F, endowed with an involution of the second kind, i.e., an anti-automorphism $*$ of order 2:

$$(d_1 d_2)^* = d_2^* d_1^*$$

inducing on $F = \mathrm{Cent}(D)$ the Galois action of $\mathrm{Gal}(F/\mathbb{Q})$. We then define G by

$$G(\mathbb{Q}) = \{d \in D \mid dd^* = 1\}$$

and **assume** that $G(\mathbb{R})$ is compact. It is assumed that there is a rational prime p such that F splits at p, $D^\times(\mathbb{Q}_p \otimes F) \cong G(\mathbb{Q}_p) \times G(\mathbb{Q}_p)$ and (for the two primes π, π' dividing p) $D(F_\pi) \cong D(F_\pi)^{\mathrm{opp}}$ is a division algebra. The base change Π will then be an automorphic representation of $D^\times(\mathbb{A}_F)$, which transports to $\mathrm{GL}(n, \mathbb{A}_F)$ (say, into Π') by a known form of "Jacquet-Langlands". Then Π' will be a discrete series at the two primes π, π' and we obtain a Galois representation of G_F as in § 2 [6].

Taking suitably large rings of coefficients, we now have an ℓ-adic field (L, $\mathcal{O}$, k, $\lambda =$ prime ideal), an n-dimensional representation R (over $\mathcal{O}$ or L) and its irreducible reduction $\bar{R}$. Following Wiles, the next step is to define a universal deformation algebra $\mathcal{R}(\bar{R})$. It should have the property that, for S a complete, local, topolgically finitely generated $\mathcal{O}$-algebra, with residue field k:

$$\mathrm{Hom}(\mathcal{R}(\bar{R}), S) \cong \left\{ \begin{matrix} \text{representations } R' : G_F \to \mathrm{GL}(n, S) \\ \text{of unitary type, such that } \bar{R}' = \bar{R} \end{matrix} \right\}. \tag{6.1}$$

(In the right-hand side we have to identify representations conjugate by $1 + M_n(\lambda_S)$ is the maximal ideal.)

There are difficult problems in defining $\mathcal{R}$. First, the representations of G_F that will come from automorphic forms must be "of unitary type", i.e. satisfy condition (b) in § 3. Since (b) means "there exists an isomorphism between R^c and $\tilde{R}\ \varepsilon^{1-n}$" it does not lead to a good deformation problem: the isomorphism must be specified on the data. This leads to the consideration, not of representations of G_F, but of morphisms of $G_\mathbb{Q}$ into an extension of $\mathrm{GL}(n)$ that incorporates the intertwining: this is a variant of Langlands' dual group of a unitary group.

Under a suitable parity condition on the image of the generator c of $\mathrm{Gal}(F/\mathbb{Q})$, the deformation theory of such "representations" can be studied as in Wiles' paper (for the parity condition see [**CHT**, Lemma

[6] Here the reader may think that we could have **chosen** f so Π has this property (such f exist). However the deformation theory, as described in [**CHT**] seems to require this for all f.

1.1.2 and Theorem 3.1.1]. As usual, one must impose suitable conditions on R' in (6.1), and then compute the "cotangent space" $\mathcal{R}/ < \mathfrak{m}_{\mathcal{R}}^2, \lambda >$. The essential condition that one must impose is to consider only:

- crystalline deformations (for the restriction to G_ℓ).

This is described using Fontaine-Laffaille theory – which imposes $\ell > n$.

- "Steinberg" or more generally "discrete series" deformations at some $p \neq \ell$ (the ramification prime specified above).

Once these conditions are specified, it is possible as in the classical theory to compute the cotangent space

$$\mathcal{R}/ < \mathfrak{m}_{\mathcal{R}}^2, \lambda) = H^1_{\mathcal{S}}(G_S, \mathrm{ad}\ \bar{R}).$$

Here S is a finite set of primes, $G_S \subset G_{\mathbb{Q}}$ is the Galois group of the maximal extension of $\mathbb{Q}$ ramified only at S, and S must contain at least ℓ and p; $H^1_{\mathcal{S}}$ is a subspace of H^1 specified by the family $\mathcal{S}$ of local conditions. See [**CHT**, § 1.4].

Now remember that we have chosen an eigenform f for $\mathcal{T}$, which gives rise to a Galois representation R (over L). It may be seen, as in the classical theory (see [**6**, Lemme 3.27]) that there is actually a naturally defined representation $R_{\mathcal{T}}$ **over** $\mathcal{T}$, where $\mathcal{T}$ is now restricted in the following manner: the full algebra $\mathcal{T}$ is endowed with

$$\mathcal{T} \to \mathcal{O}$$

defined by f; the prime ideal λ then defines a maximal ideal by f; the prime ideal λ then defines a maximal ideal $\mathfrak{m}$ in $\mathcal{T}$ and we localize at $\mathfrak{m}$. We denote by $\mathcal{T}$ this new, localized algebra. (It is a semi-local ring, free over $\mathcal{O}$.) See [**CHT**, § 2.4].

The local properties of $R_{\mathcal{T}}$ can be checked [**CHT**, Prop. 2.4.2] and by the universal property we get a uniquely defined morphism

$$\mathcal{R} \to \mathcal{T}.$$

As "usual", the point if to check that this is an isomorphism. Just as in the classical situation there are now two different cases.

6.2. Recall that the constructions in § 6.1 used modular forms of fixed level, denoted by K. (Now the "level" groups considered must be precisely described, see [**CHT**, § 2.1].) The form f comes with a natural level $K = K(f)$. We would like to show that any (suitable) representation R' with $\bar{R}' = \bar{R}$ is automorphic, associated to a form f'. However the ramification K' of f' is not determined by that of $\bar{R}$. Deformations in **minimal level** are those where K is kept fixed; but the ramification of f' may be deeper.

The **minimal level** case can be proved directly: one proves an identity

$$\mathcal{R}_\emptyset \underset{\approx}{\to} \mathcal{T}_\emptyset \tag{6.2}$$

where $\mathcal{T}_\emptyset$ is our original $\mathcal{T}$ (no level added) and $\mathcal{R}_\emptyset$ is defined by the set $\mathcal{S}$ of local conditions defined by r_f (in particular, no additional ramification on the Galois side). This is difficult but, except for the complications dues to the higher dimension, is proved as in Taylor-Wiles – the proof incorporates the elaboration of their method dues to Diamond and Fujiwara. A very clear description of the commutative algebra involved has been given by Fujiwara [**9**].

The next step consists in passing from the minimal to the general case, and it is this step which is conditional. We need a variant of Ihara's classical lemma for modular forms, see [**23**]. We will not give the classical formulation of Ihara's lemma (for $\mathrm{GL}(2)/\mathbb{Q}$) but explain its variant for a group $G = U(2)/\mathbb{Q}$ (as usual, compact at the Archimedean prime).

So far we have been using spaces $\mathcal{A}_K$ of automorphic forms of fixed level. Now we may vary K and pass to the limit, obtaining for any R:

$$\mathcal{A}(R) = \varinjlim_{K} \mathcal{A}_K(R).$$

This is the space of R-valued functions on $G(\mathbb{Q})G(\mathbb{R})\backslash G(\mathbb{A})$, invariant by K for **some** K. It carries a representation (on the right) of $G(\mathbb{A}_f)$. We may also fix a prime $p \neq \ell$, take $K = UK^p$ with $K^p \subset G(\mathbb{A}_f^p)$ fixed, and vary U. Then

$$\mathcal{A}_{K^p}(R) = \varinjlim_{U} \mathcal{A}_{UK^p}(R)$$

is a representation of $G(\mathbb{Q}_p)$. We assume $G(\mathbb{Q}_p) \cong \mathrm{GL}(2, \mathbb{Q}_p)$.

On $\mathcal{A}_{K^p}$ we have the action of $\mathcal{H}^p = \bigotimes_{q \neq p} \mathcal{H}(G(\mathbb{Q}_q), K_q)$. The factors for $\ell \notin S$ (finite) are commutative. Assume now $R = k$ (a finite extension of $\mathbb{F}_\ell$), let $\psi : \mathcal{H}^S = \bigotimes_{q \notin S} \mathcal{H}_q \to k$ be a character, and consider the eigenspace

$$\mathcal{A}_{K^p}(k)^\psi \subset \mathcal{A}(k)$$

for $\mathcal{H}^s$. (We assume $\ell \in S$.) We say ψ is **Eisenstein** if there is a one-dimensional ($\equiv$ Abelian) subrepresentation of $\mathcal{A}(k)$ on which $\mathcal{H}^s$ acts by ψ.

Since $p \neq \ell$, there is, as in characteristic zero, a notion of **generic** representation of $\mathrm{GL}(2, \mathbb{Q}_p)$. (These are the representations which admit "Whittaker functionals". For the full theory in finite characteristic see Vignéras [**36**].)

PROPOSITION 6.1. *Assume $\ell > 2$. If ψ is not Eisenstein, any irreducible $G(\mathbb{Q}_p)$-submodule of $\mathcal{A}_{K^p}(k)^\psi$ is generic.*

In this case, in fact, an irreducible representation of $G(\mathbb{Q}_p)$ is generic or one dimensional. (Recall that $n = 2$!) However, if $G(\mathbb{Q}_p)$ acts by a character, $\mathrm{SL}(2, \mathbb{Q}_p)$ acts trivially. But

$$G'(\mathbb{Q})\mathrm{SL}(2, \mathbb{Q}_p)\mathrm{SU}(2)$$

is dense in $G'(\mathbb{A})$ (by strong approximation), where $G' = \mathrm{SU}(2)/\mathbb{Q}$ is the derived group. So a non-generic form in $\mathcal{A}(k)$ is invariant by $\mathrm{SU}(2)$, so is (one-dimensional) character.

We can now state the form of Ihara's lemma mentioned earlier in the text. We assume that G is one of our particular unitary groups. In particular Galois representations can be associated to forms on $G(\mathbb{A})$. We will say that a character ψ of $\mathcal{H}^S$ is **non-Eisenstein** if it is associated to a representation Π of $\mathrm{GL}(n, \mathbb{A}_F)$ such that the associated, reduced, Galois representation $\bar{R}$ is irreducible.

CONJECTURE I. *Assume $\psi : \mathcal{H}^S \to k$ is non-Eisenstein, and π is an irreducible representation (over $k/\mathbb{F}_\ell$) of $G(\mathbb{Q}_p)$. If π is an irreducible submodule of*

$$\mathcal{A}_{K^p}(k)^\psi$$

then π is generic.

This seems much harder to prove for higher n. Note that if k is replaced by $\mathbb{C}$, a similar statement is true (this follows from [**5**]). For different formulations of the problem, see [**CHT**, § 2.5] as well as [**34**].

Assuming the Conjecture, rather delicate arguments allow one to extend (6.2) to arbitrary level. If we now consider a representation R as in § 3, the assumptions made on $\bar{R}$ imply that $\bar{R}$ will be associated to f (for a suitable choice of G), and the assumptions on R ensure that a form f' giving rise to R will occur in the support of $\mathcal{T}_K$ for a suitable ramification group K. This implies Theorem C. (The reason for conditions a - c imposed to R should now be clear; the "bigness" condition in (d) is imposed in order to make the computation of H^1_S possible, with the correct expression in order to compare with the automorphic side; (e) is also needed for this, as well as (f).)

7. Taylor's final contribution

We will now sketch, very informally, Taylor's new idea which allows him to dispense with Conjecture I of § 6. The crucial fact, here, is that, as should be clear from § 1 and 5, we are aiming at proving potential modularity rather than modularity proper. In particular, in the arguments of § 5, a potential version of Theorem C (of § 3) will suffice.

Recall that our datum is a representation $\bar{R}$ of G_F – in order to keep things simple, F is imaginary quadratic as in § 6. We assume $\bar{R}$

automorphic; we are given a lifting R of $\bar{R}$ over an ℓ-adic field, and R, $\bar{R}$ satisfy the usual conditions (say, those of Theorem C). We are to prove that R is automorphic.

Let $F'_+/\mathbb{Q}$ be a solvable, totally real extension, and $F' = F'_+F$. Then $\bar{R}' = \bar{R}|_{G_{F'}}$ is still automorphic since solvable base change is available. (We assume $\bar{R}'$ irreducible.) As the local Galois groups are solvable, we can then impose suitable local conditions to $R' = R|_{G_{F'}}$ and $\bar{R}'$. (This idea is due to Skinner and Wiles [**29**].) We will in fact assume:

(7.1)

Except at some prime v_0 where we retain a "Steinberg condition", $\bar{R}|_{G_{F',v}}$ is semi-stable for all finite primes v of F' not dividing ℓ.

We now simply denote by F the new (CM) field F'. Recall the meaning of (7.1). If $G_v = \mathrm{Gal}(\bar{F}_v/F_v)$, we have the inertia subgroup I_v given by the exact sequence

(7.2)
$$\begin{tikzcd} 1 \arrow[r] & I_v \arrow[r] & G_v \arrow[r] & \mathrm{Gal}(\bar{k}_v/k_v) \arrow[r] \arrow[d, phantom, "\|\wr"] & 1\ . \\ & & & \hat{\mathbb{Z}} & \end{tikzcd}$$

Moreover there is a surjective homomorphism

$$t_\ell : I_v \to \mathbb{Z}_\ell.$$

The assumption is that I_v acts through t_ℓ: there is a unipotent matrix $U \in \mathrm{GL}(V)$ where V is the space of R, such that for $\sigma \in \mathrm{GL}(V)$ where V is the space of R, such that for $\sigma \in I_v$:

$$R(\sigma) = U^{t_\ell(\sigma)}.$$

We now assume that $(R, \bar{R})$ satisfies this condition. Let **Ram** be the set of ramified primes for $\bar{R}$ (not dividing ℓ). We may also assume, since for $v \in$ **Ram** $\bar{R}(G_v)$ is finite and solvable:

(7.3)
$$v \in \mathbf{Ram} \Rightarrow \bar{r}(G_v) = 1.$$

(Taylor also assumes that the primes in **Ram** are above primes of F_+ split in F; in our description it is not clear that one could assume this, but recall that for Sato-Tate the choice of quadratic imaginary field was arbitrary....)

Now, as in § 6, we can introduce a deformation algebra $\mathcal{R} = \mathcal{R}_{\mathcal{S}}(\bar{R})$ [**T**, § 2]. Here as usual $\mathcal{S}$ denotes a set of local conditions on a lift R of $\bar{R}$. They include:

- R crystalline (at primes dividing ℓ)
- and condition on its Hodge-Tate weights (at these primes)
- R semi-stable (at $v \neq v_0$)
- R "Steinberg" (at v_0, or some larger set of primes).

The precise conditions are complicated, see [**T**].

On the other hand, $\bar{R}$ was supposed to be automorphic, thus associated to an automorphic form $\bar{f}$ on some unitary group of the type described in § 6 (of course, now over F_+). Again, we can define a natural Hecke algebra $\mathcal{T}$, in a level associated to $\bar{f}$, and localized at the maximal ideal $\mathfrak{m}$ defined by $\bar{f}$.

Denote by $\mathcal{R}_{\mathrm{red}}$ the **reduced** algebra, $\mathcal{R}$ divided by its nilpotent radical. Also recall that $\mathcal{T}$ is reduced. (Extension of scalars from the ℓ-adic field L to $\mathbb{C}$ shows that it has no nilpotents, because this is true for the "classical" Hecke algebras, over $\mathbb{C}$.) As in § 6 we get a map $\mathcal{R} \to \mathcal{T}$, so in fact, $\mathcal{R}_{\mathrm{red}} \to \mathcal{T}$.

THEOREM E (Taylor). *Under suitable conditions or $\bar{r}$, the natural map*

$$\mathcal{R}_{\mathrm{red}} \to \mathcal{T} \textit{is an isomorphism.}$$

The inverted commas are here, not because the theorem is doubtful, but because the statement we give is obviously vague. It was impossible to state fully here the precise conditions imposed by Taylor. The point, of course, is that they can be imposed for the representations involved in the proof of "Sato-Tate".

Finally, the proof of Theorem E proceeds by comparing the deformation problem $\mathcal{R}$ and another deformation problem (or deformation algebra) $\mathcal{R}'$ where the local conditions, at the primes $v \in \mathbf{Ram}$, are defined as follows. Fix such a prime v. Recall that $\bar{R}(G_v) = \{1\}$, while G_v acts through t_ℓ; in particular, for $\sigma \in I_v$, the characteristic polynomial of $R(\sigma)$ (for a lifting R associated to $\mathcal{R}$) is

$$(X-1)^n. \tag{7.3}$$

Let us fix, on the other hand, n distinct characters

$$\chi_i : I_v \to 1 + \lambda\mathcal{O} \subset \mathcal{O}^\times$$

where $(\mathcal{O}, \lambda)$ is a finite extension of $\mathbb{Z}_\ell$ with χ_i of order ℓ (so $\ell \geqslant n \cdots$) and consider representations at v for which the characteristic polynomial of $R(\sigma)$ is

$$\prod_i (X - \chi_i(\sigma)). \tag{7.4}$$

Note that **over** $k/\mathbb{F}_\ell$ the two deformation problems coincide. The new problem $\mathcal{R}'$ yields a sort of desingularization of $\mathcal{R}$, which implies Theorem E by a variant of the Taylor-Wiles method. Keeping track of the relation between $\mathcal{R}$ and $\mathcal{R}'$ involves a **relative** deformation theory, which relies onthe schemes describing the representations of the local Galois groups verifying (7.3) and (7.4) respectively. These are schemes of matrices over the ring of integers $\mathcal{O}$ of the ℓ-adic field of coefficients, and a rather precise study of them is done in [**T**, § 1]. Although Taylor does not use his results, this use of the local deformation spaces is due to Kisin [**18**]. We refer the reader to Harris's notes [**11**] for a fuller description of Taylor's argument.

References

[CHT] L. Clozel, M. Harris, R. Taylor, *Automorphy for some ℓ-adic lifts of automorphic mod ℓ Galois representations*, preprint.

[HSBT] M. Harris, N. Shepherd-Barron, R. Taylor, *Ihara's lemma and potential automorphy*, preprint.

[T] R. Taylor, *Automorphy for some ℓ-adic lifts of automorphic mod ℓ Galois representations* II, preprint.

[1] J. Arthur, L. Clozel, *Simple algebras, base change, and the advanced theory of the trace formula*, Ann. Math. Studies; Princeton U. Press, 1989.

[2] A. Borel, N. Wallach, *Continuous cohomology, discrete subgroups, and representations of reductive groups*, Ann. Math. Studies; Princeton U. Press, 1980.

[3] C. Breuil, B. Conrad, F. Diamond, R. Taylor, *On the modularity of elliptic curves over $\mathbb{Q}$: wild* 3*-adic exercises*, Journal of the AMS **14** (2001) 843–939.

[4] P. Cartier, *Representations of p-adic groups: a survey*, in 'Proc. Symp. Pure Math.', **33** (Part. 1), AMS, Providence (1979) 111–155.

[5] L. Clozel, *Représentations galoisiennes associées aux représentations automorphes autoduales de $GL(n)$*, Publ. I.H.E.S. **73** (1991) 97–145.

[6] H. Darmon, F. Diamond, R. Taylor, *Fermat's last theorem*, in 'Current developments in Mathematics', International Press, Cambridge, MA, 1994.

[7] P. Deligne, in SGA 7/2, *Groupes de monodromie en géométrie algébrique*, Springer LN 340, Berlin, 1973.

[8] P. Deligne, J.S. Milne, A. Ogus, K-y. Shih, *Hodge cycles, motives, and Shimura varieties*, Springer LN 900, Berlin, 1982.

[9] K. Fujiwara, *Galois deformations and arithmetic geometry of Shimura varieties*, in 'Proc. I.C.M.', Madrid, 2006.

[10] B. Gross, *Arithmetic on elliptic curves with complex multiplication*, Springer LN 776, 1980.

[11] M. Harris, *The Sato-Tate conjecture: introduction to the proof*, preprint.

[12] M. Harris, *Potential automorphy of odd-dimensional symmetric powers of elliptic curves, and applications*, preprint.

[13] M. Harris, R. Taylor, *The geometry and cohomology of some simple Shimura varieties*, Ann. Math. Studies, Princeton U. Press, 2001.

[14] E. Hrushovski, A. Pillay, *Definable subgroups of algebraic groups over finite fields*, J. Reine Angew. Math. **462** (1995) 69–91.

[15] L. Illusie, *Crystalline cohomology*, in 'Motives', U. Jannsen, S. Keliman, J.-P. Serre eds., Proc. Symp. Pure Math. **55** (Part 1), AMS, Providence (1944) 43–70.

[16] H. Jacquet, J. Shalika, *A non-vanishing theorem for zeta functions of GL_n*, Inv. Math. **38** (1976) 1–16.

[17] C. Khare, J.-P. Wintenberger, *Serre's modularity conjecture*, (I, II), preprints.

[18] M. Kisin, *Moduli of finite flat group schemes and modularity*, preprint.

[19] R.P. Langlands, *Base change for $GL(2)$*, Ann. Math. Studies, Princeton U. Press, 1980.

[20] C.R. Matthews, L.N. Vaserstein, B. Wesfeiler, *Congruence properties of Zariski dense subgroups* I, Proc. London Math. Soc. **48** (1984) 514–532.

[21] L. Moret-Bailly, *Groupes de Picard et problèmes de Skolem* II, Ann. Scient. Ecole Norm. Sup. **22** (1980) 181–194.

[22] M. Nori, *On subgroups of $GL_n(\mathbb{F}_p)$*, Inv. Math. **88** (1987) 257–275.

[23] K. Ribet, *Congruence relations between modular forms*, Proc. I.C.M. 1982, PWN, Warsaw 1984.

[24] J.-P. Serre, *Représentations linéaires des groupes finis*, Hermann, Paris, 1978.

[25] J.-P. Serre, *Abelian ℓ-adic representations and elliptic curves*, Benjamin, 1968.

[26] J.-P. Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Inv. Math. **15** (1972), 259-331 [=Œuvres, vol. III, Springer (1986) 1–73].

[27] F. Shahidi, *On non-vanishing of L-functions*, Bulletin AMS **2** (1980) 462–464.

[28] J. Silverman, *The arithmetic of elliptic curves*, Springer, Berlin, 1986.

[29] C. Skinner, A. Wiles, *Base change and a problem of Serre*, Duke Math. J. **107** (2001) 15–25.

[30] J. Tate, *Number theoretic background*, in Automorphic forms, representations and L-functions, Proc. Symp. Pure Math. **33** (Part. 2), AMS, Providence (1979) 3–26.

[31] J. Tate, *Algebraic cycles and poles of zeta functions*, in Arithmetic Algebraic Geometry, Harper and Row, New York (1965) 93–110.

[32] R. Taylor, *Remarks on a conjecture of Fontaine and Mazur*, J. Inst. Math. Jussieu **1** (2002) 1–19.

[33] R. Taylor, *On the meromorphic continuation of degree two L-functions*, preprint.

[34] R. Taylor, *Galois representations*, Proc. ICM (Beijing, 2002), vol. I, 449–474 (complete version, Ann. Fac. Sc. Toulouse **13** (2004) 73–119).

[35] R. Taylor, A. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. Math. **141** (1995) 553–572.

[36] M.-F. Vignéras, *Représentations modulaires d'un groupe p-adique avec $\ell \neq p$*, Progress in Math, Birkhaüser, 1996.

[37] A. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. Math. **141** (1995) 443–551.

Département de Mathématiques, Bètiment 425, Faculté des Sciences d'Orsay, Université Paris-Sud 11, F-91405 Orsay Cedex
E-mail address: laurent.clozel@math.u-psud.fr