THE SATO-TATE CONJECTURE: INTRODUCTION TO THE PROOF

Michael Harris

Introduction

It would be quite simple and reasonably painless to write an expository account of the recent proof of the Sato-Tate conjecture for (a large class of) elliptic curves stressing the arithmetic and geometric aspects of the proof and treating the background from automorphic forms like the mad relative in the attic, about whom the less said, the better. Such an approach is in any case inevitable in a one-hour lecture, such as my talk at the Franco-Asian Summer School of July 2006, which serves as the pretext for this article. It would moreover be faithful to the three papers that together comprise the proof ([CHT], [HST], [T]), which contain little if any novel information about automorphic forms.

This report nevertheless begins with an introduction to Langlands' reciprocity conjectures, and their arithmetic variants, in the situation most relevant to the proof of the Sato-Tate conjecture. The reader should keep in mind that the goal is to relate L-functions of Galois representations to automorphic representations on $GL(n)$, as I recall in §1, but that this relation is mediated for technical reasons by the theory of automorphic representations of certain kinds of unitary groups. Different unitary groups are relevant at different stages of the argument. I have attempted to sort this out in §2. The actual steps in the proof of the Sato-Tate conjecture are presented in §§4 and 5, following the introduction of the relevant Hecke algebras in §3.

It remains to thank my coauthors, Laurent Clozel, Nick Shepherd-Barron, and Richard Taylor, and to thank Richard Taylor separately for sending me the notes of his MSRI lectures in advance of the MSRI "Hot Topics" workshop on modularity, held in October-November 2006, in which some of this material was presented. I also thank Jean-Benoît Bost and Jean-Marc Fontaine for inviting me to speak at the Franco-Asian Summer School, and to submit the present article to the proceedings.

1. On global Langlands correspondences for $GL(n)$

Let F be a number field. One version of the global Langlands correspondence, one of particular interest to number theorists, is the conjectural dictionary:

$$\left\{\begin{array}{l}\text{(Certain) cuspidal}\\ \text{automorphic}\\ \text{representations}\\ \Pi \text{ of } GL(n,\mathbf{A}_F)\end{array}\right\} \leftrightarrow \left\{\begin{array}{l}\text{Compatible systems}\\ \rho_{\Pi,\ell} \text{ of } \ell\text{-adic}\\ \text{representations of}\\ \Gamma_F = Gal(\overline{\mathbb{Q}}/F)\\ \text{of dimension } n\end{array}\right\} \tag{1.1}$$

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

To say that the $\rho_\ell = \rho_{\Pi,\ell}$ form a *Compatible system:* is to say that all the ρ_ℓ yield the same (Artin, Hasse, Weil) L-function $L(s,\rho)$. This is a strong version of the assertion that $L(s,\rho)$ has an analytic continuation and functional equation. Let $n = \dim \rho$, so the general Euler factor of $L(s,\rho)$ is of degree n. The form of the general Euler factor of $L(s,\Pi)$ at unramified places is recalled in §3, below.

The word "certain" in the above dictionary is crucial. Not all cuspidal automorphic representations of $GL(n)$ are of Galois type, i.e., conjecturally associated to Galois representations. Maass forms for $GL(2,\mathbb{Q})$ are the most obvious example. They include, for example, cuspidal functions on $SL(2,\mathbb{Z})\backslash\mathfrak{H}$ which are eigenfunctions for all Hecke operators and for the hyperbolic Laplacian. The collection of such forms is large (Selberg) but, even allowing the level to increase, practically none of them are supposed to be of Galois type. The Π of Galois type were identified by Clozel in his article for the Ann Arbor conference; he called them "algebraic" and characterized them as those for which the archimedean component Π_∞ has infinitesimal character (character of the center of the enveloping algebra) corresponding to an element of the weight lattice of the Lie algebra $GL(n,F_\infty)$. Call this the *archimedean weight* of Π_∞; it is well-defined modulo a twisted action of the Weyl group (product of permutation groups for the different archimedean places of F). This is an integrality condition and can naturally be interpreted in terms of p-adic Hodge theory: in the complex of the Fontaine-Mazur conjectures, one expects each ρ_ℓ in a compatible system to be *geometric* in the sense of Fontaine-Mazur – i.e., of de Rham type at each prime dividing ℓ. In particular, they are Hodge-Tate, and the dictionary predicts the Hodge-Tate weights in terms of the infinitesimal character of Π_∞.

The construction of Galois representations for general automorphic Π of Galois type remains completely open. All known methods only apply when Π_∞ is not only algebraic but *cohomological.* This means that the archimedean weight of Π_∞ is a *dominant* weight, hence is the highest weight of the dual of an irreducible finite-dimensional representation $L(\Pi_\infty)$ of $GL(n,F_\infty)$. The precise condition is expressed in terms of relative Lie algebra cohomology:

$$H^\bullet(\mathfrak{gl}(n,F_\infty),K_\infty;\Pi_\infty\otimes L(\Pi_\infty)^\vee)\neq 0. \tag{1.1.1}$$

Here K_∞ is a chosen maximal compact subgroup of $GL(n,F_\infty)$ (in practice it is multiplied by the center of $GL(n,F_\infty)$); one has to make such a choice in order to define automorphic forms in the first place.

Given additional restrictions on F, one can construct Galois representations. Let F be either totally real or a CM field, and in either case let $F^+\subset F$ be its maximal totally real subfield, so that $[F:F^+]\leq 2$. Let $c\in Gal(F/F^+)$ be complex conjugation; by transport of structure it acts on automorphic representations of $GL(n,F)$.

Theorem 1.2 ([C], [Ko], [HT], [TY]). *There is an arrow from left to right $\Pi\mapsto\{\rho_{\Pi,\lambda}\}$, as λ runs through non-archimedean completions of a certain number* *field $E(\Pi)$ when F is totally real or a CM field, under the following hypotheses:*

$$\left\{\begin{array}{l}\textit{(1) The factor } \Pi_\infty \\ \textit{is cohomological} \\ \textit{(2) } \Pi\circ c \cong \Pi^\vee \\ \textit{(3) } \exists v_0, \Pi_{v_0} \\ \quad \textit{discrete series}\end{array}\right\} \Rightarrow \left\{\begin{array}{l}\textit{(a) } \rho = \rho_{\Pi,\ell} \textit{ geometric,} \\ \textit{HT regular} \\ \textit{(b) } \rho\otimes\rho\circ c \ \to \mathbb{Q}_\ell(1-n) \\ \textit{(c) local condition} \\ \quad \textit{at } v_0\end{array}\right\}$$

This correspondence has the following properties:

(i) *For any finite place v prime to the residue characteristic ℓ of λ,*

$$\rho_{\Pi,\lambda}\mid_{G_v} = \mathcal{L}(\Pi_v \otimes |\bullet|_v^{\frac{1-n}{2}}).$$

Here G_v is a decomposition group at v and $\mathcal{L}$ is the **normalized** *local Langlands correspondence;*

(ii) *The representation $\rho_{\Pi,\lambda}\mid_{G_v}$ is potentially semistable for any v dividing ℓ and the Hodge-Tate numbers at v are explicitly determined by the archimedean weight of Π_∞.*

The local Langlands correspondence is given the unitary normalization. This means that we need to introduce twists by half-powers of the norm, so that the functional equations always exchange values at s and $1-s$.

The term "geometric" (Fontaine-Mazur) means that each $\rho_{\Pi,\lambda}$ looks like the representation on (a piece of) the middle-dimensional cohomology of a smooth d-dimensional projective variety over F; in fact $d = n-1$.[1] The condition "HT regular" (Hodge-Tate) corresponds to the hypothesis that this piece of the cohomology has a Hodge structure pure of weight d with Hodge numbers

$$\forall p \ h^{p,d-p} \leq 1. \tag{1.3}$$

Fontaine's theory assigns Hodge numbers to geometric Galois representations (Fontaine-Messing, Faltings, Tsuji. To facilitate comparison with the archimedean theory, I will discount all conventions and index Hodge-Tate weights by the usual Hodge numbers so that that the dimension of the Hodge-Tate component corresponding to $H^q(\Omega^p)$, and thus to the qth power of the cyclotomic character, is denoted $h^{p,q}$.

For the local condition (c), we can take the condition that the representation of the decomposition group at v_0 is indecomposable as long as v_0 is prime to the residue characteristic of λ, or equivalently that this representation of the decomposition group at v_0 corresponds to a discrete series representation of $GL(n, F_{v_0})$. The conditions on both sides of the diagram match: (1) $\leftrightarrow$ (a), (2) $\leftrightarrow$ (b), (3) $\leftrightarrow$ (c).

When $F = F^+$, condition (2) just means that Π is self-dual. In the unitary normalization, this includes the case of all Hilbert modular forms with real Nebentypus character. Thus the above theorem does not include the most general modular Galois representations even in the elliptic modular case. There are various ways to

[1] Taylor has recently proposed that "geometric" be replaced by "algebraic," so that the same term would be used on both sides of the dictionary (1.1). In this article I have chosen to conform to the published literature, but the reader should be aware that the terminology may soon change.

weaken the condition on the central character, which is imposed by the technique of base change to unitary groups, but none of them seems optimal.

The Galois representations $\rho_{\Pi,\lambda}$ are constructed in the cohomology of Shimura varieties, in fact of Shimura varieties attached to unitary groups. Say F is really a CM field and V is a hermitian space of dimension n over F. We can associate to the unitary group $G(V)$ of V a Shimura variety. In general this can be done in more than one way, and the Shimura variety is really associated to the group of unitary similitudes of V but our interest here is just to present an approximate theorem, so we denote the Shimura variety $Sh(G(V))$. Any finite-dimensional representation L of $GL(n, F_\infty)$ satisfying the analogue of condition (2) of Theorem 1.2:

$$L \circ c \xrightarrow{\sim} L^\vee \tag{1.4}$$

defines an ℓ-adic local system $\tilde{L}_\ell$ on $Sh(G(V))$ by a standard construction.

Theorem 1.5 [HL]. *Under the hypotheses of Theorem 1.2, suppose either n or $[F^+ : \mathbb{Q}]$ is odd or $n \equiv 2 \pmod 4$. Then ρ_λ can be realized (up to dualizing and twisting by an abelian character) in $H^{n-1}(Sh(G(V)), L(\Pi_\infty)^\vee_\ell)$ (cohomology with twisted coefficients), where V is a hermitian space of signature $(n-1,1)$ at one real place τ of F^+, definite at all other real places, and quasi-split at all finite places .*

The proof is by stable base change. As we will see in §2, the parity hypothesis implies that there exists a hermitian space V with the indicated local properties at all places. Hypotheses (1) and (2) of Theorem 1.2 should then suffice to imply that the automorphic representation Π gives rise (by functorial descent, the inverse of base change) to a packet of automorphic representations of $G(V)$ that contribute to $H^{n-1}(Sh(G(V)), L(\Pi_\infty)^\vee_\ell)$. Working at finite level, this amounts to saying that the system of eigenvalues of Hecke operators attached to the finite part Π_f of Π can be realized on a component, say $H[\Pi_f]$, of the above ℓ-adic cohomology group. Now $Gal(\overline{\mathbb{Q}}/F)$ commutes with all Hecke operators, hence acts on $H[\Pi_f]$, and the relation between the Galois action and the eigenvalues of Hecke operators is determined by applying the method of Langlands and Kottwitz ("counting points"). Actually, this method is applied, in a variant adapted to bad reduction, not to $Sh(G(V))$ but to the Shimura variety attached to an inner twist G' of $G(V)$, in [HT]. The result of [HL] is proved by applying stable base change twice, once to descend from $GL(n)$ to $G(V)$, and then a second time to compare this descent with the functorial transfer (*generalized Jacquet-Langlands correspondence*) between automorphic forms on $G(V)$ and G', to reduce to the earlier results of [HT].

Stable base change is based entirely on the theory of the stable trace formula, which is central in the proofs of Theorems 1.2 and 1.5. The trace formula will not be discussed at all in these notes.

1.6. Reciprocity and examples.

The arrow in diagram (1.1) is double-headed. This can be made into a precise conjecture.

Reciprocity Conjecture (Langlands, Fontaine-Mazur). *The arrow in Theorem 1.2 is an equivalence; i.e., every absolutely irreducible ρ satisfying (a), (b), (c) is automorphic of the indicated type.*

Implicit in this conjecture is the preliminary conjecture that the representations obtained in Theorem 1.2 are all absolutely irreducible. This is only known for $n \leq 3$

(in some cases for $n = 4$) and constitutes one of the main open questions in the field. By Chebotarev density, ρ can be reconstructed from $L(s, \rho)$, provided it is semisimple, which we will always assume. In most cases we will only consider Π for which it is known *a priori* that ρ_ℓ is irreducible for some ℓ – indeed, that ρ_ℓ is even absolutely irreducible modulo ℓ.

Conditions (c) and (3) should be irrelevant and remain as a reminder that the stable trace formula for unitary groups is not yet available. Progress has been made recently toward eliminating these conditions, however, and in these notes I will have little to say about the relevance of condition (c).

I can think of two classes of examples that arise elsewhere than in the theory of automorphic forms.

1.7. Let E be an elliptic curve over F^+, and let $\rho_{E,\ell} : Gal(\overline{\mathbb{Q}}/F^+) \to GL(2, \mathbb{Q}_\ell)$ denote the representation on $H^1(E_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)$, i.e. the dual of the ℓ-adic Tate module. Suppose it is known that E is automorphic: that there exists a cuspidal automorphic representation Π_E of $GL(2, F^+)$ such that (up to normalization) $L(s, \Pi_E) = L(s, E)$ as Euler products. That this is the case when $F^+ = \mathbb{Q}$ is the theorem proved in the series of papers initiated by Wiles and Taylor-Wiles and completed by Breuil, Conrad, Diamond, and Taylor. For $n \geq 1$ let

$$\rho^n_{E,\ell} = Sym^{n-1}\rho_{E,\ell} : Gal(\overline{\mathbb{Q}}/F^+) \to GL(n, \mathbb{Q}_\ell).$$

The corresponding non-zero Hodge numbers are $h^{i,n-1-i} = 1, i = 0, \ldots, n-1$. If E has no complex multiplication, then $\rho^n_{E,\ell}$ is irreducible by a theorem of Serre, for all n. It is then obvious that $\rho^n_{E,\ell}$ satisfies conditions (a) and (b) of Theorem 1.2. Moreover, $\rho^n_{E,\ell}$ is locally indecomposable at v_0 for all n (condition (c)) if and only if E has multiplicative reduction at v_0. Thus the symmetric powers of elliptic curves over totally real fields with non-integral j-invariants all provide examples of Galois representations on the right-hand side of the diagram in Theorem 1.2. The main result of the three papers I discuss here is that the *even-dimensional* ρ^n_E for such E are *potentially automorphic* : they become automorphic over appropriate totally real Galois extensions of F^+. This suffices for the proof of the Sato-Tate conjecture.

Suppose f is an elliptic modular form of weight $k > 2$, or more generally a Hilbert modular form whose archimedean components are in the discrete series. Let Π be the corresponding automorphic representation of $GL(2, F^+)$. Let $\rho_{f,\lambda}$ be the corresponding two-dimensional λ-adic representation of $Gal(\overline{\mathbb{Q}}/F^+)$. Assume f is not obtained by automorphic induction from a Hecke character of a quadratic extension of F^+. Then the $Sym^{n-1}(\rho_{f,\lambda})$ satisfy conditions (a) and (b) of 1.2 (provided the Nebentypus character is real), and satisfy (c) as well if there is some finite place v_0 such that Π_{v_0} is a Steinberg representation. We have thus constructed another class of examples for the Reciprocity Conjecture. However, it cannot honestly be claimed that these examples arise outside the theory of automorphic forms.[2]

[2]Now that the Serre conjecture on two-dimensional modular representations of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ has been proved in nearly all cases (Khare, Wintenberger), and the Fontaine-Mazur conjecture for two-dimensional representations is not far behind (Kisin, Emerton), one could reformulate this class of examples in terms of symmetric powers of general two-dimensional representations of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ that are geometric in the Fontaine-Mazur sense).

1.8 A family of Calabi-Yau hypersurfaces. Let $Y = \mathbb{P}^1 - \{\mu^{n+1}, \infty\}$. For $t \in Y(\mathbb{C})$, consider $X_t \subset \mathbb{P}^n$ defined by

$$F_{t,n}(X_0, \ldots, X_n) = \sum_{i=0}^{n} X_i^{n+1} - (n+1)tX_0 \cdot X_1 \cdot \ldots \cdot X_n.$$

$n \geq 2$: the *Dwork pencil* of Calabi-Yau hypersurfaces familiar from mirror symmetry. It is known that X_t smooth for $t \in Y$, dim $n-1$ (if $n = 2$, X_t is an elliptic curve).

$$\text{(P)} \qquad H^{n-1}_{\acute{e}t}(X_{t,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \otimes H^{n-1}_{\acute{e}t}(X_{t,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell) \to \mathbb{Q}_\ell(1-n)$$

(Poincaré duality) preserved by $Gal(\overline{\mathbb{Q}}/F^+)$ if $t \in F^+$. The finite group

$$H = \{(\zeta_0, \ldots, \zeta_n) \in \mu_{n+1}^{n+1} \mid \prod_j \zeta_j = 1\}/\Delta(\mu_{n+1}),$$

acts on the family (multiply each X_i by ζ_i). Then

$$V_{t,\ell} = H^{n-1}_{\acute{e}t}(X_{t,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)^H \subset H^{n-1}_{\acute{e}t}(X_{t,\overline{\mathbb{Q}}}, \mathbb{Q}_\ell)$$

is an n-dimensional, polarized (via P) representation $\rho_{t,\ell}$ of $Gal(\overline{\mathbb{Q}}/F^+)$. Moreover, $\rho_{t,\ell}$ is HT regular:

$$h^{i,n-1-i} = 1, 0 \leq i \leq n-1, \ = 0 \text{ otherwise}.$$

The complex Hodge numbers for the corresponding de Rham cohomology $V_{t,DR}$, and the Gauss-Manin connection for the family of $V_{t,DR}$ have been calculated in the physics literature, and again in [HST], and are central in the discussion there; one obtains the Hodge-Tate weights by applying p-adic comparison theorems (here $p = \ell \ldots$).

As in the previous example, the representations $\rho_{t,\ell}$ satisfy conditions (a) and (b) of Theorem 1.2. It is shown in [HST] that $\rho_{t,\ell}$ satisfies condition (c) provided t is not an integer in F^+. This proof is by comparison with the complex calculation of monodromy, and in particular makes use of the fact that the monodromy around the point ∞ is unipotent and of maximal order of unipotence; letting γ denote a in $\mathbb{P}^1(\mathbb{C})$ around ∞, the minimal polynomial of the monodromy representation of γ is $(T-1)^n$.

In particular, the Reciprocity Conjecture applies to $\{\rho_{t,\ell}\}$ provided $t \in F^+ - \mathcal{O}_F$.

I conclude this section with an observation that is frequently invoked in proofs of modularity.

Proposition 1.9. *Let ρ be an n-dimensional ℓ-adic representation of $Gal(\overline{\mathbb{Q}}/F)$. Suppose F'/F is a solvable Galois extension (also CM or totally real) and Π' is a cuspidal automorphic representation of $GL(n, F')$ that is associated to $\rho' = \rho \mid_{Gal(\overline{\mathbb{Q}}/F')}$ under the arrow of Theorem 1.2. In other words, suppose ρ becomes automorphic, of the type considered in Theorem 1.2, after restriction to F'. Then ρ is already automorphic over F, i.e. there exists a cuspidal automorphic representation Π of $GL(n, F)$ such that $L(s, \rho) = L(s, \Pi)$.*

This follows from the Arthur-Clozel theory of base change: because ρ' extends to a Galois representation of F, Π' is invariant under $Gal(F'/F)$, and hence can be descended to $GL(n,F)$. The reader should be warned that not every Galois-invariant representation can be so descended, and that incorrect proofs of the Artin conjecture for solvable Galois representations, and of some cases of non-Galois base change, can be constructed by an incautious use of descent. However, our hypotheses guarantee that the descents for intermediate cyclic extensions are well-behaved.

1.10. The polarization condition in the L-group formalism.

The techniques for proving modularity do not apply directly to the representation ρ as in the statement of the Reciprocity Conjecture but rather to a version that takes account of the polarization condition (b). Let

$$\mathcal{G}_n = (GL(n) \times GL(1)) \rtimes Gal(F/F^+)$$

with the element $c \in Gal(F/F^+)$ acting by

$$c(g,\mu)c^{-1} = (\mu \cdot {}^t g^{-1}, \mu). \tag{1.10.1}$$

There is a homomorphism

$$\nu : CalG_n \to GL(1); \quad \nu((g,\mu)) = \mu,\ \nu(c) = -1.$$

This is similar to but not quite the same as the L-group of $GU(n)$, in which an extra inner automorphism is added to the action of c in order to preserve the splitting data (*épinglage*). Condition (b) can be rephrased:

Lemma 1.10.2. *(i) Let k be a field and let $\rho : Gal(\overline{\mathbb{Q}}/F) \to GL(n,k)$ be a representation satisfying the following version of (b):*

$$\rho \circ c \xrightarrow{\sim} \rho^\vee \otimes \omega_k^{1-n}, \tag{1.10.3}$$

where ω_k is the cyclotomic character over k. Suppose ρ is absolutely irreducible. Then there is an extension of ρ to a homomorphism

$$r : Gal(\overline{\mathbb{Q}}/F^+) \to \mathcal{G}_n(k) \tag{1.10.4}$$

such that $\nu \circ r \mid_{Gal(\overline{\mathbb{Q}}/F)} = \omega_k^{1-n}$ and $r(c)$ belongs to the non-trivial coset of $GL(n,k) \times GL(1,k)$ in $\mathcal{G}_n(k)$. The set of such extensions, up to $GL(n,k)$-conjugacy, is in bijection with $k^\times/k^{\times,2}$.

(ii) Suppose $\mathcal{O}$ is an ℓ-adic integer ring with maximal ideal $\mathfrak{m}$. Let $\rho : Gal(\overline{\mathbb{Q}}/F) \to GL(n,\mathcal{O})$ be a representation satisfying (1.10.3) and suppose $\bar\rho = \rho \pmod{\mathfrak{m}}$ is absolutely irreducible. Then ρ admits an extension $r : Gal(\overline{\mathbb{Q}}/F^+) \to \mathcal{G}_n(\mathcal{O})$ with $\nu \circ r$ and $r(c)$ as above.

2. Unitary groups and base change

Henceforward, we always assume F totally imaginary, so $[F : F^+] = 2$ and the Galois conjugation c is non-trivial. It's also convenient henceforward to assume n to be even, since this is the only case to which our potential modularity methods apply, although the modular lifting theorems work as well for odd n. We let $d = [F^+ : \mathbb{Q}]$.

As indicated in §1, the construction of the arrow from left to right:

$$\left\{\begin{array}{l}\text{automorphic}\\ \text{representations}\\ \text{of } GL(n)\end{array}\right\} \to \left\{\begin{array}{l}\text{Compatible systems of}\\ n\text{-dimensional } \ell\text{-adic}\\ \text{representations}\end{array}\right\} \tag{2.1}$$

proceeds by replacing the automorphic representation Π of $GL(n)$ by a packet $\{\pi\}$ of representations of some unitary group $G(V)$ that contribute to the cohomology in middle degree $n-1$ of a Shimura variety we denoted $Sh(G(V))$. There is no known way to construct the packet $\{\pi\}$ directly, starting with Π. The existence of $\{\pi\}$, which we can call the $G(V)$-avatar of Π, is deduced from the comparison of a (stable) twisted trace formula attached to $GL(n)$ to the stable trace formula for $G(V)$, or more generally to the stable trace formula for an inner twist G' of $G(V)$ that also defines a Shimura variety of dimension $n-1$. The automorphic representation Π factors over the places of F:

$$\Pi = \otimes'_v \Pi_v$$

where almost all Π_v are unramified, and in particular are determined by the eigenvalues of the corresponding Hecke operators. There are n basic Hecke operators for $GL(n)$, say $T_{1,v}, \dots, T_{n,v}$, whose eigenvalues, suitably normalized, define the coefficients of the local factor $L(s, \Pi_v)$ at v. For the purposes of these notes we hardly need to know more about them than that. Compatibility with the local Langlands correspondence implies that $\rho_{\Pi,\ell}$ is unramified at any prime v, $v \not\div \ell$, such that Π_v is unramified. The identity $L(s, \Pi_v) = L_v(s, \rho_\Pi)$ (a special case of (i) of Theorem 1.2) then implies that the eigenvalues of the $T_{i,v}$ on Π_v determine the traces of the Frobenius at v under $\rho_{\Pi,\ell}$. As for $GL(2)$, the compatible system of (semisimple) ℓ-adic representations attached to Π is thus determined up to isomorphism by the systems of eigenvalues of these Hecke operators, and indeed by the eigenvalues of the $T_{i,v}$ for a subset of places v of F of Dirichlet density 1. For example, the $T_{i,v}$ for v split over F^+ suffice.

The $G(V)$ and G' avatars are no longer needed for the construction of the arrow from right to left in (2.1). The Taylor-Wiles method depends on certain cohomology groups being free over finite subgroups of $GL(n, \mathcal{O}_v)$, and this is easiest to arrange when the cohomology is in degree 0 (or in degree 1, as in the Taylor-Wiles paper). Thus we work with a *totally definite* unitary group over F^+, which we call G_0. Thus

$$G_0(F_\infty) = G_0(F \otimes \mathbb{R}) \xrightarrow{\sim} U(n)^d,$$

where $U(n)$ is the compact unitary group. In order for this to be of any use we need Π to descend to a G_0-avatar, i.e. a packet π_0 of automorphic representations (in practice a singleton) of G_0.

How to think about about automorphic representations of unitary groups (avatars).

From the theoretical point of view, an automorphic representation π of a unitary group G over F^+ is supposed to be parametrized possibly up to ambiguity in its L-packet, by a global Langlands parameter, which in the case that concerns us is a homomorphism $\phi = \phi_\pi$ from $Gal(\overline{\mathbb{Q}}/F^+)$ to the L-group of G,

$$^LG = GL(n, \mathcal{K}) \rtimes Gal(F/F^+)$$

where $c \in Gal(F/F^+)$ acts on $GL(n, \mathcal{K})$ by an appropriately normalized outer automorphism. Here $\mathcal{K}$ is an algebraically closed field, often taken to be $\mathbb{C}$; in this Galois formulation we take $\mathcal{K} = \bar{\mathbb{Q}}_\ell$. The homomorphism should commute with projection of both sides to $Gal(F/F^+)$, and so

$$\phi \mid_F : Gal(\overline{\mathbb{Q}}/F)) \to GL(n, \bar{\mathbb{Q}}_\ell)$$

is just our ℓ-adic representation ρ. Recall that ρ was attached to an automorphic representation Π of $GL(n, F)$, and plays the role of the Langlands parameter of Π. Conditions (a) and (b) on ρ imply precisely that ρ extends to a Langlands parameter ϕ with values in ^LG.

In practical terms. cuspidal automorphic representations of $GL(n)$ are the atoms of the theory of automorphic forms, from which all automorphic representations of other groups ultimately derive, as in Langlands' hypothetical Tannakian formalism. They are classified by their L-functions, whose basic properties were established by Jacquet, Shalika, and Piatetski-Shapiro on the one hand, and Shahidi on the other, in the 1970s and 1980s. The arithmetic properties of automorphic representations of $GL(n)$ with $n > 2$ are generally accessible only indirectly, by means of operations involving their avatars on other classical groups, usually unitary groups. By way of analogy one might consider the tradition according to which the Bhagavad-Gita was transmitted to Arjuna not directly by Vishnu but rather by his avatar Krishna.

In order to understand automorphic representations of unitary groups it is best not to think of them as matrix groups but rather as abstract groups related in a certain way to $GL(n)$. This relation is exploited by the trace formula but is irrelevant for the present exposition. An automorphic representation π_0 of G_0 should be understood in terms of its local factors $\pi_{0,v}$ for primes v of F^+. This is how the relation with Π is defined. For example, suppose v splits as $w \cdot w^c$ in F. Then for any hermitian space V/F, $G(V)_v \xrightarrow{\sim} GL(n, F_w) \xrightarrow{\sim} GL(n, F_v^+)$. Condition (2) (polarization) implies that

$$\Pi_w^\vee \xrightarrow{\sim} \Pi_{w^c}.$$

Then the avatar π_0 has the property

$$\pi_{0,v} \xrightarrow{\sim} \Pi_w \tag{2.2}$$

where the isomorphism with Π_w, rather than with Π_{w^c}, depends on some implicit choices. At non-split places it is not so easy to write $\pi_{0,v}$ in terms of Π_w; there is a formula when Π_w is unramified, or at real places, but at other places this is still an open question. We solve this question by reducing to the situation where there are no such places; see below.

Any two unitary groups over F^+, relative to the quadratic extension F/F^+, are inner forms of one another; in particular, their Langlands L-groups are isomorphic. Assuming we have already descended Π to $\{\pi\}$, one can view π_0 as a functorial transfer of $\{\pi\}$ corresponding to the isomorphism ${}^LG(V) \xrightarrow{\sim} {}^LG_0$. Here we invoke the following consequence of Langlands functoriality:

Vague general principle. *(i) The only obstructions to transfer of L-packets between inner forms are local.*

(ii) Let K be a local field, G_1 a reductive group over K, G_2 an inner form of G_1 over K. Assume G_1 is quasi-split. Then there are no local obstructions to transfer of L-packets from G_2 to G_1.

(iii) Let G_1 be a reductive group over $\mathbb{R}$, G_2 an inner form, and suppose G_1 has a discrete series (in which case so does G_2). Then there is no local obstruction to transfer of discrete series L-packets from G_2 to G_1.

In other words, (i) asserts that, if one can transfer π_v to some irreducible admissible representation $\pi_{0,v}$ of $G(V_0)_v$ for every place v, then there is a global π_0. This is what happens with the Jacquet-Langlands correspondence between automorphic representations of $GL(2)$ and automorphic representations of division algebras. Of course, this is a principle, not a theorem, and until there is a completely general and explicit stable trace formula it has to be proved anew in each individual case. For the twisted unitary groups considered in [HT], this was proved in [HT] and in my earlier paper on p-adic uniformization; for untwisted unitary groups there are partial results in [HL] and in Labesse's book [L].

As for (ii), we will not encounter the quasi-split inner form of $U(n)$ over $\mathbb{R}$. Suppose v is a place of F^+ that splits in F; then we have already seen that $G(V)_v$ is a general linear group, which is certainly quasi-split. If v does not split in F, then there are two non-isomorphic unitary groups over F_v^+, the quasi-split one G_v^+, for which the hermitian form is anti-diagonal, and the non-quasi-split one G_v^- (when n is odd every unitary group over a p-adic field is quasi-split). We set $\epsilon(G_v^{\pm}) = \pm 1$; this is the *Hasse invariant* of $G_v^{\pm}$.

Finally, unitary groups over $\mathbb{R}$ all have discrete series, and the $G(*)$-avatars of our cohomological representation Π of $GL(n,F)$ are always of discrete series type at real places, so there is never a local obstruction at ∞. The discrete series is a local L-packet, but for definite groups it contains a single element. If we know $\Pi_\infty = \prod_{w \div \infty} \Pi_w$, where w runs over the (conjugate pairs of) complex prime(s) w of F, then we can determine the corresponding $\pi_{0,\infty}$, as follows. First of all, since Π is cuspidal, Shalika's theorem implies that Π_w is *generic* (has a Whittaker model) for every prime w of F. It is known that, for each irreducible finite-dimensional representation L of

$$Lie(GL(n,F_\infty))_{\mathbb{C}} = \prod_{w \div \infty} \mathfrak{gl}(n,F_w) \times \mathfrak{gl}(n,F_{w^c})$$

satisfying $L^c \xrightarrow{\sim} L^\vee$, there is a unique Π_∞ satisfying (1.1.1) for L

$$H^\bullet(\mathfrak{gl}(n,F_\infty),K_\infty;\Pi_\infty \otimes L^\vee) \neq 0, \tag{2.2.2}$$

such that all factors Π_w are generic. (See Clozel's Ann Arbor article for a discussion of this.) Say

$$L = \prod_v L_w \otimes L_w^\vee$$

where if v is the restriction of w to F^+ then $\mathfrak{gl}(n, F_w)$ acts on the first factor and $\mathfrak{gl}(n, F_{w^c})$ on the second. Then the representation π_0 of $G_0(\mathbb{R}) = \prod_v U(n)$, v running over real places of F^+, is just $\otimes_w L_w$, where there is again an implicit choice of an extension of each real v to a place w of F.

In fact, I have been concealing from you the existence of an important global obstruction, namely the obstruction to the existence of a totally definite G_0 that is quasi-split at all finite primes, hence creates no local obstructions to transfer at finite primes. Actually, we are not yet in a position to work with such a G_0, because the stable trace formula does not yet apply in this case. Instead, we recall the two non-c-conjugate places v_0 and v_1 of Theorem 1.5, and consider only Π (resp. ρ) that satisfy condition (3) (resp. (c)) at *both* v_0 and v_1 (hence at four places in all, counting the complex conjugates). We will soon show that doubling condition (3) entails no loss of generality.

Let B be a central division algebra over F of dimension n^2 whose opposite algebra isomorphic to its c-conjugate. Assume it is split outside a finite set of primes of F, all of which split over F^+, and let $S(B)$ be the corresponding set of primes of F^+. We assume the divisors of $S(B)$ include $v_0, c(v_0), v_1, c(v_1)$ and at each prime of F dividing a prime of $S(B)$ B is a division algebra. The hyothesis on B^{op} implies that the local Hasse invariants of B at the two primes of F above any $v \in S(B)$, so in particular such a B exists. Now let $\ddagger$ be an involution of B of the *second kind*, i.e. $\ddagger$ restricts to complex conjugation c on F. Let G_0 (denoted G in [CHT] and [T]) be the reductive algebraic group over F^+ such that, for any F^+-algebra R,

$$G_0(R) = \{g \in (B \otimes_{F^+} R)^\times \mid g^{\ddagger \otimes 1} \cdot g = 1\}. \tag{2.3}$$

At non-split primes v of F^+ (or F) B_v is a matrix algebra, $\ddagger$ is a c-antilinear automorphism of B_v, and by the classification of such anti-automorphisms $G_0(F_v^+)$ can be identified to the unitary group of some hermitian form on F_v^n. The general classification of unitary groups (outer forms of $GL(n)$) (see [C], §2) yields

Fact 2.4. *Suppose $d = [F^+ : \mathbb{Q}]$ is even. Then $\ddagger$ can be chosen so that $G_0(F_v^+)$ is quasi-split at every finite prime of F^+ not in $S(B)$, and such that $G_0(F^+ \otimes_{\mathbb{Q}} \mathbb{R})$ is compact, and isomorphic to the product of d copies of the compact unitary group $U(n)$.*

By invoking Proposition 1.9, we see that, in order to prove modularity of an n-dimensional representation ρ of $Gal(\overline{\mathbb{Q}}/F)$ satisfying hypotheses (a), (b), (c) we can always reduce to the case in which

(2.4.1) d is even;
(2.4.2) ρ satisfies local hypothesis (c) at two places not conjugate under c;
(2.4.3) F/F^+ is unramified at all finite places;
(2.4.4) Every prime v at which ρ ramifies (including primes of residue characteristic ℓ) splits in F/F^+.

Indeed, we can replace F^+ by a totally real quadratic extension F_1^+ and F by $F_1 = F{\cdot}F_1^+$); we can assume v_0 splits in F_1; we can let F_1^+ absorb all the ramification of F/F^+ and split F/F^+ at every prime at which ρ ramifies. By using more general solvable extensions, we can eliminate most of the remaining ramification; only unipotent ramification cannot be absorbed by an appropriately chosen finite solvable extension.

Let Π be a cuspidal automorphic representation of $GL(n,F)$ satisfying conditions (1), (2), and (3) of Theorem 1.2. By replacing F^+ by F_1^+ as above one can arrange that Π satisfies a set of hypotheses strictly analogous to (2.4.1)-(2.4.4), and we do so henceforward, even though only (2.4.2) is really necessary, in light of Fact 2.4 (on the other hand, (2.4.3) and (2.4.4) do appear to be very helpful if one removes condition (3)).

Theorem 2.5. *Under these hypotheses,* Π *has a* G_0*-avatar, i.e. descends to an automorphic representation* π_0 *of* G_0*.*

This is due to Clozel and Labesse [CL] for the group G_0 introduced above. One way of comparing π_0 to Π is by means of their L-functions. An automorphic representation π of a unitary group G has so-called standard L-function $L(s,\pi)$, associated to the standard $2n$-dimensional representation of the L-group of G, introduced above. In Langlands' formalism it is then tautological that $L(s,\pi) = L(s,\Pi)$, at least at unramified places (and under the analogue of (2.4.4) this can be arranged at all places). Both the Langlands-Shahidi method and the doubling method of Piatetski-Shapiro and Rallis (studied in more detail by Shimura, and more generally by Lapid and Rallis) can be used to prove that $L(s,\pi)$ admits an analytic continuation and a functional equation of the expected type, without reference to its relation to the standard L-function of an automorphic representation of $GL(n)$.

Our simplifying hypotheses determine π_0 up to isomorphism as a representation of $G_0(\mathbf{A})$. This is because compact groups over $\mathbb{R}$ have no non-trivial L-packets, and (2.4.3) and (2.4.4) have removed the potential for L-indistinguishability at finite primes. It's almost certain that Labesse's methods show that π_0 is also unique as an automorphic representation, that is, that the abstract representation occurs with multiplicity one in the automorphic spectrum of G_0. A multiplicity one theorem would improve certain results but is unnecessary for our main applications.

Up to now we have encountered automorphic forms only in the plural, as elements of automorphic representations. In the following section we work with modules of actual automorphic forms on over Hecke algebras of mixed characteristic. This theory is available for $GL(n)$ as well as for the various unitary groups we have introduced, but it works best over the totally definite unitary group G_0.

3. Hecke algebras and unitary groups

Let Π and ρ be as in the previous sections. For now we work on the Π side. Henceforward we make the following simplifying assumption:

$$L(\Pi_\infty) = \mathbb{C}. \tag{3.1}$$

It follows that, if π_0 is the G_0-avatar of Π, then $\pi_{0,\infty}$ is the trivial representation of $G_0(F_\infty^+)$; moreover, the Galois representation ρ_ℓ will ultimately be realized in the middle-dimensional cohomology of an $n-1$-dimensional Shimura variety with $\mathbb{Q}_\ell$-coefficients. This hypothesis is irrelevant to the modularity theorems but it suffices for the applications to the Sato-Tate Conjecture, and it spares us a lot of notation. In particular, the non-trivial Hodge-Tate numbers are all of the form $h^{i,n-1-i} = 1$ with $0 \leq i \leq n-1$.

As is customary we begin by introducing an ℓ-adic integer ring $\mathcal{O}$, with fraction field K and residue field k, a finite extension of $\mathbb{F}_\ell$. Our Hecke algebras and deformation rings will all be $\mathcal{O}$-algebras.

The subspace of the space of $\mathbb{C}$-valued automorphic forms on G_0 generated by automorphic representations π_0 with $\pi_{0,\infty} = \mathbb{C}$ is just the space of automorphic forms on G_0 on which $G_0(\mathbb{R})$ acts trivially, namely

$$S(G_0, \mathbb{C})) = S_{triv}(G_0, \mathbb{C})) := C^{\infty}(G_0(F^+)\backslash G_0(\mathbf{A})/G_0(\mathbb{R})), \mathbb{C}). \tag{3.2}$$

The space $Sh(G_0) = G_0(F^+)\backslash G_0(\mathbf{A})/G_0(\mathbb{R}))$ is a profinite set, in fact a zero-dimensional Shimura variety, and the notation C^{∞} denotes the space of locally constant functions. In (3.2) these functions are taken with values in $\mathbb{C}$, but we could just as well take values in $\mathcal{O}$, or more generally in any $\mathcal{O}$-algebra A:

$$S(G_0, A)) := C^{\infty}(Sh(G_0), A). \tag{3.3}$$

This can be viewed as the cohomology in degree zero of $Sh(G_0)$, and obviously behaves well with respect to base change: if $A \to B$ is a homomorphism of $\mathcal{O}$-algebras, then $S(G_0, A) \otimes_{\mathcal{O}} B \xrightarrow{\sim} S(G_0, B)$ under the natural map. This is not always true for cohomology in higher degrees of more general Shimura varieties, not to mention the locally symmetric spaces attached to $GL(n)$, and is one of the advantages of working with G_0.

We fix a set T of primes of F^+ which will be the primes at which our π_0 (or Π, or ρ) will be allowed to ramify. We assume

$$T = S(B) \cup S_{\ell} \cup S_1 \cup R$$

where S_{ℓ} is the set of divisors of ℓ, S_1 is a *non-empty* set of auxiliary primes (descended from the $\mathfrak{r}$ of the original Taylor-Wiles paper) which allows us to eliminate elliptic fixed points in $Sh(G_0)$, and R is the set of primes at which Taylor studies possible level-raising in [T]. There will also be sets of primes disjoint from T, denoted Q_N, as N varies among positive integers; these are the Taylor-Wiles primes, used in the patching method. We let $T(Q_N) = T \cup Q_N$. These primes have the following properties:

3.4.1 All primes in $T(Q_N)$ split in F/F^+.
3.4.2 If $v \in S_1$ lies above a rational prime p then $[F(\zeta_p) : F] > n$.
3.4.3 If $v \in R$ then $\mathbf{N}v \equiv 1 \pmod{\ell}$.
3.4.4 If $v \in Q_N$ then $\mathbf{N}v \equiv 1 \pmod{\ell^N}$.

Let $\tilde{T}$ denote a set of liftings of T to primes of F, so that $\tilde{T} \coprod \tilde{T}^c$ is the set of all primes of F above T; if $v \in T$ let $\tilde{v}$ be the corresponding element of $\tilde{T}$. For any Q_N we define $\tilde{T}(Q_N)$ in the same way.

For split primes v we identify $G_0(F_v^+)$ with $GL(n, F_w)$ for some w dividing v (we choose $\tilde{v}$ for $v \in T$). Now choose an open compact subgroup U of $G_0(\mathbf{A}^f)$, $U = \prod_v U_v$, where v runs over finite primes of F^+, such that

3.5.1 If $v \notin T$, or if $v \in S_{\ell}$, then U_v is a hyperspecial maximal compact subgroup of $G_0(F_v^+)$.
3.5.2 If $v \in S(B) \cup R$ then U_v is an Iwahori subgroup.
3.5.3 If $v \in S_1$ then U_v is the principal congruence subgroup of level v:

$$U_v = \{g \in GL(n, \mathcal{O}_{F,\tilde{v}} \mid g \equiv 1 \pmod{\mathfrak{m}_{\tilde{v}}})\}$$

where $\mathfrak{m}_{\tilde{v}}$ is the maximal ideal of $\mathcal{O}_{F,\tilde{v}}$.

3.5.4 If $v \in Q_N$ then

$$U_v = U_{1,v} := \{g \in GL(n, \mathcal{O}_{F,\tilde{v}}) \mid g \equiv \begin{pmatrix} g_{n-1} & *_{n-1} \\ 0_{n-1} & 1 \end{pmatrix} \pmod{\mathfrak{m}_{\tilde{v}}}\}$$

where $g_{n-1} \in GL(n-1, \mathcal{O}_{F,\tilde{v}})$ and $*_{n-1}$ (resp. 0_{n-1}) is an arbitrary column matrix of height $n-1$ (resp. the zero row matrix of width $n-1$).

We write $\mathcal{O}_v = \mathcal{O}_{F,\tilde{v}}$ for simplicity, and let $k(v)$ denote its residue field. One likewise defines $U_{0,v} \supset U_{1,v}$ by weaking the condition in (3.5.4) so that the lower right-hand entry is an arbitrary element of $k(v)^\times$. For $v \in S(B)$, (3.5.2) implies that U_v is a maximal compact subgroup, the multiplicative group of a maximal order of B_v; for $v \in R$ U_v can be identified with integral matrices whose reduction modulo $\tilde{v}$ is upper-triangular, which we denote I_v (Iw_v in [CHT,T]). Let q_v be the order of the residue field $k(v)$, a power of the prime p_v. Let $I(1)_v \subset I_v$ be the p_v-Sylow subgroup, the matrices whose reduction modulo $\tilde{v}$ is upper-triangular unipotent; mapping to the diagonal entries thus identifies

$$I_v/I(1)_v \xrightarrow{\sim} (k(v)^\times)^n \tag{3.6}$$

A character of $I_v/I(1)_v$ is denoted $\chi_v = (\chi_{1,v}, \ldots, \chi_{n,v})$ where each $\chi_{i,v}$ is a character of $k(v)^\times$. A character of $U_{0,v}/U_{1,v} \xrightarrow{\sim} k(v)^\times$ is denoted ψ_v^0.

Let χ_v be as above, for $v \in R$, and define

$$S_{\{\chi_v\}}(U, A) = \{f \in S(G_0, A) \mid f(gu) = \prod_{v \in R} \chi_v^{-1}(u_v) f(g)\} \tag{3.7}$$

for all $g \in G_0(\mathbf{A}^f)$ and $u = \prod u_v \in U$. This is the module on which our Hecke algebras act. Suppose $A = \mathbb{C}$ (don't worry about its $\mathcal{O}$-algebra structure); then $S_{\{\chi_v\}}(U, \mathbb{C})$ is the space of vectors in the space of automorphic forms on

$$G_0(F^+)\backslash G_0(\mathbf{A})/G_0(\mathbb{R}) \cdot \prod_{v \notin R} U_v$$

on which $\prod_{v \in R} U_v$ acts by the indicated character. In particular, the only automorphic representations π_0 that contribute to $S_{\{\chi_v\}}(U, \mathbb{C})$ are those with non-trivial fixed vectors under $\prod_{v \notin R} U_v \times \prod_{v \in R} I(1)_v$ Our choice of U_v for $v \in S(B)$ implies that any π_0 has a base change Π to $GL(n, F)$ for which Π_v is an abelian twist of the Steinberg representation. In order to allow more discrete series local factors at $v \in S(B)$ (as required by condition (3)) we would need to allow representations of U_v of dimension > 1 and consider vector-valued forms with values in these representations, tensored over the places in $S(B)$. This is the point of view of [CHT] and [T]. For simplicity we prefer not to work with vector-valued forms in these notes. However, the reader is advised that certain steps in the proof of the Sato-Tate conjecture require the use of such vector-valued forms; we will point this out when appropriate.

For a place v of F^+ we let Γ_v denote a decomposition group at v. Here is how the conditions on primes in $T(Q_N)$ translate into conditions on the Galois representation $\rho = \rho_{\Pi_\ell}$, which we assume takes values in $GL(n, \mathcal{O})$[3] We write $\bar{\rho}$ for the reduction of ρ modulo the maximal ideal of $\mathcal{O}$.

3.8.1(a) If $v \notin T$, then $\rho \mid_{\Gamma_v}$ is unramified;

[3] This is true when ρ is viewed as a representation of $Gal(\overline{\mathbb{Q}}/F)$; ρ extends to a representation of $Gal(\overline{\mathbb{Q}}/F^+)$ with values in the L-group of G_0), as discussed above.

3.8.1(b) If $v \in S_\ell$, then $\rho \mid_{\Gamma_v}$ is crystalline.

3.8.2(a) If $v \in S(B)$ then the image of $\rho \mid_{\Gamma_v}$ contains a unipotent matrix τ such that $\tau - 1$ is nilpotent of maximal rank; i.e. $(\tau-1)^n) = 0$, $(\tau-1)^{n-1} \neq 0$.

3.8.2(b) If $v \in R$ then $\rho \mid_{\Gamma_v}$ may be more ramified than $\bar{\rho}$, and that is the issue resolved in [T].

3.8.3 If $v \in S_1$ then ρ is unramified at v and $\bar{\rho}$ has no deformation to a representation ramified at v.

3.8.4 If $v \in Q_N$ then ρ is unramified at v but $\bar{\rho}$ has certain deformations to representations ramified at v, and the point of the Taylor-Wiles method, as generalized in [CHT] and [T], is to use these additional deformations to bound the size of the ring of all deformations in terms of the Hecke algebra; see §3.13 for details.

The assertions (3.8.1) and (3.8.2) can be justified on the basis of the information presented up to now. This is not true of (3.8.3) and (3.8.4). The need to choose sets S_1 and Q_N with these properties requires us to impose additional hypotheses on $Im(\bar{\rho})$. That such choices are possible then follows from an argument using Chebotarev density, as in the original article of Taylor-Wiles. This will be explained below.

Now let $A = \mathcal{O}$. Our Hecke algebra is a finite free $\mathcal{O}$-algebra, given with an explicit infinite family of generators. Each *split* prime $v \notin T(Q_N)$ contributes n generators. One can also include generators at non-split primes outside T, but these are unnecessary, basically because the split primes of F have Dirichlet density 1 (this is not true of the primes of F^+ that split in F!).

Let w be a prime of F split over F^+, v its restriction to F^+, so that v factors as $w \cdot w^c$. Then $G_0(F_v^+) \xrightarrow{\sim} GL(n, F_w)$, and $\pi_{0,v} \xrightarrow{\sim} \Pi_w$, as we saw above. The Hecke polynomial attached to the unramified representation Π_w of $GL(n, F_w)$ was originally determined by Shimura and is presented in his red book on modular forms. The coefficients of this polynomial define the n Hecke operators at v; replacing the Hecke operators by their eigenvalues on the 1-dimensional $U_v \simeq GL(n, \mathcal{O}_w)$-fixed subspace of $\pi_{0,v}$, and the variable by q_v^{-s}, yields the inverse of the local Euler factor $L(s, \Pi_w)$.

Expliticly, let ϖ_w be a uniformizer at w. The Hecke operators are double coset operators

$$T_v^{(j)} = U_v(diag(\varpi_w I_j, I_{n-j}))U_v \subset G_0(F_v^+), j = 1, \dots, n, \tag{3.9}$$

These operators act in the usual way on $S_{\{\chi_v\}}(U, A)$ for any A, as does $(T_v^{(n)})^{-1}$, which is just translation by an element of the center. We denote

$$\mathbb{T}^T_{\{\chi_v\}}(U)$$

the $\mathcal{O}$-subalgebra of $End(S_{\{\chi_v\}}(U, \mathcal{O}))$ generated by the $T_v^{(j)}$, $j = 1, \dots, n$, together with $(T_v^{(n)})^{-1}$, for all split unramified $v \notin T$. The unramified local Langlands correspondence considered in §1 is normalized so that

$$P_w(X) = X^n + \sum_{j=1}^{n} (-1)^j q_w^{j(j-1)/2} T_v^{(j)} X^{n-j}, \tag{3.10}$$

with each $T_v^{(j)}$ specialized to its eigenvalue for Π_w, is the characteristic polynomial of $\rho_{\Pi,\ell}(Frob_w)$ for any ℓ, where $Frob_w$ is geometric Frobenius.

We will be using

Proposition 3.11. *The algebra* $\mathbb{T}^T_{\{\chi_v\}}(U)$ *is reduced.*

This follows in the usual way from the semisimplicity of the space of automorphic forms over $\mathbb{C}$ as an admissible $G_0(\mathbf{A}^f)$-module; recall that G_0 is anisotropic.

3.12. Surjectivity of the map $R \to \mathbb{T}$.. Now recall the automorphic representation Π of $GL(n)$. As in the earlier lectures on modularity, we associate a maximal ideal $\mathfrak{m} = \mathfrak{m}_\Pi \subset \mathbb{T}^T_{\{\chi_v\}}(U)$ to Π: first we descend Π to π_0 as in §2. The Hecke algebra acts on the U-invariants in π_0 by a character λ_Π and we let $\mathfrak{m} \subset \mathbb{T}^T_{\{\chi_v\}}(U)$ be the maximal ideal containing $ker\ \lambda_\Pi$. The localization $S_{\{\chi_v\}}(U, \mathcal{O})$ at $\mathfrak{m}$ consists roughly of those forms congruent to (the U-invariants of) of π_0, and form a module for $\mathbb{T}^T_{\{\chi_v\}}(U)_\mathfrak{m}$. In the next section we introduce the deformation ring $R^{univ} = R_{\bar{\rho}}$ of the residual representation $\bar{\rho}$. Theorem 1.2 implies there is a map R^{univ} to $\mathbb{T}^T_{\{\chi_v\}}(U)_\mathfrak{m} \otimes \overline{\mathbb{Q}}$. We will assume

Hypothesis 3.12.1. *The ideal* $\mathfrak{m}$ *is not Eisenstein; i.e.* $\bar{\rho}$ *is absolutely irreducible.*

This hypothesis, together with a result of Carayol, implies that this factors through a map

$$R^{univ} \to \mathbb{T}^T_{\{\chi_v\}}(U)_\mathfrak{m} \tag{3.12.2}$$

and the information contained in the previous paragraph implies that (3.12.2) is surjective. The arguments are identical to the familiar case of $n = 2$. If we could show that (3.12.2) is an isomorphism, then the reciprocity conjecture of (1.6) would follow for any lifting of $\bar{\rho}$ satisfying the conditions used to define $R_{\bar{\rho}}$. In fact, it is not known in general that (3.12.2) is an isomorphism, and specifically it is not known in the cases relevant to the Sato-Tate conjecture. But it is known that the map on irreducible components in characteristic zero is a bijection, and this is sufficient.

3.13. The Taylor-Wiles primes Q_N.

As indicated in 3.8.4, the primes $v \in Q_N$ are chosen so that ρ is unramified at v. It is assumed more pertinently that

$$\bar{\rho}\ \mid \Gamma_v = \bar{\psi}_v \oplus \bar{s}_v \tag{3.13.1}$$

where $\bar{\psi}_v$ is an unramified character that **does not occur as a subquotient** of $\bar{s}_v$. Since $\mathbf{N}_v \equiv 1 \pmod{\ell}$, this means in particular that

3.13.2. *There are no non-trivial* Γ_v*-extensions between* $\bar{\psi}_v$ *and* $\bar{s}_v$.

This is one of the conditions that makes the Taylor-Wiles method work. The existence of sets Q_N with these properties as well as the properties of trivializing the dual Selmer group depends on the size of the image of $\bar{\rho}$, and will be briefly discussed in §4.

4. The Taylor-Wiles-Kisin method

4.0. The Taylor-Wiles-Kisin twelve-step program to modularity.

Our goal is to prove that, for any elliptic curve E over $\mathbb{Q}$ with non-integral j-invariant, and any even integer n, the representation $\rho^n_{E,\ell}$ becomes modular over some totally real Galois extension $F^+/\mathbb{Q}$. In §5 we will explain how to prove that, for appropriate ℓ, $\bar{\rho}^n_{E,\ell}$ is (residually) modular over F^+, assuming one has a good enough modular lifting theorem over F^+.

In what follows, we always assume $\ell > n$ and ℓ is unramified in F^+. To go further – indeed, to prove the Sato-Tate conjecture for modular forms of weight > 2 – we will probably have to relax at least the second assumption, but for now the modular lifting theorems we consider only apply in this situation. Here is an outline of the main steps of the method introduced in the Taylor-Wiles paper, enhanced by Kisin, and then generalized by Taylor.

(1) Definition of a deformation problem
(2) Verification of local liftability and global representability (and non-obstruction, if relevant)
(3) Framed deformations, Galois-cohomological identification of cotangent space
(4) Local lifting conditions at easy primes (including crystalline lifting), local cohomology computations
(5) Local lifting conditions at difficult primes and framed deformations; dimensions and irreducible components of local lifting spaces.
(6) Wiles global duality argument and dimension count
(7) Local conditions for Taylor-Wiles systems
(8) Modules of modular forms, Hecke algebras, proof that $R \to \mathbb{T}$ is surjective
(9) Verification of global Taylor-Wiles axioms for modular forms (free over diamond operators, identification of coinvariants under diamond operators)
(10) Axioms on size of image of residual representation, Chebotarev arguments, auxiliary prime $\mathfrak{r}$ (now called S_1)
(11) Base change to eliminate problematic primes
(12) Patching argument (standard, but there are complications if one doesn't use level-raising).

These steps are only an outline of the program. Each implementation emphasizes different problems, and the steps do not have to be followed in order. In this case it is most judicious to start with Step 5, for the following reason.

We have a representation $\bar{\rho}$ which we know is modular over F, a quadratic CM extension of F^+, specifically that it comes from an automorphic representation Π of $GL(n, F)$, satisfying the usual three conditions. We want to show that every lifting $\tilde{\rho}$ of $\bar{\rho}$, satisfying the axioms of the deformation problem introduced in Step 1, is again modular. This will be applied in the end to $\rho^n_{E,\ell}$, whose behavior we know at all places, but there is an intermediate step involving a residually monomial representation, in which $\tilde{\rho}$ may be more ramified than $\bar{\rho}$ at an unspecified set of places R. The methods we use do not allow us to control this ramification, and so we either have to prove a level-raising theorem – this was the approach anticipated in [CHT], which depended on a conjecture called Ihara's Lemma – or to find a way to avoid level-raising altogether, as Taylor was ultimately successful in doing in [T]. By the base change principle (Proposition 1.9) we can assume the set R consists of places split in F/F^+, and that the initial representation Π such that $\bar{\rho} = \bar{\rho_\Pi}$ satisfies conditions (4.5.1) and (4.5.2), below.

4.5. Step 5: Local deformation rings in the degenerate classical limit.

When Mumford wrote an introduction to his approach to moduli via geometric invariant theory in 1970 [MS], his first example to show the importance of his stability criterion was the classification of endomorphisms of vector spaces; the presence of unipotents implies that no coarse moduli space exists. For the same reason, the functor of deformations of ℓ-adic representations of $\Gamma_K := Gal(\bar{K}/K)$, where K is a q-adic field, $q \neq \ell$, is in general not representable. The worst case is the one that arises in the problem of level raising. We consider an upper-triangular representation

$$\beta : \Gamma_K \to GL(n, \mathcal{O}); g \mapsto \begin{pmatrix} \chi_1(g) & * & \ldots\ldots & * \\ 0 & \chi_2(g) & \ldots & * \\ 0 & 0 & \ldots\ldots & * \\ 0 & 0 & 0\ldots & \chi_n(g) \end{pmatrix}$$

for some ℓ-adic integer ring $\mathcal{O}$. The diagonal entries are $\mathcal{O}^\times$-valued characters of Γ_K, whose reductions modulo $\mathfrak{m}$ are denoted $\bar{\chi}_i$, $i = 1, \ldots, n$. N.B.: The deformation problem (introduced in (4.1)) imposes a restriction only on the inertial representation $\beta \mid_{I_K}$

We will be assuming $K = F_v^+$, for some $v \in R$, so in particular (3.4.3) implies

$$q = q_v := \mathbf{N}_v \equiv 1 \pmod{\ell}. \tag{4.5.1}$$

We call (4.5.1) plus the running assumption $\ell > n$, *the classical limit mod* ℓ. After a finite cyclic extension – which makes no difference to the modularity problem, by Proposition 1. 9 – we can assume

$$\bar{\chi}_i = 1, 1, \ldots, n; \tag{4.5.2}$$

this is the *degenerate case*. By hypothesis β is tamely ramified, and hence is determined up to isomorphism by an upper-triangular representation β_I of the tame inertia group I_K^{tame} and an upper-triangular invertible Frobenius element $\Phi = \beta(Frob_K)$, satisfying

$$\Phi \beta_I(x) \Phi^{-1} = \beta_I(x^q), \forall x \in I_K^{tame}$$

Again, one can assume (after a finite solvable extension) that tame inertia is purely ℓ-adic, and letting $x_0 \in I_K^{tame}$ denote a generator of ℓ-adic tame inertia, $\Sigma = \beta_I(x_0)$, the above equality becomes

$$\Phi \Sigma \Phi^{-1} = \Sigma^q. \tag{4.5.3}$$

We are thus led to consider the moduli space of pairs of matrices (Φ, Σ) satisfying (4.5.3). More precisely, for any monic polynomial $P \in \mathcal{O}[X]$ of degree n, we let $\mathcal{M}(P, q)$ be the affine scheme over $\mathcal{O}$ representing pairs (Φ, Σ) as above, with Φ invertible, such that Σ has characteristic polynomial P. Note that (4.5.3) implies that, if $\mathcal{M}(P, q)$ is non-empty, P is invariant under the $q - th$ power operation applied to its roots. The following lemma is clear:

Lemma 4.5.4. *Suppose* $q \equiv 1 \pmod{\ell}$ *and* $P = \prod_{i=1}^n (X - \zeta_i)$, *where the* ζ_i *are* ℓ*th roots of unity in* $\mathcal{O}$. *Then*

- $\mathcal{M}(P, q) \simeq \mathcal{M}(P, 1)$
- $\mathcal{M}(P, q) \times Spec(k) \xrightarrow{\sim} \mathcal{M}((X - 1)^n, q) \times Spec(k)$.

Note that $\mathcal{M}(P,1)$ just parametrizes pairs of commuting matrices, one of which has fixed characteristic polynomial. The moduli problem makes sense over $(\mu_\ell^n)_{\mathbb{Z}}$ but only becomes interesting over $Spec(\mathbb{Z}_\ell)$, where μ_ℓ becomes connected over the closed point. The observation behind [T] is that the most degenerate case $P = (X-1)^n$ deforms to the least degenerate case $P = \prod_{i=1}^n (X-\zeta_i)$ with all ζ_i distinct. The affine algebra of $\mathcal{M}(P,q)$ is the ring of local liftings at $v \in R$ used in Kisin's version of the Taylor-Wiles method. To describe its geometric properties, we relate it to a Lie algebra variant. Let $\mathcal{N}(q)$ denote the moduli space of pairs of matrices (Φ, N), with Φ invertible, N nilpotent (characteristic polynomial X^n) and

$$\Phi N \Phi^{-1} = qN \tag{4.5.5.}$$

Lemma 4.5.6. *Assume $\ell > n$. Then*

(i) $\mathcal{N}(q)^{red}$ is a union of reduced irreducible components parametrized by nilpotent conjugacy classes in $Lie(GL(n))$; i.e. by partitions of n (Jordan block decomposition).

(ii) Each reduced irreducible component Z of $\mathcal{N}(q)$ is equidimensional of dimension n^2+1, Z_k is irreducible of dimension n^2 and generically reduced, and each irreducible component of $\mathcal{N}(q) \times Spec(k)$ is contained in a unique irreducible component of $\mathcal{N}(q)$ which is not purely of characteristic ℓ.

(iii) The logarithm and exponential (applied to Σ) identify

$$\mathcal{M}((X-1)^n, q)^{red} \xrightarrow{\sim} \mathcal{N}(q)^{red}.$$

In particular, the reduced irreducible components of $\mathcal{M}((X-1)^n, q)^{red}$ have the properties (ii).

At the other extreme:

Lemma 4.5.7. *Let $P = \prod_{i=1}^n (X-\zeta_i)$ with all ζ_i distinct. Then $\mathcal{M}(P,1) \times K$ is smooth and irreducible of dimension n^2, whereas $\mathcal{M}(P,1) \times k \xrightarrow{\sim} \mathcal{N}(1) \times k$, and hence has components indexed by partitions of n as in the previous lemma.*

Moreover, the completion of the affine ring of $\mathcal{M}(P,1)$ at the closed point of the special fiber corresponding to $\Sigma = 1$ and $\Phi = 1$ has a unique minimal prime.

In the second statement we just send Σ to $\Sigma - 1$, which is why there is no need to consider reduced components.

The ζ_i are the eigenvalues of $\beta(x_0)$. We can identify $Syl_\ell(k_v^\times)$ (ℓ-Sylow subgroup) with the subgroup of $Gal(K^{ab}/K)$ generated by x_0, and so we define χ_i to be the character of $k_v^\times$ of ℓ-power order whose image on x_0 is ζ_i. We let

$$R_{v,\chi}^{loc}$$

be the affine $\mathcal{O}$-algebra of $\mathcal{M}(P_\chi, 1)$, where $P_\chi = \prod_{i=1}^n (X-\zeta_i)$ as above. Thus χ and ζ are alternative notation for the same thing; we have already seen χ in the discussion of R in §3 in connection with Hecke algebras. The notation R^{loc} will be explained in the following section.

4.1: Step 1: Definition of a deformation problem.

In this section we recast the conditions already used to define the Hecke algebra $\mathbb{T}^T_{\{\chi_v\}}(U) \otimes \overline{\mathbb{Q}}$ in Galois-theoretic terms, defining a deformation problem. Write Γ_F

and Γ_{F^+} for $Gal(\overline{\mathbb{Q}}/F)$ and $Gal(\overline{\mathbb{Q}}/F^+)$, respectively. We start with a representation $\rho : \Gamma_F \to GL(n, \mathcal{O})$, satisfying condition (b) of Theorem 1.2. We always assume $\bar{\rho} = \rho \pmod{\mathfrak{m}}$ is absolutely irreducible. Then Lemma 1.10.2 implies that ρ extends to a homomorphism $r : \Gamma_{F^+} \to \mathcal{G}_n(\mathcal{O})$, with $r(c)$ not in the identity component of $\mathcal{G}_n$. Let $\bar{r} = r \pmod{\mathfrak{m}}$. Our deformation problem concerns lifts of $\bar{r}$ to homomorphisms $\tilde{r} : \Gamma_{F^+} \to \mathcal{G}(A)$, where A runs over the category $\mathcal{C}$ of finitely generated local $\mathcal{O}$-algebras with residue field k. These liftings satisfy the following local properties at places v of F^+, indexed for comparison with (3.8.?). We write $\tilde{\rho}$ for the restriction of $\tilde{r}$ to a homomorphism $\Gamma_F \to GL(n, A)$, $\tilde{r}_v$ for $\tilde{r} \mid_{\Gamma_v}$, and define $\tilde{\rho}_v$ analogously. .

4.1.1(a) If $v \notin T$, then $\tilde{r}_v$ is unramified;

4.1.1(b) If $v \in S_\ell$, then $\tilde{\rho}_v$ is crystalline, in the sense of being in the essential image of the Fontaine-Laffaille functor. Moreover, $\tilde{\rho}_v$ has Hodge-Tate weights such that $h^{i,n-1-i} = 1$ for $0 \leq i \leq n-1$.

4.1.2(a) If $v \in S(B)$ then $\tilde{\rho}_v$ fixes a flag of a certain form (condition to which we allude in §4.4).

4.1.2(b χ) If $v \in R$ then $\tilde{\rho}_v \mid_{I_v}$ satisfies the condition χ described at the end of §4.5, see below.

4.1.3 If $v \in S_1$ then there is no restriction on $\tilde{r}_v$ (but none is necessary).

4.1.4 If $v \in Q_N$ then $\tilde{\rho}_v$ breaks up as a direct sum $\psi \oplus s$ where ψ lifts $\bar{\psi}_v$ and s is an unramified lifting of $\bar{s}_v$. (The possible additional tame ramification of ψ adds one degree of freedom, as required.)

Condition (4.1.2(b χ)) is interpreted by considering the moduli space of all liftings of $\tilde{\rho}_v$, which is represented by the ring of coefficients of matrices subject to certain relations determined by $\bar{r}$, and only allowing the liftings in a certain closed subset (the characteristic polynomial of the element Σ defined in §4.5) determined by χ. This closed subset is the spectrum of $R^{loc}_{v,\chi}$.

4.2: Step 2: Verification of local liftability and global representability.

A *deformation* of $\bar{r}$ of type (4.1.?) is a lifting satisfying conditions (4.1.1-4.1.4), considered up to conjugation by a matrix in $1 + \mathfrak{m}_A M(n, A)$.

Proposition 4.2.1. *Assume $\bar{\rho}$ is absolutely irreducible. Then the functor of deformations is representable in $\mathcal{C}$ by a ring $R^{univ}_{\chi,N}$, and there is a universal deformation*

$$r^{univ} : \Gamma_{F^+} \to \mathcal{G}_n(R^{univ}_{\chi,N}).$$

If Q_N is empty we just write R^{univ}_χ.

The notation $R^{univ}_{\chi,N}$ should really be R^{univ}_{χ,Q_N}, but the Q_N will be understood. Because condition (4.1.2)(b χ) is not liftable, the usual Galois cohomological techniques do not allow us to calculate the number of generators of the maximal ideal of $R^{univ}_{\chi,N}$. Instead, one needs to consider framed deformations.

4.3. Step 3: Framed deformations, Galois-cohomological identification of cotangent space.

A *framed deformation* of $\bar{r}$ to a ring A in $\mathcal{C}$, of type χ, is a lifting $\tilde{r}$ of type χ as in (4.1.1-4.1.4), together with a local representation r_v of Γ_v of type (4.1.2 (b) χ), and isomorphisms (local framings)

$$\alpha_v : \tilde{r} \mid_{\Gamma_v} \xrightarrow{\sim} r_v,$$

for all $v \in R$, all taken up to $1 + \mathfrak{m}_A M(n, A)$-conjugacy.

There is a universal framed deformation of $\bar{r}$ of type χ over an object $R^{\square}_{\chi,N}$ of $\mathcal{C}$. Forgetting the framings defines a map $R^{univ}_{\chi,N} \to R^{\square}_{\chi,N}$. We drop the subscript N for the remainder of this section.

On the other hand, let

$$R^{loc}_{\chi} = \widehat{\otimes}_{v \in R} R^{loc}_{v,\chi}$$

(denoted R_R in [T]):

$$\mathcal{T}_R = \mathcal{O}[[X_{v,i,j} \mid v \in R\ , i, j = 1, \ldots, n]]$$

(matrix coefficients of liftings for each v to parametrize r_v). If r^{univ} is the universal deformation of $\bar{r}$, choose a lifting that represents it (i.e. in its equivalence class). Such a choice of matrix coordinates gives rise to maps

$$R^{loc}_{\chi} \to R^{\square}_{\chi,N};\ \mathcal{T}_R \to R^{\square}_{\chi,N}$$

(the latter defined by setting $\alpha_v = I_n + (X_{v,i,j})$ in the given matrix coordinates) and one has the fundamental

Lemma 4.3.1. Smoothing lemma for framings.

$$R^{univ}_{\chi,N} \hat{\otimes} \mathcal{T}_R \xrightarrow{\sim} R^{\square}_{\chi,N}.$$

Now we are in a position to calculate the embedding dimension of $R^{\square}_{\chi}$ as an R^{loc}_{χ}-algebra

Lemma 4.3.2. *$R^{\square}_{\chi}$ can be topologically generated over R^{loc}_{χ} by*

$$\dim_k H^1_{(4.1.\bullet)}(\Gamma_{F^+}, ad\bar{r}) + \sum_{v \in R} \dim_k H^0(\Gamma_v, ad\bar{r})$$

elements.

The proof in [T] is analogous to the usual calculations of

$$\dim \mathfrak{m}_{R^{\square}_{\chi,N}} / (\mathfrak{m}^2_{R^{\square}_{\chi,N}} + \mathfrak{m}_{R^{loc}_{\chi}} R^{\square}_{\chi,N}),$$

which are deformations over $k[\epsilon](\epsilon)^2$ with the R-data fixed (equal to $\bar{r}\ |_{\Gamma_v}$ at v).

4.4. Step 4: Local lifting conditions at easy primes (including crystalline lifting), local cohomology computations.

The "easy" primes in question are those in S_ℓ and archimedean primes on the one hand, and $S(B)$ on the other. In Richard Taylor's lectures at MSRI in October 2006, he incorporated local data about S_ℓ and ∞ into the singular coefficient ring $R^{loc}_{?}$. Moreover, by relying on base change, he implicitly eliminated the distinction between the primes in $S(B)$ and the primes in R. Thus all local cohomology computations at ramified primes are placed on the same footing and incorporated into an appropriate R^{loc}. The article [T] appears to have solved all problems concerning primes other than those in S_ℓ, the archimedean primes being genuinely easy to understand in all situations. I will say no more about this here except to mention that

(a) Under hypotheses (4.1.1)(b), whether we incorporate local data at ℓ into R^{loc} or whether we make (4.1.1)(b) a condition defining a Selmer group, as in [CHT] and [T], this step is based on calculations in the Fontaine-Laffaille theory that have been known for years, and are considered standard;
(b) the calculation at ∞ depends on the parity of the polarization of $\bar{\rho}_{\Pi}$ for an automorphic representation satisfying conditions (1)-(3), defined by the action of c and the pairing given by condition (b). Basically, the image $r(c) \in \mathcal{G}_n(k)$ acts by conjugation on the Lie algebra of $GL(n)$ and one needs to determine its signature. This must be in Langlands' general conjectures on zeta functions of Shimura varieties, but it does not appear to be accessible to available techniques. So the calculations in [CHT] and [T] drag along an unknown term denoted $\frac{1+\chi_v(c)}{2}$ for each real place v. At the end it turns out that they all vanish, which completes the local reciprocity calculation at archimedean primes; we know no other way to carry out this calculation.
(c) In the end, the calculation depends on a certain numerical coincidence in which the contributions of primes in S_ℓ and of archimedean primes balance each other. We return to this point in (4.7.2). The spurious terms $\frac{1+\chi_v(c)}{2}$ mentioned in (b) does not disturb this balance; on the contrary, the vanishing of these terms follows from an a priori global inequality (roughly, that the dimension of local ring is bounded below by 0).

In [CHT] and [T], the local condition at $S(B)$ is expressed in terms of filtrations, as indicated in (4.1.2)(a). As mentioned above, Taylor now thinks this is unnecessary, so I omit the details.

4.6. Step 6: Local conditions at Q_N..

This is contained in (4.1.4). The extra degree of freedom reappears in the following step.

4.7. Step 7: Wiles global duality argument and dimension count.

As in Wiles' original paper, one applies Poitou-Tate duality and the local Euler characteristic formula at the primes considered in Steps 4-6 to rewrite the formula in Lemma 4.3.2 as follows:

Proposition 4.7.1. *$R^{\square}_{\chi,N}$ can be topologically generated over R^{loc}_{χ} by*

$$|Q_N| + \dim_k H^1_{(4.1.\bullet)^\perp}(\Gamma_{F^+}, ad\bar{r}(1)) - \delta_{F^+} - n \sum_{v \div \infty} \delta_v,$$

elements where (1) *denotes Tate twist and for real v, $\delta_v = \frac{1+\chi_v(c)}{2}$ as discussed in Step 4 (b), and*

$$\delta_{F^+} = \dim_k H^0(\Gamma_{F^+}, ad(\bar{r})(1)).$$

The global term will vanish by our hypothesis on $im(\bar{\rho})$, and the archimedean terms will be dragged along until the end, as mentioned above.

4.7.2. Explanation. The subscript in $H^1_{(4.1.\bullet)^\perp}(\Gamma_{F^+}, \bullet)$ is a reference to the notation of Wiles and Taylor-Wiles. The conditions (4.1.1-4.1.4) translate, for each

$v \in T$, to a Selmer condition on the global Galois cohomology group defined by a subspace $L_v \subset H^1(\Gamma_v, \mathrm{ad}(\bar{r}))$. By Tate local duality there is a dual subspace

$$L_v^{\perp} \subset\subset H^1(\Gamma_v, \mathrm{ad}(\bar{r})(1))$$

which is used to define the global H^1 term appearing in (4.7.1)

A few words: [CHT] and [T] follow the original approach of Wiles, and the duality argument, together with the local Euler characteristic formula, introduces local terms at S_ℓ and ∞ which almost cancel, leaving the spurious term $n \sum_{v \div \infty} \delta_v$. This numerical coincidence is crucial to the success of the method. If we worked over F instead of F^+ there would be no cancellation, and the deformation ring would be too big relative to the Hecke algebra, another reason we are forced to assume the polarization hypothesis (b).

In the approach followed in Taylor's MSRI lectures, the numerical coincidence is absent from the application of global duality but reappears in the calculation of the relative dimensions (over $\mathcal{O}$) of the R^{loc}. The coincidence is that the extra dimensions one obtains for R^{loc} (at the primes in S_ℓ, ∞, and R) is exactly the relative dimension of $R_\chi^\square$ over R_χ^{univ}, as in the Smoothing Lemma 4.3.1. This is what makes the patching argument work at the end.

4.8. Step 8: Modules of modular forms, Hecke algebras, proof that $R \to \mathbb{T}$ is surjective.

To emphasize the role of the set Q_N of Taylor-Wiles primes we henceforward write $U = U(Q_N)$ for the level subgroup, and write U for the level subgroup with $Q_N = \emptyset$. The modules of modular forms and their Hecke algebras were discussed in §3, as was the surjectivity of

$$R_{\chi,N}^{univ} \to \mathbb{T}_{\chi_v}^T(U(Q_N))_{\mathfrak{m}}.$$

For the purposes of the Taylor-Wiles argument there are too many forms locally at $v \in Q_N$. Recall that at such v we let ψ_v^0 denote a character of $U_{0,v}/U_{1,v} \xrightarrow{\sim} k(v)^\times$. We drop the index v for the moment and let K be a non-archimedean local field.

Lemma 4.8.1. *Let π be a principal series representation $I(\underline{\psi})$ of $GL(n,K)$ induced by an n-tuple $\underline{\psi} = (\psi_1, \ldots, \psi_n)$ of characters of $K^\times$. Suppose ψ_i is unramified for $i > 1$ and ψ_1 is at most tamely ramified. Let ψ_1^0 be the restriction to $\mathcal{O}_K^\times$, or equivalently to $k(v)^\times$. Then $\dim Hom_{U_0}(\psi_1^0, \pi) = n$ if ψ_1^0 is trivial, $= 1$ otherwise.*

The above Lemma is independent of the coefficients; it remains valid for (smooth) principal series representations with coefficients in $\bar{\mathbb{F}}_\ell$. Since we are in the classical limit, the representation theory over $\bar{\mathbb{F}}_\ell$ is particularly simple: Vignéras proved in [V] that every $\pmod{\ell}$ principal series representation is completely reducible, and the factors are the same as the factors of the corresponding module over the Hecke algebra, which are easy to identify. In particular, for appropriate $\underline{\psi}$, the $\pmod{\ell}$ Steinberg is a direct summand of $I(\underline{\psi})$. These facts are used in what follows.

The Taylor-Wiles method involves patching spaces of modular forms of level $U(Q_N)$, localized at the ideal $\mathfrak{m}$ of $\mathbb{T}_{\chi_v}^T(Q_N)$, for a set of Q_N with $N \to \infty$. The starting point is forms of level U, which are unramified at Q_N. For each N, we need to compare $S_{\chi_v}(U, \mathcal{O})_{\mathfrak{m}}$ with $S_{\chi_v}(U(Q_N), \mathcal{O})_{\mathfrak{m}}$ as $\mathbb{T}$-modules. Write $U_0(Q_N) \supset U(Q_N)$ the level subgroup with $U_{1,v}$ replaced by $U_{0,v}$ for all $v \in Q_N$.

There are actually two comparisons made. The first is between $S_{\chi_v}(U(Q_N), \mathcal{O})_{\mathfrak{m}}$ and $S_{\chi_v}(U_0(Q_N), \mathcal{O})_{\mathfrak{m}}$. The former contains global π with π_v tamely ramified at $v \in Q_N$ – and with only one degree of freedom for the ramification – whereas the latter contains only those global π with $\pi_v^{U_{0,v}} \neq 0$.

Lemma 4.8.2. *With Q_N as in 3.13, $S_{\chi_v}(U_0(Q_N), \mathcal{O})_{\mathfrak{m}}$ consists only of π with π_v unramified at all $v \in Q_N$.*

One actually uses something stronger: that each π_v is an unramified principal series that is residually irreducible as far as the character $\bar{\psi}_v$ is concerned (cf. (3.13.1)). In other words, any reducibility comes from reducibility of the unramified principal series of $GL(n-1)$ corresponding to the summand $\bar{s}_v$. I will not attempt to make this more precise. Assume for simplicity that $\pi_v \pmod{\ell}$ is an irreducible unramified principal series. Then we have seen that $\dim(\bar{\pi}_v^{U_{0,v}}) = n$, whereas the tamely ramified constituents of $S_{\chi_v}(U(Q_N), \mathcal{O})_{\mathfrak{m}}$ have only a one-dimensional $U_{0,v}$-invariant subspace. We use a Hecke operator for $U_{0,v}$, or for $U_{1,v}$, to cut out a 1-dimensional subspace of $\dim(\bar{\pi}_v^{U_{0,v}}) = n$. Namely, there is a Hecke operator

$$V_v = U_{?,v}(diag(I_{n-1}, \varpi_v))U_{?,v}, \ ? = 0, 1 \tag{4.8.3}$$

that acts on the $U_{0,v}$-fixed subspace and decomposes it as a sum of generalized eigenspaces with eigenvalues equal to the $\psi_{i,v}(\varpi_v))$, $i = 1, \ldots, n$. Our assumption that $\pi_v \pmod{\ell}$ is basically equivalent (in the classical limit) to the hypothesis that the V_v-eigenvalues are multiplicity free ($\psi_{i,v} \neq \psi_{j,v}$ if $i \neq j$).

Let H_{1,Q_N} denote the $\mathcal{O}$-submodule of $S_{\chi_v}(U(Q_N), \mathcal{O})_{\mathfrak{m}}$ on which V_v acts as $\psi_{1,v}(\varpi_v)$, and define $H_{0,Q_N} \subset S_{\{\chi_v\}}(U_0(Q_N), \mathcal{O})_{\mathfrak{m}}$ likewise. One has to be careful in making sense of this: $\psi_{1,v}$ varies with the different automorphic representations π contributing to $S_{\{\chi_v\}}(U(Q_N), \mathcal{O})_{\mathfrak{m}}$, but they are all congruent modulo $\mathfrak{m}$, by construction. This comes down to showing that the characteristic polynomial of V_v over $\mathbb{T}$ has a linear factor. One sees similarly that H_1 and H_0 are direct factors of the appropriate S_{χ_v}. The comparison between H_1 and H_0 is the subject of Step 9.

The second comparison is between the $\mathbb{T}$-modules H_0 and $H := S_{\{\chi_v\}}(U, \mathcal{O})_{\mathfrak{m}}$. There is a subtle point, because the former is a module over $\mathbb{T}_{\mathfrak{m}}^{T(Q_N)}$, whereas the latter is a module over $\mathbb{T}_{\mathfrak{m}}^{T}$. In principle $\mathbb{T}_{\mathfrak{m}}^{T}$ contains extra Hecke operators for the places in Q_N, but a density argument using the surjectivity $R^{univ} \rightarrow \mathbb{T}_{\mathfrak{m}}$ shows that the map $\mathbb{T}_{\mathfrak{m}}^{T(Q_N)} \rightarrow \mathbb{T}_{\mathfrak{m}}^{T(Q_N)}$ ("forget the level at Q_N") is surjective.

Now the $\mathbb{T}$ isomorphism

$$H \rightarrow H_0 \tag{4.8.4}$$

is given by an explicit level-raising operator constructed as a polynomial in the V_v. The construction of such an operator is again ultimately based on Vignéras' classification [V] of irreducible spherical representations of $GL(n)$ in the classical limit, as well as some standard constructions from the theory of types in representations of p-adic groups.

N.B. Level-raising works at the primes in Q_N because they are chosen for that purpose. The existence of such primes is guaranteed by the Chebotarev arguments in Step 10. The primes in R, by contrast, are part of the initial data; we cannot control them.

4.9. Step 9: Verification of global Taylor-Wiles axioms for modular forms.

The sets Q_N will all be chosen to have u elements, where

$$u = \dim_k H^1_{(4.1.1-3)^\perp}(\Gamma_{F^+}, \mathrm{ad}\bar{r}(1)). \tag{4.9.1}$$

N.B. This dimension is the integer we encountered in Proposition 4.7.1 when $Q_N = \emptyset$. Let

$$\Delta_{Q_N} = \prod_{v \in Q_N} Syl_\ell(k(v)^\times) \simeq (\mathbb{Z}/\ell^N\mathbb{Z})^u.$$

(This last $\simeq$ is a bit of poetic license. In general there is a surjective map from $Syl_\ell(k(v)^\times)$ to $\mathbb{Z}/\ell^N\mathbb{Z}$ for each $v \in Q_N$, and we use this map to define the diamond operators Δ_{Q_N}.) The deformation ring $R^{univ}_{\chi,N}$ is an $\mathcal{O}[\Delta_N]$-algebra through the action on the liftings ψ_v of $\bar{\psi}_v$. The map (3.12.2) makes H_{1,Q_N} into an $\mathcal{O}[\Delta_N]$-module. Property (i) of the global Langlands correspondence of Theorem 1.2 (local-global compatibility), combined with the properties of base change discussed (or not discussed) in §2, imply that

Taylor-Wiles property 4.9.2. *This is exactly the same as the natural action on H_{1,Q_N} of*

$$\Delta_N \subset \prod_{v \in Q_N} U_{0,v}/U_{1,v}$$

Taylor-Wiles property 4.9.3. *H_{1,Q_N} is a (finite) free $\mathcal{O}[\Delta_N]$-module.*

Actually we show that $S_{\{\chi_v\}}(U(Q_N), \mathcal{O})$ is a (finite) free $\mathcal{O}[\Delta_N]$-module because it is just the module of $\mathcal{O}$-valued functions on a space on which Δ_N acts freely. It is here that we use the auxiliary set S_1, specifically properties (3.4.2) and (3.5.3), which imply that the level $U(Q_N)$ is always sufficiently small that Δ_N acts without fixed points. Since H_{1,Q_N} is an $\mathcal{O}[\Delta_N]$-direct factor of $S_{\{\chi_v\}}(U(Q_N), \mathcal{O})$, it is also free.

In the approach of [T], as in Kisin's work, one does not apply the Taylor-Wiles method directly to H_{1,Q_N} but rather to an artifically enhanced version:

$$H_N^\square = H_{1,Q_N} \otimes_{R^{univ}_{\chi,N}} R^\square_{\chi,N}.$$

Let $\mathcal{T}_{R,N} = \mathcal{T}_R \otimes \mathcal{O}[\Delta_N]$, with $\mathcal{T}_R$ as in the smoothing lemma 4.3.1. Combining the smoothing lemma with (4.9.3), we find that

Corollary 4.9.4. *For any N, $H_N^\square$ is a finite free $\mathcal{T}_{R,N}$-module.*

In [T] $\mathcal{T}_{R,N}$ is called $S_\infty/\mathfrak{a}_N$. Note that the action of $\mathcal{T}_{R,N}$ factors through the action of $R^\square_{\chi,N}$ (the action of Δ_N is already contained in $R^{univ}_{\chi,N}$).

There is an augmentation

$$\mathcal{T}_{R,N} \to \mathcal{O} \tag{4.9.5}$$

defined by sending all the variables $X_{v,i,j}$ to 0 and all the elements of Δ_N to 1. Let $\mathfrak{a}$ denote the kernel of this augmentation.

Taylor-Wiles property 4.9.6. *There are natural isomorphisms*

$$\mathcal{T}_{R,N}/\mathfrak{a} \otimes_{\mathcal{T}_{R,N}} H_N^{\square} \xrightarrow{\sim} H_{0,Q_N} \xrightarrow{\sim} H,$$

where the second isomorphism is the inverse of (4.8.4).

This is an easy consequence of (4.9.2) and (4.8.4).

4.10. Step 10: Axioms on size of image of residual representation, Chebotarev arguments, auxiliary prime $\mathfrak{r}$ (now called S_1).

As in the original Wiles and Taylor-Wiles articles, the theorems of [CHT] and [T] require that the image of the residual representation be "big" in order to apply Chebotarev-type arguments to guarantee the existence of the Q_N and S_1 satisfying the requirements of the earlier steps. There is no optimal definition of "big," but the following axioms suffice.

(4.10.1) The extension of F^+ fixed by ker $\mathrm{ad}(\bar{r})$ does not contain $\mathbb{Q}(\mu_\ell)$;
(4.10.2) $H^i(Gal(\overline{\mathbb{Q}}/F^+(\mu_\ell), \mathrm{ad}^0\bar{r}) = 0$, $i = 0, 1$;
(4.10.3) Let $W \subset \mathrm{ad}\bar{r}$ be an irreducible $Gal(\overline{\mathbb{Q}}/F^+(\mu_\ell)$-submodule. Then there exists $h \in Gal(\overline{\mathbb{Q}}/F^+(\mu_\ell)$ and $\alpha \in k$ with the following properties: (1) The α-generalized eigenspace $V_{h,\alpha}$ of h in $\bar{r}$ is 1-dimensional; (2) If $\pi_{h,\alpha} : \bar{r} \to V_{h,\alpha}$, resp. $i_{h,\alpha} : V_{h,\alpha} \to \bar{r}$, is the h-equivariant projection (resp. injection), then $\pi_{h,\alpha} \circ W \circ i_{h,\alpha} \neq 0$.

The first two are fairly familiar, and (4.10.2) in particular includes the hypothesis that certain global terms in the dimension formula vanish. All that can be said of (4.10.3) is that it can frequently be verified, in particular when the image of $\bar{r}$ contains $Sym^{n-1}(SL(2,k))$ or when $\bar{r}$ is a monomial representation induced from a sufficiently regular character of a cyclic extension of F^+ of degree n. All three axioms are satisfied by $\rho^n_{E,\ell}$ for almost all ℓ.

There is a more problematic notion of "big" representation that enters into the application of a generalization of Ramakrishna's level-raising theorem to make sure that some of the auxiliary Galois representations satisfy condition (c); this is perhaps the most technical point of [CHT] and I will say no more about it.

4.11. Step 11: Base change to eliminate problematic primes.

We have been doing this all along. All I want to add here is that we also use the Skinner-Wiles trick of successive cyclic base changes to avoid dealing with the problem of *level-lowering* as in Ribet's paper on the level in Serre's conjecture.

4.12. Step 12: Patching argument.

I will be brief and refer the reader to the original articles. One simultaneously patches for the problems $R_\chi^\square$ and $R_1^\square$, where χ is generic, as in Lemma 4.5.7. Lemma 4.5.7, plus the dimension calculation and the Auslander-Buchsbaum theorem, implies that $H_{\chi,\infty}^\square$ is a nearly faithful $R_{\chi,\infty}^\square$-module. Since

$$R_{\chi,\infty}^\square/\lambda \simeq R_{1,\infty}^\square/\lambda$$

and likewise for the H's, this implies that $H_{1,\infty}^\square/\lambda$ is a nearly faithful $R_{1,\infty}^\square/\lambda$-module. But then the comparison of irreducible components in characteristic zero and characteristic ℓ (Lemma 4.5.6) implies that $H_{1,\infty}^\square$ is a nearly faithful $R_{1,\infty}^\square$-module, hence (by 4.3.1 and 4.9.6) that H is a nearly faithful R_1^{univ}-module, hence that

$$(R_1^{univ})^{red} \to \mathbb{T}_1$$

is an isomorphism.

4.13 Conclusion.

Here is the modular lifting theorem proved in [CHT] and [T] for representations of Γ_{F^+}. There is a similar theorem for $Gal(\overline{\mathbb{Q}}/F)$. Conditions (a)-(c) are as in Theorem 1.2.

Modular lifting theorem. *Let $\ell > n$ be a prime unramified in F^+ and let*

$$r:\ \Gamma_{F^+} \to GL(n, \bar{\mathbb{Q}}_\ell)$$

be a continuous irreducible representation satisfying the following properties:

(a) *r ramifies at only finitely many primes, is crystalline at all primes dividing ℓ, and is Hodge-Tate regular*
(b) *$r \simeq r^\vee(1-n)\cdot\chi$ where $(1-n)$ is the Tate twist and χ is a character whose value is constant on all complex conjugations;*
(c) *At some finite place v not dividing ℓ r_v corresponds to a square-integrable representation of $GL(n, F_v^+)$ under the local Langlands correspondence.*

In addition, we assume that $\bar{r}$

(d) *has "big" image in the sense of (4.10) above;*
(e) *is absolutely irreducible;*
(f) *is of the form $\rho_{\bar{\Pi},\ell}$ for some cuspidal automorphic representation Π of $GL(n, F^+)$ satisfying conditions (i)-(iii) of Theorem 1.2.*

Hodge-Tate regularity is as explained in §1. I repeat that condition (c), which causes most of the headaches in [CHT] and [HST], can probably be removed once there is a sufficiently explicit version of the stable trace formula for cohomological representations of unitary groups. The method really breaks down if we don't know condition (d), although it must be irrelevant; the approach of Skinner-Wiles to residually reducible representations looks very hard for $n > 2$; as for (f), one could hope to formulate a generalization of the Serre conjecture. Meanwhile, in the following section, I will explain how to remove this hypothesis if one is willing to settle for potential modularity, which suffices for the Sato-Tate conjecture.

5. Potential modularity of even-dimensional symmetric powers

The Langlands functoriality conjectures, applied to $GL(n)$, include the following prediction.

Conjecture 5.1. *Let*

$$\tau : GL(n_1) \times GL(n_2) \times \ldots GL(n_r) \longrightarrow GL(N)$$

be an irreducible algebraic representation. Let F be a number field and let $\pi_1, \ldots, \pi_r$ be cuspidal automorphic representations of $GL(n_i, F)$, $i = 1, \ldots, r$. Then there is an automorphic representation (functorial transfer) $\tau(\pi_1 \boxtimes \pi_2 \boxtimes \cdots \boxtimes \pi_r)$ of $GL(N, F)$ such that, at almost all (all?) places v,

$$\mathcal{L}(\tau(\pi_1 \boxtimes \pi_2 \boxtimes \cdots \boxtimes \pi_r)_v = \tau \circ (\mathcal{L}(\pi_{1,v} \otimes \cdots \otimes \pi_{r,v}),$$

where $\mathcal{L}(\pi_{i,v})$ is the n_i-dimensional representation of the Weil group at v given by the local Langlands correspondence.

There are of course predictions of functoriality for more general homomorphisms of L-groups. I'm not sure what question to ask about the exceptional places; one might ask that the transfer under τ be *isobaric* in Langlands' sense.

Whether or not the transfer is cuspidal depends both on τ and on the original π_i, and the answer does not admit a simple description. However, when $r = 1$, one expects that its transfer will be cuspidal provided $\pi = \pi_1$ is sufficiently general; then the standard L-function $L(s, \tau(\bullet))$ is entire and satisfies the functional equation for $GL(n)$. In the case of automorphic representations of $GL(2)$ attached to elliptic curves " sufficiently general" excludes only elliptic curves with complex multiplication.

The case $r = 2$, with $N = n_1 \cdot n_2$ and τ the standard tensor product representation, may be considered the main open question in automorphic forms. The following list exhausts all the cases I know (excluding the trivial cases, where some of the $n_i = 1$):

(1) $r = 1$, $n(= n_1) = 2$, $N = 3$ (symmetric square): Gelbart-Jacquet.
(2) $r = 1$, $n(= n_1) = 2$, $N = 4, 5$ (symmetric cube and fourth power) Kim-Shahidi, Kim.
(3) $r = 2$, $n_1 = n_2 = 2$, τ the tensor product: Ramakrishnan.
(4) $r = 2$, $n_1 = 2$, $n_2 = 3$, τ the tensor product: Kim.

These results are *unconditional* and apply to all automorphic representations over all number fields. The articles [CHT], [HST], and [T] together prove a weak version of what is expected when F is totally real, $r = 1$, $n = 2$, and N is even. They apply only to representations of discrete series type at archimedean places, and indeed only to holomorphic Hilbert modular forms of weight $(2, \ldots, 2)$. There is moreover a local condition (Steinberg at one finite place) that it should be possible to remove. Most importantly, the functorial transfer is not constructed over F itself but rather over an unknown Galois extension of F.

5.1. Potential modularity theorems and the Sato-Tate conjecture.

We return to the notation of (1.7). Let E be an elliptic curve over F^+, and assume it is known to be modular (e.g., $F^+ = \mathbb{Q}$); let Π_E be the corresponding automorphic representation of $GL(2, E)$, so that $L(s, \Pi_E) = L(s, E)$, with the L-function normalized to have center of symmetry at $s = \frac{1}{2}$. We let $\rho^n_{E,\ell} = Sym^{n-1}\rho_{E,\ell}$ as before. For almost all p, the local factor $\Pi_{E,p}$ is unramified; let α_p, β_p be the Satake parameters with unitary normalization.

Eichler-Shimura ("Ramanujan conjecture").

$$|\alpha_p| = |\beta_p| = 1.$$

Up to permutation have $\alpha_p = e^{i\theta_p}, \beta_p = e^{-i\theta_p}$, say, with $0 \leq \theta_p \leq \pi$.

Sato-Tate Conjecture. *Assume E has no complex multiplication. Then the θ_p are equidistributed in $[0, \pi]$ with respect to the measure $dST(\theta) := \frac{2}{\pi} sin^2\theta \; d\theta$.*

(The Sato-Tate measure is the push-forward of the Haar measure on $SU(2)$ to a measure on the set of conjugacy classes in $SU(2)$, which can be identified with $[0, \pi]$.)

Theorem. *Suppose E is an elliptic curve over $\mathbb{Q}$ (or any totally real field) with multiplicative reduction at some prime; i.e. $j(E)$ is not integral. Then the Sato-Tate Conjecture holds for E.*

A simple argument involving Shahidi's theorem on non-vanishing of Rankin-Selberg L-functions on the line $Re(s) = 1$, together with an application of Brauer's theorem on induced characters, as in Taylor's original article on potential modularity [T02], shows it suffices to prove

Theorem 5.1.1. *For every even n, there exists a totally real Galois extension F_n/F^+ such that $\rho_E^m \mid_{Gal(\overline{\mathbb{Q}}/F_n)}$ is automorphic over F_n – there exists a cuspidal automorphic representation Π_E^m of $GL(m, F_n)$, with $L(s, \Pi_E^m) = L(s, (\rho_E^m)_{F_n})$ for all* **even** *$m \leq n$.*

In particular, for all n (even or odd), $L(s, \rho_E^n)$ (over the original F^+) has a meromorphic continuation and functional equation and is non-vanishing for Re $s \geq 1$.

In other words, all the even-dimensional ρ_E^m with $m \leq n$ are potentially modular and have modular realizations over the same totally real field.

The remainder of this section describes the steps of the proof contained in [HST], but the application of the results of [HST] depend on the modular lifting theorems proved in [CST] and [T]. The same techniques apply to special n-dimensional representations discussed in (1.8):

Theorem 5.1.2. *Let $t \in \mathbb{P}^1(F^+)$, $t \notin \{\mu^{n+1}, \infty\}$, and let $\rho_{t,\ell}$ be the corresponding n-dimensional polarized representation arising in the middle-dimensional cohomology of the Calabi-Yau hypersurface X_t (1.8). Assume $t \notin \mathcal{O}_{F^+}$. Then $\rho_{t,\ell}$ is potentially modular: there exists a totally real Galois extension F'/F^+ and a cuspidal automorphic representation Π_t of $GL(n, F')$ with $L(s, \rho_{t,\ell} \mid_{\Gamma_{F'}}) = L(s, \Pi_t)$. In particular, $L(s, \rho_{t,\ell})$ admits a meromorphic continuation and the usual functional equation.*

Henceforward we place ourselves in the situation of Theorem 5.1.1. The proof of Theorem 5.1.2 is different in that it requires a separate argument, based on a theorem of Larsen and Pink, to prove that for a set of ℓ with Dirichlet density 1, the residual representations $\bar{\rho}_{t,\ell}$ have "big" image, which permits application of the modular lifting theorems of §4.

5.1.3. The condition $j(E) \notin \mathbb{Z}$.

Recall condition (3) on Π:

$$(3)\ \exists\ v_0,\ \Pi_{v_0} \text{ discrete series}$$

$$\Leftrightarrow$$

$$(c)\ \rho_\Pi \text{ satisfies a local condition at } v_0$$

Restriction to condition (c) explains why we need to assume E has multiplicative reduction for Sato-Tate. If ρ_v is a 2-dimensional (Frobenius semi-simple) representation of a local Weil-Deligne group such that $Sym^{n-1}\rho_v$ is indecomposable for all $n > 1$, then $\rho_v = \mathcal{L}(St)$, where St is some Steinberg representation of the local $GL(2)$. A similar consideration explains why $t \notin \mathcal{O}_{F^+}$ in the Calabi-Yau case:

Lemma 5.1.4. *At a place v dividing the denominator of t, if ℓ is prime to v, then $\rho_{t,\ell}(\Gamma_v)$ contains a unipotent element of maximal rank.*

This Lemma is proved in [HST] by comparing the local monodromy with the transcendental monodromy. The latter is the subject of the B-model of interest in mirror symmetry; more importantly, its calculation (due to physicists?) is essential in the proof of the Sato-Tate conjecture, as we see below.

5.2. The idea of potential modularity. How to prove reciprocity?

In Wiles' proof. the prime $\ell = 3$ plays a particularly important role.

The starting point of Wiles' approach: reduce ρ modulo ℓ to obtain a representation

$$\bar{\rho} \ = \ \rho \pmod{\ell} : Gal(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL(n, \mathbb{F}_\ell) \text{ (or } \bar{\mathbb{F}}_\ell)$$

(For Wiles $n = 2$.) Deformation theory provides a first-order classification of possible liftings of $\bar{\rho}$ to characteristic zero in terms of Galois cohomology.

Key Definition. *ρ is* **residually automorphic** *if $\bar{\rho}$ admits one reasonable lifting, say*

$$\tilde{\rho} : Gal(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL(n, \mathcal{O}),$$

$\mathcal{O}$ an ℓ-adic integer ring with residue field $\mathbb{F}_\ell$, such that $\tilde{\rho}$ "comes from" modular forms.

Wiles (following Serre): when $\ell = 3$, there is always at least one automorphic lifting $\tilde{\rho}$. Thus an appropriate modular lifting theorem implies every lifting is automorphic, notably the original ρ. Apart from one potential complication at the end, this completes the argument.

Three steps.

(1) (Taylor-Wiles) Modular Lifting Theorem for "minimal" liftings.
(2) (Wiles level-raising) Modular Lifting Theorem for *all* liftings.
(3) Trick at $\ell = 3$ to get started.

For $n > 2$:

(1) Generalized in [CHT]
(2) An alternative provided in [T]
(3) Impossible to generalize; a substitute in [HST]

Wiles' trick is peculiar to $\ell = 3$, $n = 2$, forms of weight 2. For $n = 2$, Taylor (2002, 2004) found a way to construct an automorphic lifting of any odd $\pmod{\ell}$ representation for general $\ell > 2$ but at the cost of extending the ground field to an unspecified totally real field. This applies to modular forms of higher weight and Hilbert modular forms.. The article [HST] generalizes this to all even n.

Idea: One class of n-dimensional Galois representations is always modular. Let $M/\mathbb{Q}$ be an abelian (cyclotomic) extension of degree n, totally imaginary (so n is even).

$$\chi : M_{\mathbf{A}}^\times / M^\times \to \mathbb{C}^\times$$

an *algebraic Hecke character* which is trivial on the idèles of the totally real subfield $M^+ \subset M$; thus χ is actually a character of the group $U(1)$ relative to M/M^+. Then by Weil (the case n = 1 of Theorem 1.2!) we get a compatible system of characters

$$\chi_\ell : \Gamma_M \to \bar{\mathbb{Q}}_\ell^\times.$$

Induction up to $\Gamma_{\mathbb{Q}}$ yields the monomial representation

$$I(\chi_\ell) : \Gamma_{\mathbb{Q}} \to GL(n, \bar{\mathbb{Q}}_\ell).$$

Because χ comes from $U(1)$, and 1 is odd, one actually needs to twist χ by a half-integral power of the norm, in order to obtain the Galois characters χ_ℓ. With the appropriate choice of χ_∞, we find that $I(\chi_\ell)$ has Hodge-Tate numbers corresponding to $h^{i,n-1-i} = 1$ for $0 \leq i \leq n-1$, as we have been assuming.

Proposition 5.2.1 (Reciprocity for $I(\chi_\ell)$). *The representation $I(\chi_\ell)$ is modular. Moreover, it is attached to a representation $\Pi(\chi)$ satisfying the local discrete series condition (3) at some place v_0, provided v_0 is inert in $\mathcal{K}$ and $\chi\,|_{\Gamma_{v_0}}$ is distinct from all its conjugates under $Gal(\mathcal{K}/\mathbb{Q})$.*

Indeed, for $n = 2$, χ defines a modular form (binary theta function) $\Theta(\chi)$, and $I(\chi_\ell)$ is naturally associated to $\Theta(\chi)$. For general n, use automorphic induction (Kazhdan, Arthur-Clozel) to obtain a self-dual automorphic representation $\Pi(\chi)$ of $GL(n, \mathbb{Q})$, which is associated to $I(\chi_\ell)$ for all ℓ. These automorphic representations satisfy an important property that is unknown for any other kind of cohomological automorphic representation: they admit base change to *any* Galois extension (not only solvable extensions). No substitute is known, and this is why so far we can only prove modularity of even-dimensional symmetric powers.

5.2.2. Deformation to monomial representations.

Suppose you have a compatible system $\{\rho_\ell\}$ and suppose there is an ℓ' such that, for some χ and $\mathcal{K}$ as above,

(Mo) $$\rho_{\ell'} \equiv I(\chi_{\ell'}) \pmod{\ell'}$$

Thus $\bar{\rho}_{\ell'} = \rho_{\ell'} \pmod{\ell'}$ admits at least one modular lifting. Then under the usual restrictions (big residual image, $\ell > n$, etc.) we can apply [CHT] and [T] to conclude that every lifting, in particular $\rho_{\ell'}$ is modular. Since the system is compatible, ρ_ℓ is also modular.

The situation (Mo) is rare and unpredictable.

Taylor's idea: one can often find something weaker: $\exists F/\mathbb{Q}$, a compatible family of Galois representations

$$\sigma_\ell : \Gamma_F \rightarrow GL(n, \bar{\mathbb{Q}}_\ell),$$

and two primes ℓ and ℓ', such that

(A_F) $$\sigma_\ell \equiv \rho_\ell\,|_{\Gamma_F} \pmod{\ell}.$$

(B_F) $$\sigma_{\ell'} \equiv I(\chi_{\ell'})\,|_{\Gamma_F} \pmod{\ell'}.$$

Or written solely in terms of the residual representations:

$$(A_F) \qquad \bar{\sigma}_\ell = \bar{\rho}_\ell \mid_{\Gamma_F} .$$

$$(B_F) \qquad \bar{\sigma}_{\ell'} = \bar{I}(\chi_{\ell'}) \mid_{\Gamma_F} .$$

Now suppose Taylor-Wiles-Kisin applies to n-dim representations of Γ_F.

(1) (B_F) + previous discussion $\Rightarrow \sigma_{\ell'}$ is modular – *over F!*.
(2) σ is a compatible family $\Rightarrow \sigma_\ell$ is also modular (over F),
(3) (A_F) + Wiles/Taylor-Wiles $\Rightarrow \rho_\ell$ is residually automorphic over F, hence ρ_ℓ is modular over F.

As mentioned above, an argument based on Brauer's theorem then implies that $L(s, \rho)$ has at least a meromorphic continuation and functional equation.

The articles [CHT] and [T] develop the Taylor-Wiles-Kisin argument over F, provided

(α) F is totally real, and
(β) ℓ and ℓ' are unramified in F.

Undoubtedly (β) will eventually be removed, but this presumably requires new insight into the p-adic (ℓ-adic?) local Langlands correspondence.

5.3 How to find σ_ℓ?.

The only obvious source of compatible families is in the ℓ-adic cohomology of an algebraic variety over F.

$$\{\sigma\} \leftrightarrow \left\{ \begin{matrix} F\text{-Rational points on a moduli space } \mathcal{M} \\ \text{for some } F \text{ satisfying } (\alpha), (\beta) \end{matrix} \right\}$$

$\mathcal{M}$ parametrizes a certain family of algebraic varieties.

Note that we are deforming in three directions: vertically (lifting from *mod* ℓ to ℓ-adic representations), horizontally (by varying ℓ) and geometrically (over the moduli space $\mathcal{M}$).

Conditions (α) and (β) are local.

5.3.1. Rumely's local-global principle. *If $\mathcal{M}$ is geometrically irreducible and has points locally over $\mathbb{R}$ and over unramified extensions of $\mathbb{Q}_\ell$ and $\mathbb{Q}_{\ell'}$, then it has points over global F satisfying (α) and (β).*

Taylor used a precise version due to Moret-Bailly [MB], though there are versions due to Green-Pop-Roquette and others.

Now recall conditions (a), (b), (c) from the beginning of the talk. We need a $\mathcal{M}$ parametrizing varieties whose cohomology satisfies (a)-(c).

Actually, we will look for $\mathcal{M}$ as a moduli space for *motives*: collections of pieces of cohomology in various theories (topological, de Rham, Hodge, crystalline, ℓ-adic) with comparison maps. We only really want families of ℓ-adic étale cohomology groups, equipped with Galois representations, but with a trace of the other theories (for example, need Hodge-Tate numbers, so regularity can be defined).

By general principles (Griffiths transversality), one can't expect to find continuous families of motives with regular Hodge-Tate numbers if these are too spread out. I only know of one family of motives whose associated ℓ-adic representations are regular has attracted much attention, namely the Calabi-Yau family of $V_{t,\ell} \subset H^{n-1}(X_t, \mathbb{Q}_\ell)^H$, parametrized by $Y = \mathbb{P}^1 - \{\mu^{n+1}, \infty\}$.

As a first approximation to $\mathcal{M}$: for $M \geq 1$, $t \in Y$, n even, let

$$V[M](X_t) := H^{n-1}(X_t, \mathbb{Z}/M\mathbb{Z})^H$$

and let Y_M be the covering of Y parametrizing polarized level N-structures:

$$Y_M = \{\lambda : (\mathbb{Z}/M\mathbb{Z})^n \xrightarrow{\sim} V[M](X_t), t \in Y\}$$

where λ is a symplectic isomorphism. This is étale over Y, in fact over $\mathbb{Z}[\frac{1}{M(n+1)}]$.

Theorem 5.3.2. *There is an integer N_0 such that, if every prime factor of M is larger than N_0, then $Y_M(\mathbb{C})$ is an irreducible étale cover of $Y(\mathbb{C})$ with covering group $Sp(n, \mathbb{Z}/M\mathbb{Z})$.*

This theorem was explained to us by Nick Katz, who gave us the references to his book [Ka]. In [HST] it is derived as a consequence of

(1) the explicit determination of the monodromy of the Gauss-Manin connection for $H^{n-1}(Y, \mathbb{C})^H$ (cf. physics literature). The result is an explicit hypergeometric differential equation of degree n.
(2) Beukers-Heckman: the differential Galois group of this hypergeometric equation is $Sp(n)$.
(3) Matthews, Vaserstein, and Weisfeiler: the monodromy mod ℓ is full for almost all ℓ (also proved by M. Nori).

However, Katz has more recently explained to us that one can replace the final step (which in the version of Matthews, Vaserstein, and Weisfeiler depends on the classification of finite groups) by a more elementary argument. The Beukers-Heckman result implies that the image of monodromy – i.e., the image of the fundamental group of Y in the monodromy representation on the solutions to the Gauss-Manin connection – is Zariski dense in $Sp(n)$. The problem is thus to show that the image of a Zariski dense subgroup upon reduction mod ℓ contains $Sp(n, \mathbb{F}_\ell)$ for sufficiently large ℓ. But we know more than the Zariski density of monodromy. The calculation of the hypergeometric equation is based on the fact that the image of monodromy contains a principal unipotent element of $Sp(n)$, i.e., a unipotent matrix with minimal polynomial $(T-1)^n$. It follows that the image of the Lie algebra of monodromy mod ℓ contains a principal nilpotent element for almost all ℓ. Now the result follows easily Zariski density of monodromy.

Let ℓ, ℓ' be two primes larger than N_0. Fix a CM field M cyclic of degree n over $\mathbb{Q}$, an algebraic Hecke character χ of $\mathbf{A}_M^\times/M^\times$, and $I(\chi\ell')$ as defined above

$$= ind_{\Gamma_M}^{\Gamma_\mathbb{Q}} \chi_{\ell'} \ \rightarrow \ GL(n, \bar{\mathbb{Q}}_{\ell'}).$$

Consider

$$\mathcal{M}_{\ell,\ell'} = \left\{ \begin{array}{c} \text{C-Y hypersurfaces } X \text{ in the Dwork pencil} \\ \text{with polarization-preserving isomorphisms} \end{array} \right\}$$

$$V[\ell](X) \xrightarrow{\sim} \rho^n_{E,\ell};\ V[\ell'](X) \xrightarrow{\sim} \bar{I}(\chi_{\ell'}).$$

By Theorem 5.3.2 $\mathcal{M}_{\ell,\ell'}$ is geometrically irreducible.

Theorem 5.3.3. Local-global principle (cf. [MB]). *Let $S = S_1 \coprod S_2$ be a finite set of places of $\mathbb{Q}$, with $\infty \in S_1$. Let $\mathcal{M}/\mathbb{Q}$ be geometrically irreducible, and assume*

(*) *For all $v \in S_1$ (resp. $w \in S_2$) the set $\mathcal{M}(\mathbb{Q}_v)$ (resp. $\mathcal{M}(\mathbb{Q}_w^{unr})$) is non-empty.*

Then there is a finite Galois extension $F_1/\mathbb{Q}$ in which all $v \in S_1$ split completely – in particular, F_1 is totally real – and which is unramified at all $w \in S_2$, such that $\mathcal{M}(F_1) \neq \emptyset$. Moreover, if L is any fixed finite extension of $\mathbb{Q}$, we can assume F_1 and L are linearly disjoint over $\mathbb{Q}$.

Recall Wiles' three steps:

(1) (Taylor-Wiles) Modular Lifting Theorem for "minimal" liftings.
(2) (Wiles level-raising) Modular Lifting Theorem for *all* liftings.
(3) Trick at $\ell = 3$ to get started.

If we can verify condition (*) for appropriate sets S_1 and S_2, we will find a point $t_1 \in \mathcal{M}_{\ell,\ell'}(F_1)$ for which $V[\ell'](X_{t_1})$ is *monomial.* This is our substitute for Wiles' trick at $\ell = 3$.

Then we can argue exactly as for $n = 2$. Our point $t_1 \in \mathcal{M}_{\ell,\ell'}(F_1)$ provides a compatible family

$$\sigma_q : \Gamma_{F_1} \to GSp(n, \overline{\mathbb{Q}}_q)$$

with

$$(A_{F_1}) \qquad \bar{\sigma}_\ell = \bar{\rho}^n_{E,\ell} \mid_{\Gamma_{F_1}} .$$

$$(B_{F_1}) \qquad \bar{\sigma}_{\ell'} = \bar{I}(\chi\ell') \mid_{\Gamma_{F_1}} .$$

Assume we have generalized modular lifting theorems under the conditions already seen for $n = 2$.

(α) F is totally real, and
(β) ℓ and ℓ' are unramified in F.

We need $\infty \in S_1$ for condition (α), $\ell, \ell' \in S_2$ for condition (β). In the absence of these two conditions, nothing is known. Then as before

(1) (B_{F_1}) + modular lifting thm. $\Rightarrow \sigma_{\ell'}$ is automorphic – *over* F_1.
(2) σ is a compatible family $\Rightarrow \sigma_\ell$ is also modular (over F_1),
(3) (A_{F_1}) + modular lifting thm. $\Rightarrow \rho^n_{E,\ell}$ is residually automorphic over F_1, hence $\rho^n_{E,\ell}$ is automorphic over F_1.

As mentioned before, proof of (3) for all even n suffices for meromorphic continuation of all symmetric powers, and for the Sato-Tate conjecture.

5.4. The fine print.

The existence of points of $\mathcal{M}_{\ell,\ell'}$ over unramified extensions of $\mathbb{Q}_\ell$ and $\mathbb{Q}_{\ell'}$ is non-trivial, and it is also insufficient to complete the proof.

5.4.1. Existence at ℓ, ℓ', and ∞ is insufficient – there are other primes. I start with the latter point, which has to do with the condition (c) that has been plaguing us since the beginning. Remember that when we introduced the Hecke character χ we guaranteed condition (c) by requiring that $I(\chi_\ell)$ be irreducible locally at a place

v_0 inert in M. In order to guarantee that $I(\chi_\ell)$ remains not only modular over F_1 but also modular and satisfying condition (c), we require v_0 to split completely in F_1.

So far, so good. However, $\bar{\sigma}_{\ell'}$ has to be isomorphic to some $V[\ell](X_t)$, and there is no reason to expect $V[\ell](X_t)$, which is the cohomology of a specific kind of variety, to have local representations with the required kind of ramification at v_0. So in order to preserve condition (B_{F_1}), we would have to allow v_0 to be highly ramified in F_1 – but then we lose condition (c).

In the (not so) long run, we should be able to scrap condition (c). In the meantime, [CHT] proves the existence of ℓ' (enough of them) for which $\bar{I}(\chi_\ell)$ admits a lift of Steinberg type at some place q'. This argument uses a generalization of Ramakrishna's level-raising theorem, and is the most technical point in [CHT]. Together with a descent and base change argument, this implies a quite general modular lifting theorem for $\bar{I}(\chi_\ell)$ that remains in the framework of condition (3) – but only for very large ℓ (on the order of $(2n)^{n/2}$).

5.4.2. Existence at ℓ and ℓ' is non-trivial. I'll consider the problem for ℓ, that for ℓ' being somewhat similar. Our condition is

$$(A_{F_1}) \qquad V[\ell](X_t)\ |_{I_\ell} \xrightarrow{\sim} \bar{\rho}^n_{E,\ell}\,|_{I_\ell}\ .$$

where I_ℓ is the inertia subgroup of $Gal(\bar{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)$. Now E can either be supersingular or ordinary at ℓ. I don't know whether or not it is impossible for the $\mathbb{Z}/\ell\mathbb{Z}$-cohomology of some X_t, with $t \in Y(\mathbb{Q}_\ell{}^{unr})$, to contain an n-dimensional piece that looks like a symmetric power of a supersingular elliptic curve, but it's more prudent to assume ℓ is an ordinary prime for E. This is a fateful assumption: it is known that the set of ordinary primes of an elliptic curve over a number field has Dirichlet density one, but nothing of the kind is known for modular forms of weight $k > 2$, and this limits the application of our method to forms of weight 2 for the moment.

If $E[\ell]$ is ordinary, then the action of I_ℓ on $E[\ell]$ is upper triangular, so $\bar{\rho}^n_{E,\ell}$ is concentrated along the diagonal and first superdiagonal:

$$\begin{pmatrix} 1 & \bullet & 0 & 0 & \dots & 0 & 0 \\ 0 & \omega_\ell^{-1} & \bullet & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \omega_\ell^{2-n} & \bullet \\ 0 & 0 & 0 & 0 & \dots & 0 & \omega_\ell^{1-n} \end{pmatrix}$$

If X_t is an ordinary hypersurface, which we can assume for the sake of argument, then the action of I_ℓ on $V[\ell](X_t)$ is upper-triangular with the same diagonal entries but there is no reason for the entries above the diagonal to be concentrated on one superdiagonal.

The solution is to assume I_ℓ acts diagonally on $E[\ell]$. Since this is not always possible we introduce a new prime ℓ'' and a new elliptic curve E' with $E'[\ell''] \xrightarrow{\sim} E[\ell'']$ as Galois modules and $E'[\ell]$ diagonal for I_ℓ (a canonical lift at ℓ). This requires an additional application of Moret-Bailly's theorem, but to simplify we assume $E[\ell]$ is already diagonal. This means we need to find X_t with the same property. Fortunately, one is staring us in the face: the Fermat hypersurface, with $t = 0$, whose cohomology is diagonal over I_ℓ as long as ℓ splits completely in $\mathbb{Q}(\mu_{n+1})$ ($\ell \equiv 1 \pmod{n+1}$). Note that in the usual study of hypergeometric functions, 0 is a singular point, as it is in the usual study of moduli of the Dwork family. This is why our moduli space is Y, a cyclic cover of $\mathbb{P}^1 - \{1, \infty\}$ with Galois group μ_{n+1}, rather than the usual moduli space $\mathbb{P}^1 - \{0, 1, \infty\}$.

Once we have a diagonal action of I_ℓ on $E[\ell]$, with powers of the cyclotomic character along the diagonal, the set of points lying above some ℓ-adic neighborhood of 0 in $Y(\mathbb{Q}_\ell{}^{unr})$ provides points of $\mathcal{M}_{\ell,\ell'}$ over $\mathbb{Q}_\ell{}^{unr}$, provided $\ell \equiv 1 \pmod{n+1}$. The argument for ℓ' is similar, except that one needs ℓ' split in the cyclic CM field M as well; for good measure, we take ℓ split in M, to avoid conflicts with condition (B_{F_1}) at ℓ.

5.4.3 Wrapping up. We also need to keep track of at least one of the primes q dividing the denominator of $j(E)$, to make sure it splits completely in F_1, in order to remain in the range of applicability of condition (c). We need moreover that the local monodromy at q – and also at the q' introduced in (5.4.1) – is non-trivial mod ℓ and ℓ', which forces some of the ℓ's to be big relative to q and q' at some stages in the argument. Most, if not all, of these complications, would disappear if one could scrap condition (c).

References

[BR] D. Blasius and J. Rogawski, Motives for Hilbert modular forms, *Inventiones Math.*, **114** (1993) 55-87.

[C] L. Clozel, Représentations Galoisiennes associées aux représentations automorphes autoduales de GL(n), *Publ. Math. I.H.E.S.*, **73**, 97-145 (1991).

[CHT] L. Clozel, M. Harris, and R. Taylor, Automorphy for some ℓ-adic lifts of automorphic mod ℓ Galois representations, preprint (2005)

[CL] L. Clozel and J.-P. Labesse, Changement de base pour les représentations cohomologiques de certains groupes unitaires, in [L], 119-133.

[HL] M. Harris and J.-P. Labesse, Conditional base change for unitary groups, *Asian J. Math*, **8** (2004) 653-684.

[HST] M. Harris, N. Shepherd-Barron, and R. Taylor, A family of Calabi-Yau varieties and potential automorphy, preprint (2005, 2006)

[HT] M. Harris and R. Taylor, The geometry and cohomology of some simple Shimura varieties, Annals of Math. Studies, **151** (2001).

[L] J.-P. Labesse, *Cohomologie, stabilisation et changement de base*, *Astérisque*, **257** (1999).

[Ka] N. M. Katz, *Exponential Sums and Differential Equations*, *Annals of Math. Studies*, **124** Princeton: Princeton University Press (1990).

[K] M. Kisin, Moduli of finite flat group schemes and modularity, preprint (2004).

[Ko] R. Kottwitz, On the λ-adic representations associated to some simple Shimura varieties, *Invent. Math.*, **108**, 653-665 (1992).

[MB] L. Moret-Bailly, Groupes de Picard et problèmes de Skolem II, *Ann. scient. Ec. Norm. Sup.*, **22** (1989) 181–194.

[MS] D. Mumford and K. Suominen, Introduction to the theory of moduli, in F. Oort, ed, *Algebraic Geometry, Oslo 1970*, Groningen: Wolters-Noordhoff (1972) 171-222.

[T] R. Taylor, Automorphy for some ℓ-adic lifts of automorphic mod ℓ Galois representations, II, preprint (2006).

[T02] R. Taylor, Remarks on a conjecture of Fontaine and Mazur, *J. Inst. Math. Jussieu*, **1** (2002) 1-19.

[TY] R. Taylor and T. Yoshida, Compatibility of local and global Langlands correspondences, *J. Am. Math. Soc.*, in press.

[V] M.-F. Vignéras, Induced R-representations of p-adic reductive groups, *Selecta Math.*, **4** (1998), 549-623.