GALOIS REPRESENTATIONS, AUTOMORPHIC FORMS, AND THE SATO-TATE CONJECTURE

Michael Harris

UFR de Mathématiques
Université Paris 7
2 Pl. Jussieu
75251 Paris cedex 05, FRANCE

1. Introduction

An elliptic curve is the set E of solutions of a cubic curve in two variables, for example

$$E: \ y^2 + y = x^3 + x.$$

I will only consider elliptic curves with rational coefficients, which after a change of variables can be written

$$y^2 = x^3 + Ax + B.$$

As abstract algebraic curves, these are not all distinct, and one can isolate two invariants: the *discriminant*

$$\Delta_E = -16(4A^3 + 27B^2)$$

which is not really an invariant of E, but which has the following property: if $\Delta_E \neq 0$ then E is non-singular, which we always assume. There is also the j-invariant, which really depends on E and not just on the equation:

$$j(E) = 1728 \cdot \frac{4A^3}{4A^3 + 27B^2}.$$

The quantity $j(E)$ determines E up to isomorphism over an algebraically closed field.

Institut des Mathématiques de Jussieu, U.M.R. 7586 du CNRS. Membre, Institut Universitaire de France.

Without (much) loss of generality, we may assume $A, B \in \mathbb{Z}$. It then makes sense to reduce the equation modulo a prime p and ask how many solutions E has modulo p:

$$N_p(E) = |E(\mathbb{F}_p)|.$$

Suppose for the moment we replace E by a line L, given by a *linear* equation

$$L: \ y = ax + b.$$

Then the number of solutions of L in the plane ${\mathbb{F}_p}^2$ obviously equals p, to which we add 1 for the point at infinity:

$$|L(\mathbb{F}_p)| = p + 1.$$

It turns out that $p+1$ is in a natural sense the optimal number of points for a curve of *any* genus (or degree). Skipping over quadric curves, we define an integer $a_p(E)$, for each prime p, by

$$N_p(E) = p + 1 - a_p(E).$$

We only consider p for which E remains nonsingular modulo p, which is somewhat weaker than the condition that $\Delta_E \neq 0 \pmod{p}$. Such a p is called a *prime of good reduction.*

One can date the beginning of arithmetic algebraic geometry to Hasse's discovery that

$$|a_p(E)| \leq 2\sqrt{p}$$

for any prime of good reduction. In other words, $p + 1$ is a good approximation to $N_p(E)$ to square-root order. This can be compared to the square-root good approximation to $\pi(x)$, the number of primes less than x:

$$\pi(x) = \int_2^x \frac{dx}{logx} + \ \mathrm{Error}(x)$$

where the Riemann hypothesis is the assertion that

$$\mathrm{Error}(x) = O(x^{\frac{1}{2}})$$

and indeed Hasse's theorem was generalized by Weil to a version of the Riemann hypothesis valid for all curves over finite fields.

The next question is whether anything can be said about the behavior of the $a_p(E)$ as p varies. Is $a_p(E)$ more likely to be positive or negative? Is it more likely to cluster around 0 or around $\pm 2\sqrt{p}$? The rough answer is that it is as random as possible, but it is not immediately obvious how to make sense of this. We normalize all the $a_p(E)$ simultaneously to allow them to be compared:

$$a_p^{norm}(E) = \frac{1}{2\sqrt{p}} a_p(E) \in [-1, 1].$$

Thus there is a unique $\theta_p = \theta_p(E) \in [0, \pi]$ such that $a_p^{norm}(E) = cos(\theta_p)$. We ask about the distribution of the a_p^{norm} in $[-1, 1]$, or equivalently of the $\theta_p \in [0, \pi]$. Over forty years ago, Sato and Tate independently formulated the following conjecture:

Sato-Tate Conjecture. *Suppose E has no complex multiplication. Then the $a_p^{norm}(E)$ (resp. the θ_p) are equidistributed in $[-1,1]$ (resp. $[0,\pi]$) with respect to the probability measure*

$$\frac{2}{\pi}\sqrt{1-t^2}dt \quad (\text{ resp. } \frac{2}{\pi}\sin^2(\theta)d\theta \text{ }) .$$

Regarding the initial hypothesis most E have no complex multiplication. In particular, if $j(E) \in \mathbb{Q} - \mathbb{Z}$, then $j(E)$ has no complex multiplication. In this setting the conjecture was proved in 2006:

Theorem 1.1 (Clozel, Harris, Shepherd-Barron, Taylor)[1]. *Suppose $j(E)$ is not an integer. Then the Sato-Tate Conjecture is valid for E.*

For E with complex multiplication the distribution of the $a_p^{norm}(E)$ is different: for half of p, $a_p(E) = 0$. The distribution has a more natural description in terms of the quadratic field defining the complex multiplication, and is an easy consequence of class field theory.

Serre explained in his 1968 notes on elliptic curves [S1] how to derive the Sato-Tate conjecture from Tauberian theorems of the sort used to prove the prime number theorem. The main input is the following. One needs to rewrite the expression for $N_p(E)$:

$$N_p(E) = (1 - \sqrt{p}e^{i\theta_p})(1 - \sqrt{p}e^{-i\theta_p}) = (1 - \alpha_p)(1 - \beta_p)$$

and to define more generally

$$L_p(s,E) = [(1 - \alpha_p p^{-s})(1 - \beta_p p^{-s})]^{-1};\ N_p(E) = L_p(0,E)^{-1} \tag{1.2}$$

Let S be the set of primes of bad reduction for E, and define

$$L(s,E) = \prod_{p \notin S} L_p(s,E) \times \prod_{p \in S} L_p(s,E)$$

with explicit factors $L_p(s,E)$ for $p \in S$, simpler than those for primes of good reduction.

It follows from Hasse's estimate that $L(s,E)$ converges absolutely for $Re(s) > \frac{3}{2}$. Now one can attach a very similar Dirichlet series to a holomorphic modular (cusp) form f of weight 2. This is a holomorphic function on the upper half plane $\{x+iy \in \mathbb{C} \mid y > 0\}$ that satisfies the following functional equation:

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

[1]Some references, e.g. [G, p. 347], have chosen to save space by offering an abbreviated version of this list.

for all z and all $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ belonging to an appropriate subgroup $\Gamma \subset SL(2,\mathbb{Z})$ of finite index. Such a modular cusp form is also assumed to satisfy an appropriate growth condition that implies that it admits a Fourier expansion

$$f(z) = \sum_{n=1}^{\infty} a_n(f) q^n,\ q = e^{2\pi i z}$$

whose corresponding Dirichlet series is defined by

$$L(s,f) = \sum_{n=1}^{\infty} b_n n^{-s}.$$

It is known that $L(s,f)$ extends to an entire analytic function that satisfies a functional equation relating $L(s,f)$ to $L(2-s,f)$, and such that $L(s,f) \neq 0$ for $Re(s) \geq \frac{3}{2}$. The most striking development of number theory in recent years was Wiles' discovery of a technique for proving that any $L(s,E)$ is also an $L(s,f)$, specifically

$$L(s,E) = L(s,f_E),\ f_E = q + \sum_{n \geq 2} a_n(E) q^n \tag{1.3}$$

where when $n = p$ is a prime not in S, the coefficient $a_n(E)$ is the $a_p(E)$ defined above.

With help from Taylor, Wiles applied this technique to a sufficiently large family of E to prove Fermat's Last Theorem. A few years later, Taylor, together with Breuil, Conrad, and Diamond, proved that every $L(s,E)$ is an $L(s,f)$. In particular,

Theorem [BCDT]. *L(s,E) extends to an entire analytic function with no zeroes on the half-plane $Re(s) \geq \frac{3}{2}$.*

From the information contained in the $N_p(E)$ one can construct an infinite family of L-functions. For each $n \geq 0$, define

$$L_p(s,E,Sym^n) = [\prod_{j=0}^{n} (1 - \alpha_p^j \beta_p^{n-j} p^{-s})]^{-1}$$

if $p \notin S$; I again omit the definition for $p \in S$. We define

$$L(s,E,Sym^n) = \prod_p L_p(s,E,Sym^n).$$

Thus for $n = 0$ we find the Riemann zeta function and for $n = 1$ we have $L(s,E)$. Hasse's estimates imply that $L(s,E,Sym^n)$ converges absolutely for $Re(s) > 1 + \frac{n}{2}$.

Theorem 1.4 (Serre, [S1]). *Suppose E is an elliptic curve and, for all $n > 0$, $L(s,E,Sym^n)$ extends to a meromorphic function that is holomorphic and non-vanishing for $Re(s) \geq 1 + \frac{n}{2}$. Then the Sato-Tate Conjecture holds for E.*

The results of the three papers [CHT], [HST], and [T] together imply this sufficient condition:

Theorem 1.5 (Clozel, Harris, Shepherd-Barron, Taylor). *Suppose E is an elliptic curve over $\mathbb{Q}$ with non-integral j-invariant. Then for all $n > 0$, $L(s, E, Sym^n)$ extends to a meromorphic function that is holomorphic and non-vanishing for $Re(s) \geq 1 + \frac{n}{2}$.*

The next section reviews the proof of Theorem 1.4 in the language of Galois representations, the perspective emphasized in [S1].

2. Equidistribution

2.1. Galois representations and associated L functions. Let F^+ be a totally real field. Those who prefer can assume $F^+ = \mathbb{Q}$.

Let E be an elliptic curve over F^+. To E we associate a 2-dimensional ℓ-adic representation for any prime ℓ: 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. Assume $F^+ = \mathbb{Q}$ for the time being. Then this representation encodes all information about $|E(\mathbb{F}_p)|$ for almost all p, in the following sense. Remember that in the previous section we mentioned primes of good reduction. Suppose p is a prime of good reduction for E, and suppose also $p \neq \ell$. Then we can recover $a_p(E)$ from $\rho_{E,\ell}$. Let $Frob_p \in Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ be a Frobenius element for p. This is defined at the beginning of a course in algebraic number theory. We know that $Gal(\overline{\mathbb{F}}_p/\mathbb{F}_p)$ is generated by an element ϕ_p with the property that,

$$\forall x \in \overline{\mathbb{F}}_p, \phi_p(x) = x^p.$$

For technical reasons, we prefer to work with $Frob_p = \phi_p^{-1}$. This is an element of $Gal(\overline{\mathbb{F}}_p/\mathbb{F}_p)$, but if we extend the p-adic valuation on $\mathbb{Q}$ to a valuation v on $\overline{\mathbb{Q}}$, the decomposition subgroup $\Gamma_v \subset Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ fixing v is isomorphic to $Gal(\bar{\mathbb{Q}}_p/\mathbb{Q}_p)$. Since E has good reduction at p and $p \neq \ell$, the representation $\rho_{E,\ell}$ is *unramified* at p, which means in particular that it is trivial on the inertia subgroup $I_v \subset \Gamma_v$, hence factors through $\Gamma_v/I_v \simeq Gal(\overline{\mathbb{F}}_p/\mathbb{F}_p)$. Thus we can define $\rho_{E,\ell}(Frob_p)$. This depends on the choice of extension v of the p-adic valuation, but any two extensions are conjugate by an element of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$. In particular, the characteristic polynomial

$$P_{p,E}(T) = det(I - \rho_{E,\ell}(Frob_p)T) \in \mathbb{Q}_\ell[T] \tag{2.1.1}$$

depends only on p. Moreover, it is well known that $P_{p,E}(T)$ has coefficients in $\mathbb{Q}$ and is independent of $\ell \neq p$. Thus the complex function $P_{p,E}(p^{-s})$ is well defined for all primes of good reduction. In fact, we know that

$$P_{p,E}(p^{-s}) = 1 - a_p(E)p^{-s} + p^{1-2s}.$$

Let

$$L_p(E, s) = P_{p,E}(p^{-\frac{1}{2}-s})^{-1} = 1 - a_p^{norm}(E)p^{-s} + p^{-2s}.$$

The complex L-function of E is

$$L(s, E) = \prod_p L_p(E, s), \tag{2.1.2}$$

where for primes of bad reduction one has another definition of $L_p(E,s)$. With our chosen normalization, this function converges absolutely for $Re(s) > 1$.

A general conjecture is that $L(s,E)$ extends to an entire function and satisfies a functional equation. This is known for $F^+ = \mathbb{Q}$ (Wiles, [BCDT]) and meromorphic continuation is known for general totally real F^+ [TFM]. One proves that $L(s,E)$ is an entire function, for $F^+ = \mathbb{Q}$, by proving that it is the L-function of the modular form f_E defined in (1.3) . For more general F^+, one wants to prove that it is the L-function of a *cuspidal automorphic representation* of $GL(2,F^+)$ (cf. the Appendix for more details). In that case one says that E is *automorphic* over F^+.

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. For $n \geq 1$ let

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

With $\rho = \rho^n_{E,\ell}$, we define

$$L(s,\rho) = \prod_v L_v(s,\rho) \tag{2.1.3}$$

where v runs over finite places of the field F^+ and with the local factors defined for almost all v by analogy with (2.1.1):

$$L_v(s,\rho) = \det(I - \rho(Frob_v)Nv^{-s})^{-1}. \tag{2.1.4}$$

Here Nv is the number of elements in the residue field of the place v (so $Nv = p$ if v is the rational prime p) and ρ can be taken to be any representation for which the coefficients of the characteristic polynomials of Frobenius elements can be identified with complex numbers.

For almost all p, the elliptic curve E has good reduction at p, which means that the local factor $\Pi_{E,p}$ is unramified. This representation is determined by its *Satake parameters* α_p, β_p, an unordered pair of complex numbers that can be expressed explicitly in terms of the number of points modulo p:

$$a_p(E) = p^{\frac{1}{2}}(\alpha_p + \beta_p),\ \alpha_p \cdot \beta_p = 1.$$

Hasse's theorem mentioned above has a meaning in terms of Π_E.

Hasse, Eichler-Shimura ("Ramanujan conjecture").

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

Up to permutation we have $\alpha_p = e^{i\theta_p}, \beta_p = e^{-i\theta_p}$, say, with $0 \leq \theta_p \leq \pi$. I restate the Sato-Tate Conjecture:

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 $[0,\pi]$. The conjecture makes sense for the automorphic representation Π_E, without reference to elliptic curves, and also makes sense for modular forms of higher weight.

Let $X = [0,\pi]$. For any $f \in C(X)$ and $x > 0$ define

$$S(f,x) = \sum_{p \le x} f(\theta_p).$$

The Sato-Tate conjecture asserts the following: for any continuous function $f \in C(X)$, we have

$$(*) \qquad \lim_{x\to\infty} \frac{S(f,x)}{S(1,x)} = \lim_{x\to\infty} \frac{\sum_{p\le x} f(\theta_p)}{\sum_{p\le x} 1} = \int_X f(\theta) dST(\theta).$$

Now the diagonal matrix $diag(\alpha_p, \beta_p)$ belongs to $SU(2)$. There is an obvious map $\phi : SU(2) \to X$ identifying X with the space of conjugacy classes in $SU(2)$, and $dST(\theta)$ is the image with respect to ϕ of the Haar measure on $SU(2)$ with total mass 1. It suffices to prove (*) for f in an orthogonal basis of $L_2(X)$. Such an orthogonal basis is given by the characters χ_n of the irreducible representations Sym^n of $SU(2)$. For $f = \chi_0$, which is the trivial representation, (*) is obvious. For $f = \chi_n$ with $n > 0$, we have

$$\int_X \chi_n(\theta) dST(\theta) = \int_X \chi_n(\theta) \cdot 1 \; dST(\theta) =< \chi_n, \chi_0 >= 0$$

because the characters form an orthogonal basis.

In general

$$\chi_n(\theta_p) = \sum_{j=0}^{n} \alpha_p^j \beta_p^{n-j}.$$

This is why it is convenient to use the Satake parameters of $\Pi_{E,p}$. So we need to show

$$(2.2) \qquad \lim_{x\to\infty} \sum_{p\le x} \chi_n(\theta_p) = o(\pi(x)).$$

Now we use a standard argument from analytic number theory. With $\rho_E = \rho_{E,\ell}$ (any ℓ), let

$$L^*(s, E, Sym^n) = L^*(s, \rho_E^{n+1}) = L(s + \frac{n}{2}, \rho_E^{n+1})$$

normalized to be absolutely convergent for $Re(s) > 1$. In other words,

$$L^*(s, E, Sym^n) = \prod_p L_p^*(s, \rho_E^{n+1})$$

where for $p \notin S$,

$$L_p^*(s, \rho_E^{n+1}) = \prod_{j=0}^{n} (1 - \alpha_p^j \beta_p^{n-j} p^{-s})^{-1}.$$

Comparing this with (1.1), we find

$$\begin{aligned} \frac{d}{ds} log(L^*(s, E, Sym^n)) &= -\sum_p \sum_m \frac{\chi_n(\theta_p^m) log\ p}{p^{ms}} \\ &= -\sum_p \frac{\chi_n(\theta_p) log\ p}{p^s} + \varphi(s) \end{aligned}$$

where $\varphi(s)$ is holomorphic for $Re(s) > \frac{1}{2}$ and the first equality is only up to a finite set of bad factors which are irrelevant for the second equality.

Let $L^*(s)$ be a Dirichlet series absolutely convergent for $Re(s) > 1$. We say $L^*(s)$ is *invertible* if it extends to a meromorphic function on $\mathbb{C}$ and if $L^*(s)$ has no zeroes for $Re(s) \geq 1$ and no poles for $Re(s) \geq 1$ except for a possible pole at $s = 1$.

Suppose E is an elliptic curve over $\mathbb{Q}$ with non-integral j-invariant. Then Theorem 1.5 [CHT, T, HST] implies that, for all $n \geq 0$, the function $L^*(s, E, Sym^n)$ is invertible and has no pole at $s = 1$ unless $n = 0$. Thus for each $n > 0$,

$$\frac{d}{ds} log(L^*(s, E, Sym^n)) = L^{*,\prime}(s, E, Sym^n)/L^*(s, E, Sym^n)$$

is a quotient of meromorphic functions that are holomorphic and non-vanishing for $Re(s) \geq 1$.

Corollary. *Under the hypotheses of Theorem 1.5,* $\sum_p \frac{\chi_n(\theta_p) log p}{p^s}$ *has no pole for* $Re(s) \geq 1$.

The Wiener-Ikehara tauberian theorem states that if $D(s) = \sum_i \frac{b_i}{i^s}$ is a Dirichlet series convergent for $Re(s) > 1$ and non-singular except for a possible first-order pole at $s = 1$, with residue α, then

$$\sum_{i<x} b_i = \alpha \cdot x + o(x).$$

For the prime number theorem, this is applied with $b_i = p \cdot log\ p$ if $i = p$ is prime, $b_i = 0$ otherwise, to yield a form of the prime number theorem:

$$\sum_{p<x} log\ p = x + o(x).$$

For $n > 0$, Theorem 2.3 implies

$$\sum_{p \leq x} \chi_n(\theta_p) log\ p = o(x).$$

Applying Abel summation to get rid of the logs, we find

$$S(\chi_n, x) = \sum_{p \leq x} \chi_n(\theta_p) = o(x/logx)$$

and since

$$S(1, x) = S(\chi_0, x) = \sum_{p \leq x} 1 = x/logx + o(x/logx)$$

by the prime number theorem we have

$$\lim_{x \to \infty} \frac{S(\chi_n, x)}{S(1, x)} = 0$$

for all $n > 1$. This yields the estimate (2.1), and hence equidistribution.

Two elliptic curves.

One can define two elliptic curves E and E' to be *isogenous* if they are related by a non-trivial group homomorphism; then $a_p(E) = a_p(E')$ for almost all p, and this can be taken to be the definition of isogeny by a very deep theorem of Faltings (another conjecture of Tate). Now suppose E and E' are two elliptic curves over $\mathbb{Q}$. One can consider the pair

$$(a_p^{norm}(E), a_p^{norm}(E')) \subset [-1, 1] \times [-1, 1].$$

If E and E' are isogenous, this point always lies on the diagonal. If not, the conjecture is the following, a special case of a completely general Sato-Tate conjecture for motives, formulated by Serre in [S2]:

Conjecture 2.3. *Let F^+ be a totally real field, let E and E' be elliptic curves over F^+, and assume E and E' do not become isogenous over an abelian extension of F^+. For any prime v of F^+ where E and E' both have good reduction, we let*

$$|E(k_v)| = (1 - q_v^{\frac{1}{2}} e^{i\phi_v})(1 - q_v^{\frac{1}{2}} e^{-i\phi_v})$$

$$|E'(k_v)| = (1 - q_v^{\frac{1}{2}} e^{i\psi_v})(1 - q_v^{\frac{1}{2}} e^{-i\psi_v})$$

where $\phi_v, \psi_v \in [0, \pi]$.

Then the pairs $(\phi_v, \psi_v) \in [0, \pi] \times [0, \pi]$ are uniformly distributed with respect to the measure

$$\frac{4}{\pi^2} sin^2\phi \; sin^2\psi \; d\phi d\psi.$$

Now if we have two non-isogenous elliptic curves E and E' as above, we can form

$$L^*(s, \rho_E^n \otimes \rho_{E'}^m) = \prod_p L_p^*(s, \rho_E^n \otimes \rho_{E'}^m)$$

where for $p \notin S$

$$L_p^*(s, \rho_E^n \otimes \rho_{E'}^m) = \prod_{j=0}^{n} \prod_{k=0}^{m} (1 - \alpha_p^j \beta_p^{n-j} (\alpha_p')^k (\beta_p')^{m-k} p^{-s})^{-1},$$

an expression easier to understand as the determinant of a certain tensor product matrix. The above conjecture follows from

(Conditional) Theorem 2.4. *Assume the expected results of the "book project" ([Book1] and subsequent books). For all* $(m,n) \neq (0,0)$ $L^*(s, \rho_E^n \otimes \rho_{E'}^m)$ *is invertible and has no pole at* $s = 1$.

A proof of this theorem is given in [H] but it is conditional on work in progress, including notably the results of [L], [CHL1], [CHL2], [CH], and [Shin]. It is a consequence of the following result, also proved conditionally in [H]:

(Conditional) Theorem 2.5. *Assume the expected results of the "book project". For every* $m, n \geq 1$, *there is a totally real Galois extension* $F_{m,n}/\mathbb{Q}$ *such that the L-function of* $\rho^n_{E,F_{m,n}} = \rho^n_E \mid_{Gal(\overline{\mathbb{Q}}/F_{m,n})}$ *(resp.* $\rho^m_{E',F_{m,n}}$*) is the L-function of a cuspidal automorphic representation* $\Pi_{E,n}$ *of* $GL(n, F_{m,n})$ *(resp.* $\Pi_{E',m}$ *of* $GL(m, F_{m,n})$*).*

Unconditionally, it is proved in [CHT], [HST], [T] that this is true when m and n are both **even** (or when either $m = 1$ or $n = 1$). If we admit Theorem 2.5, then Theorem 2.4 follows from properties of Rankin-Selberg L-functions and an argument using Brauer's theorem on finite group characters first applied by Taylor. See the discussion in §6.

Three or more elliptic curves? For three elliptic curves we would need something like

$$L^*(s, \rho_E^n \otimes \rho_{E'}^m \otimes \rho_{E''}^r).$$

Unfortunately, at present there is no analytic theory of such L-functions. This is a major barrier to the further development of automorphic forms. The case of triples is in some sense the crucial case.

The error in the error. Barry Mazur has been interested in the question of the discrepancy between the Sato-Tate distribution and the actual distributions of the $a_p(E)$. The rate of convergence to the Sato-Tate distribution of the pointwise distributions, for $p < X$ is a statistical problem whose optimal solution is intimately related to the Generalized Riemann Hypothesis for the L-functions $L^*(s, \rho_E^n)$, cf. [M2].

Brief synopsis. The Langlands conjectures predict that $L(s, E, Sym^{n-1})$ can be associated to a cuspidal automorphic representation of $GL(n)_{\mathbb{Q}}$. This would imply that $L(s, E, Sym^{n-1})$ is entire (Godement-Jacquet) and is non-vanishing on the indicated domain. We do not prove this. Instead, we prove that for n even, $L(s, E, Sym^{n-1})$ is *potentially automorphic*; that is, it is associated to a cuspidal automorphic representation of $GL(n)$ over some totally real Galois extension of $\mathbb{Q}$. This argument involves two parts. The first is an extension of Wiles' technique for identifying L-functions of elliptic curves with L-functions of modular forms, and is based essentially on Galois cohomology and an analysis of automorphic representations of different sorts of groups, especially unitary groups. This is begun in [CHT] and completed in [T], and is described in §3 and §4, below. The second is an extension of an idea used by Taylor to prove meromorphic continuation of L-functions attached to two-dimensional Galois representations, using weak approximation on

moduli spaces. In [HST], we found a moduli space that could be used to study n-dimensional Galois representations for any even n; it is a twisted form of the moduli space of certain Calabi-Yau varieties originally studied by Dwork in certain cases, and more generally by physicists interested in mirror symmetry. These results and their applications to potential automorphy are explained in §5.

It is explained in §6 how potential automorphy of $L(s, E, Sym^{n-1})$ for all even n implies Theorem 2.3 for all n. As we have already seen, this analytic property of the L-functions implies the Sato-Tate conjecture for the elliptic curve E.

3. Deformation rings of Galois representations

For the next few sections we will forget about everything connected with elliptic curves and retain only the Galois representation ρ_E^n, which we denote simply ρ. For the moment ρ is a representation of $\Gamma_F = Gal(\overline{\mathbb{Q}}/F)$ for an arbitrary number field F. In fact, we will only retain a few of its properties. Representations like ρ given by the action of Γ_F on the ℓ-adic cohomology of a (smooth, proper) algebraic variety[2] may be called *motivic*. The general machinery of ℓ-adic cohomology implies the following important properties:

(1) There is a finite set S of primes of F such that, for all primes $v \notin S$, the restriction of ρ to the local Galois group $\Gamma_v = Gal(\bar{F}_v/F_v)$ is unramified;
(2) For all primes v of F dividing ℓ, the restriction of ρ to the local Galois group $Gal(\bar{F}_v/F_v)$ is *de Rham* in the sense of Fontaine.

One often compares ℓ-adic representations of Galois groups of number fields to finite-dimensional representations of fundamental groups of Riemann surfaces. In the geometric setting, condition (1) corresponds to the condition that the Riemann surface is algebraic, that is it is the complement in a projective algebraic curve of a finite set of points, which correspond to the points of ramification of the representation of the fundamental group of the closed curve. I know of no geometric analogue of condition (2), whose proof is one of the main theorems of p-adic Hodge theory (here $p = \ell$).

Fontaine and Mazur call an ℓ-adic Galois representation ρ of Γ_F *geometric* if it satisfies (1) and (2). The *Fontaine-Mazur conjectures* are an adaptation of the Langlands program to ℓ-adic representations of Galois groups of number fields. One of the conjectures states that any geometric ℓ-adic representation ρ is necessarily motivic. More relevant to our present discussion is another conjecture of Fontaine-Mazur that states (or would state, if it were written down for general n) that any geometric n-dimensional ℓ-adic representation ρ of Γ_F is necessarily associated to a *cuspidal automorphic representation* $\Pi(\rho)$ of $GL(n, F)$; the conjecture even gives an explicit recipe for $\Pi(\rho)$ as abstract representation. The (**commutative**) case $n = 1$ is an interpretation of class field theory combined with the Shimura-Taniyama-Weil theory of complex multiplication and Hecke characters. The set of cuspidal automorphic representations, whatever they are, have strong finiteness

[2]The representation ρ_E^n is not actually realized on the cohomology of the smooth proper variety E^n but rather on the part of the cohomology invariant under the permutations of the factors. This difference is immaterial for the purposes of the present paper.

properties and together the Fontaine-Mazur conjectures therefore imply that the set of motives with given Hodge numbers unramified outside S is finite. In this way the Langlands program has very stringent consequences for diophantine geometry.

The program initiated by Wiles reduces certain cases of the Fontaine-Mazur to a counting problem. The present section explains how to count geometric ℓ-adic Galois representations in the case of interest. The next section explains how to count those geometric ℓ-adic representations that do come from automorphic representations, together with the generalization of the Taylor-Wiles theorem that establishes a sufficiently strong version of equality.

Henceforward ℓ is an odd prime. The first step is to identify ρ as a point on an appropriate moduli space, or parametrized family of geometric ℓ-adic representations. Let $\mathcal{O}$ be a finitely generated local noetherian $\mathbb{Z}_\ell$-algebra with maximal ideal $\mathfrak{m}$ and residue field k. For the time being we assume $\mathcal{O}$ to be the integers in a finite extension of $\mathbb{Q}_\ell$. Let $\rho : \Gamma_F \to GL(n, \mathcal{O})$ be a continuous representation and let $\bar{\rho} : \Gamma_F \to GL(n, k)$ be its reduction modulo $\mathfrak{m}$. Fix a finite set of primes S including all primes dividing ℓ and all primes at which $\bar{\rho}$ is ramified, and let $\Gamma_{F,S}$ denote the Galois group of the maximal extension of F unramified outside S.

Definition 3.1. *Let A be a noetherian local $\mathcal{O}$-algebra with maximal ideal $\mathfrak{m}_A$ and residue field k. A* **lifting** *of $\bar{\rho}$ to A (understood to be unramified outside S) is a homomorphism $\tilde{\rho} : \Gamma_{F,S} \to GL(n, A)$ together with an isomorphism*

$$\tilde{\rho} \pmod{\mathfrak{m}_A} \xrightarrow{\sim} \bar{\rho}$$

compatible with the natural map $\mathcal{O}/\mathfrak{m} = A/\mathfrak{m}_A = k$. A **deformation** *of $\bar{\rho}$ to A is an equivalence class of liftings, where two liftings are equivalent if one can be obtained from the other by conjugation by an element of $Id + \mathfrak{m}_A M(n, A) \subset GL(n, A)$.*

Let $Def_{\bar{\rho},S}$ be the functor on the category of Artinian local $\mathcal{O}$-algebras that to A associates the set of equivalence classes of deformations of $\bar{\rho}$ to A (unramified outside S). This can be extended to a (pro)-functor on the category of noetherian local $\mathcal{O}$-algebras, and the starting point of the theory is Mazur's theorem

Theorem 3.2 (Mazur). *Suppose $\bar{\rho}$ is absolutely irreducible (or more generally $End(\bar{\rho} \otimes \bar{k}) = \bar{k}$). Then $Def_{\bar{\rho},S}$ is (pro)representable by a noetherian local $\mathcal{O}$-algebra $R_{\bar{\rho},S}$.*

We write $\mathfrak{m}_R$ for the maximal ideal of $R_{\bar{\rho},S}$ when this does not cause confusion. The Zariski tangent space $Hom(\mathfrak{m}_R/(\mathfrak{m}_R)^2, k) = Def_{\bar{\rho},S}(k[\varepsilon])$, with $\varepsilon^2 = 0$, has a natural interpretation in terms of Galois cohomology.

Proposition 3.3. *There is a natural isomorphism*

$$Hom(\mathfrak{m}_R/(\mathfrak{m}_R)^2, k) \xrightarrow{\sim} H^1(\Gamma_{F,S}, ad(\bar{\rho})),$$

where $ad(\bar{\rho})$ is the n^2-dimensional Galois module $Hom(\bar{\rho}, \bar{\rho})$.

The isomorphism is obtained as follows. There is a short exact sequence

$$1 \to \varepsilon M(n,k) \to GL(n,k[\varepsilon]) \to GL(n,k) \to 1.$$

Thus any two liftings r and r_0 of $\bar\rho$ to $GL(n,k[\varepsilon])$ differ by a map $[r-r_0]:\Gamma_{F,S}\to M(n,k)\xrightarrow{\sim}\varepsilon M(n,k)$, where $[r-r_0](g)=r(g)r_0(g)^{-1}-I$. Since r and r_0 are homomorphisms, one calculates easily that $[r-r_0]$ is a cocycle with values in $ad(\bar\rho)$ that is a coboundary if and only if r and r_0 are equivalent as deformations of $\bar\rho$. In what follows, we will consider r_0 to be a base point and write $[r]$ instead of $[r-r_0]$.

The vector space $H^1(\Gamma_{F,S}, ad(\bar\rho))$ is finite-dimensional, and this is used to prove that $R_{\bar\rho,S}$ is noetherian, but does not correspond in general to a counting problem that can be solved. Imposing supplementary restriction on deformations of $\bar\rho$ define alternative moduli problems. In many cases these can be proved to be representable and related to variants of $H^1(\Gamma_{F,S}, ad(\bar\rho))$ with better properties. We sketch the calculation of the dimension of the tangent spaces to the deformation rings used in the Taylor-Wiles method. The reader willing to take these on faith is invited to skip ahead to the result of the calculation, given in (3.7).

It seems the Taylor-Wiles method only works for Galois representations equipped with some sort of polarization. In particular, we need to assume from now on that F is either a totally real field or a CM field – a totally imaginary quadratic extension of a totally real field, so that in particular there is a well-defined complex conjugation $c \in Aut(F)$ of order 1 or 2, whose fixed field is a totally real field denoted F^+, as above. The representation ρ_E^n admits a non-degenerate bilinear form

$$\rho_E^n \otimes \rho_E^n \to \mathbb{Q}_\ell(1-n),$$

where $\mathbb{Q}_\ell(m)$ is the one-dimensional vector space $\mathbb{Q}_\ell$ on which Γ_F acts by the mth power of the cyclotomic character. The pairing is alternating if n is even and symmetric if n is odd. More generally, we consider ρ admitting non-degenerate pairings

$$\rho \otimes \rho\circ c \to \mathcal{O}\otimes\mathbb{Q}_\ell(1-n),$$

with c as above. Such ρ will be called *of unitary type* and will be related to automorphic forms on unitary groups.

Of course $\bar\rho$ admits a similar polarization with values in $k(1-n)$, defined analogously. One can consider deformations of $\bar\rho$ together with its polarization. This defines a representable moduli problem whose tangent space is isomorphic to to $H^1(\Gamma_{F^+,S}, ad(\bar\rho))$, where the extension of $ad(\bar\rho)$ to $\Gamma_{F^+,S}\supset\Gamma_{F,S}$ is defined in terms of the pairing.[3]

The next step is to impose local conditions at primes in S, which is now considered a set of primes of F^+ all of which split in the quadratic extension F/F^+; in particular this is true of all primes dividing ℓ. A *deformation problem* is then a collection $\mathcal{S}=\{\mathcal{D}_v, v\in S\}$ of conditions on the restriction of a deformation $\tilde\rho$ to Γ_v for $v\in S$ that

(i) define a representable moduli problem (so it can be studied abstractly) and

[3]In [CHT] we instead consider homomorphisms from $\Gamma_{F^+,S}$ with values in a certain algebraic group with two connected components, but there is no need to go into that level of detail here.

(ii) can be expressed in terms of the natural restriction map

$$loc_v : H^1(\Gamma_{F^+,S}, ad(\bar\rho)) \to H^1(\Gamma_v, ad(\bar\rho)),\ v \in S \tag{3.4}$$

(so that its numerical invariants can be calculated)

In the setting of [CHT], the most important local conditions are of two types.

($v \nmid \ell$) In this case, we want $\mathcal{D}_v$ to be *minimal*, which roughly means that $\tilde\rho$ is no more ramified than $\bar\rho$. More precisely, we define a k-subspace $L_v \subset H^1(\Gamma_v, ad(\bar\rho))$ so that the $\tilde\rho$ of type $\mathcal{D}_v$ are precisely those such that the corresponding cohomology class $loc_v([\tilde\rho]) \in L_v$ under the restriction map (3.4). Then $\mathcal{D}_v$ is minimal if and only if $\dim L_v = h^0_v := \dim H^0(\Gamma_v, ad(\bar\rho))$ (cf. [CHT, 2.4.21]).

($v \mid \ell$) When $\rho = \rho^n_E$ and ℓ is a prime of good reduction for E, then $\rho\ \ |_{\Gamma_v}$ is not only de Rham, in Fontaine's sense, but is crystalline and has Hodge-Tate weights corresponding to the classical Hodge decomposition of the subspace of $H^{n-1}(E^n, \mathbb{C})$ fixed under permutation of the factors. Under the standard convention, the Hodge-Tate weights are thus $0, -1, \dots, 1-n$, each with multiplicity one. The local condition we impose is thus that ρ be crystalline at all primes v dividing ℓ with the same Hodge-Tate weights as ρ^n_E. More generally, it suffices to assume that ρ is crystalline at each such v with n distinct Hodge-Tate weights; such a ρ is called *HT regular*. In order to make sense of this we need to assume $\ell > n$ and ℓ unramified in F. Then $\bar\rho$ and its liftings can be analyzed in terms of Fontaine-Laffaille modules [FL]. All that is needed in what follows is that, if $L_v \subset H^1(\Gamma_v, ad(\bar\rho))$ is the subspace corresponding to this deformation condition, then

$$\dim L_v - h^0_v = \frac{n(n-1)}{2}.$$

The *Selmer group* $H^1_{\mathcal{S}}(\Gamma_{F^+,S}, ad(\bar\rho)) \subset H^1(\Gamma_{F^+,S}, ad(\bar\rho))$ is defined to be $\cap_{v\in S} loc_v^{-1}(L_v)$. Now *Tate's local duality* defines a perfect pairing for every $v \in S$:

$$H^1(\Gamma_v, ad(\bar\rho)) \otimes H^1(\Gamma_v, ad(\bar\rho)(1)) \to k,$$

where $ad(\bar\rho)(1)$ is $ad(\bar\rho)$ tensored with the cyclotomic character. Let $L_v^\perp$ denote the annihilator of L_v with respect to this pairing and define

$$H^1_{\mathcal{S}^*}(\Gamma_{F^+,S}, ad(\bar\rho)) = \cap_{v\in S} loc_v^{-1}(L_v^\perp) \subset H^1(\Gamma_{F^+,S}, ad(\bar\rho)).$$

Abbreviate $h^1_{\mathcal{S}}(ad(\bar\rho)) = \dim H^1_{\mathcal{S}}(\Gamma_{F^+,S}, ad(\bar\rho))$ and so on, and write

$$\chi_v(\mathcal{S}) = \dim L_v - h^0_v.$$

For v real, we set $\chi_v(\mathcal{S}) = -\dim H^0(\Gamma_v, ad(\bar\rho)$. For v finite not in S, we set $\chi_v(\mathcal{S}) = 0$. This allows us to sum over all primes.

3.5. Sample calculation. An unramified representation of Γ_v is just a representation of the Frobenius element $Frob_v$. Consider the exact sequence:

$$0 \to \ker(Frob_v) \to ad(\bar\rho) \xrightarrow{Frob_v} ad(\bar\rho) \to \operatorname{coker}(Frob_v) \to 0.$$

Then $h_v^0 = \dim \ker(Frob_v)$, whereas we have $\dim(L_v) = \dim \operatorname{coker}(Frob_v)$ when $\mathcal{D}_v$ is the condition that deformations are unramified at v. Since this is our assumption for $v \notin S$, our notation is consistent.

One tries to reduce all calculations of $\chi_v(\mathcal{S})$ to considerations as simple as (3.5).

Combining Tate's global Euler characteristic formula for Galois cohomology with Poitou-Tate global duality, we obtain the following important identity:

$$h^1_{\mathcal{S}}(ad(\bar\rho)) - h^1_{\mathcal{S}^*}(ad(\bar\rho)(1)) = h^0(ad(\bar\rho)) - h^0(ad(\bar\rho)(1)) + \sum_v \chi_v(\mathcal{S}) \tag{3.6}$$

The sum is over all places v of F^+. The Taylor-Wiles method introduces a new collection of primes Q, split in F/F^+ and not overlapping S, and additional deformation conditions with the property that, for $v \in Q$, $\dim L_v - h_v^0 = 1$. The corresponding deformation problem is denoted $\mathcal{S}(Q)$, and the corresponding deformation ring is denoted $R_{\bar\rho,\mathcal{S}(Q)}$. Formula (3.6) remains valid in this generality, with every $\mathcal{S}$ replaced by $\mathcal{S}(Q)$. This allows $h^1_{\mathcal{S}}(ad(\bar\rho)$ to grow while simultaneously shrinking $h^1_{\mathcal{S}^*}(ad(\bar\rho)(1))$. In fact, in the applications we have the following simplifications:

(a) $h^0(ad(\bar\rho)) = h^0(ad(\bar\rho)(1)) = 0$.
(b) For v real[4], $\dim H^0(\Gamma_v, ad(\bar\rho)) = \dim ad(\bar\rho)^{c=1} = \frac{n(n-1)}{2}$
(c) There is a collection Q of (Taylor-Wiles) primes such that $h^1_{\mathcal{S}(Q)^*}(ad(\bar\rho)(1)) = 0$.

The existence of collections of Taylor-Wiles primes, which need to satisfy several other properties in addition to (c), is guaranteed provided the image of $\bar\rho$ is not degenerately small. We will soon need many such collections Q. For each such Q, we obtain the following very simple formula:

$$\dim \mathfrak{m}_{R_{\bar\rho,\mathcal{S}(Q)}}/\mathfrak{m}^2_{R_{\bar\rho,\mathcal{S}(Q)}} = h^1_{\mathcal{S}(Q)}(ad(\bar\rho)) = |Q|. \tag{3.7}$$

In the next section we will see that this expression yields a natural upper bound for deformations of $\bar\rho$ attached to automorphic forms. The Taylor-Wiles method shows this is also a lower bound.

Remark. Formula (3.7) says that an important property of the residual representation $\bar\rho$ of the Galois group of the number field F can be expressed in terms of its restrictions to carefully chosen local Galois groups. This is very roughly analogous to the (obvious) fact that the space of sections of a vector bundle on a compact manifold can be embedded in the sum of the fibers at carefully chosen finite sets of points. This corresponds in both cases to adding certain kinds of singularities at the chosen points. What is remarkable in the Taylor-Wiles method is that the singularities at sets of Taylor-Wiles primes can be made to approximate arbitrarily closely the full structure of the original deformation ring $R_{\bar\rho,\mathcal{S}(Q)}$.

[4]In [CHT] the equality (b) is only deduced at the end, but there is an independent proof due to Bellaïche and Chenevier.

4. Automorphic Galois representations and Hecke algebras

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

$$\text{(4.1)} \qquad \left\{ \begin{array}{l} \text{(Certain) cuspidal} \\ \text{automorphic} \\ \text{representations} \\ \Pi \text{ of } GL(n, \mathbf{A}_F) \end{array} \right\} \leftrightarrow \left\{ \begin{array}{l} \text{Compatible systems} \\ \rho_{\Pi,\ell} \text{ of irreducible } \ell\text{-adic} \\ \text{representations of} \\ \Gamma_F = Gal(\overline{\mathbb{Q}}/F) \\ \text{of dimension } n \end{array} \right\}$$

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 $L(s, \rho)$, defined as in (2.1.3, 2.1.4). 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 conjecturally associated to Galois representations. Maass forms for $GL(2, \mathbb{Q})$ are the most obvious example. They include, for example, cuspidal functions on $SL(2, \mathbb{Z})\backslash\mathfrak{H}$ that are eigenfunctions for all Hecke operators and for the hyperbolic Laplacian. Thanks to Selberg one knows the collection of such forms is large 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 [C]; 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 which, for $GL(n)$, is just the 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 setting of the Fontaine-Mazur conjectures [FM], one expects each ρ_ℓ in a compatible system to be *geometric* in the sense of Fontaine-Mazur, as defined in §3. In particular, they are *Hodge-Tate* at each prime dividing ℓ, and the dictionary predicts the Hodge-Tate weights in terms of the infinitesimal character of Π_∞.

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 $W(\Pi_\infty)$ of $GL(n, F_\infty)$. The precise condition is expressed in terms of relative Lie algebra cohomology:

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

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 $GL(n, F)$. I want to talk particularly about the following theorem whose proof is (at the time of writing) in the course of being written down [CHL2, Shin, CH]:

Theorem 4.2 (Many people). *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:*

$$\left\{ \begin{array}{l} \textit{(1) The factor } \Pi_\infty \\ \textit{ is cohomological} \\ \textit{(2) } \Pi \circ c \cong \Pi^\vee \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \textit{(a) } \rho = \rho_{\Pi,\lambda} \textit{ geometric,} \\ \textit{ HT regular} \\ \textit{(b) } \rho \otimes \rho \circ c \ \rightarrow \mathbb{Q}_\ell(1-n) \end{array} \right\}$$

This correspondence has the following properties:

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

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

Here G_v *is a decomposition group at* v *and* $\mathcal{L}$ *is the* **normalized** *local Langlands correspondence. The superscript* ^{ss} *refers to semisimplification, but*

(ii) *If* n *is odd or if the highest weight of* $L(\Pi_\infty)$ *is weakly regular, then one can replace "semisimplification" above by "Frobenius semi-simplification", which is the only sense in which* $\mathcal{L}$ *is defined; in particular, the local monodromy operator on the left-hand side is as predicted by the right-hand side.*

(iii) *(At least) under the same regularity hypothesis, one can show that* Π_v *is (essentially) tempered at all* v*;*

(iv) *The representation* $\rho_{\Pi,\lambda} \mid_{G_v}$ *is de Rham for any* v *dividing* ℓ *and the Hodge-Tate numbers at* v *are explicitly determined by the archimedean weight of* $L(\Pi_\infty)$*. If* Π_v *is unramified then* $\rho_{\Pi,\lambda} \mid_{G_v}$ *is crystalline.*

An automorphic representation of $GL(n, F)$*, with* F *a totally real or CM field, satisfying (1) and (2), or an* ℓ*-adic Galois representation satisfying (a) and (b), is said to be of* **CM type**.

Remark. It is expected that the $\rho_{\Pi,\lambda}$ are absolutely irreducible, but unfortunately this cannot be shown in general; this is the main open question concerning these representations. For the Π considered in [HT], and again in [CHT,T], one can prove by local arguments that $\rho_{\Pi,\lambda}$ is necessarily absolutely irreducible.

Corollary 4.3. *Under the hypotheses of Theorem 4.2, we have the following equality of L-functions.*

$$L(s + \frac{n-1}{2}, \rho_{\Pi,*}) = L(s, \Pi).$$

Here the right hand side is the Euler product attached by the formula of Artin and Serre to the compatible family of Galois representations $\rho_{\Pi,\lambda}$*. The left-hand side is the standard (Godement-Jacquet) L-function of the automorphic representation* Π.

The first results of this kind, for $F = \mathbb{Q}$ and $n = 2$, were proved by Eichler and Shimura in the 1950s. They worked with elliptic modular Hecke eigenforms of weight 2 rather than automorphic representations of $GL(2, \mathbb{Q})$. When $F = \mathbb{Q}$, the L-functions are Euler products whose factors are indexed by prime numbers. For all but finitely many primes p the local factor at p of $L(s, \rho_{\Pi,\lambda})$ is

$$\det(I - \rho_{\Pi,\lambda}(Frob_p)p^{-s})^{-1}$$

where $Frob_p \in Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ is the Frobenius element we already encountered in §1. This is (the inverse of) the value at p^{-s} of the polynomial of degree n $P_{p,\rho_{\Pi,\lambda}}(T) = \det(I - \rho_{\Pi,\lambda}(Frob_p)T)$. The typical Euler factor $L_p(s, \Pi)$ of $L(s, \Pi)$ is likewise (the inverse of) the value at p^{-s} of a degree n polynomial, which we write (cf. Appendix, A.3)

$$L_p(s, \Pi) = (1 - a_{1,p}(\Pi)p^{-s} + a_{2,p}(\Pi)p^{-2s} - \cdots + (-1)^n a_{n,p}(\Pi)p^{-ns})^{-1}. \tag{4.4}$$

The $a_{i,p}(\Pi)$ are integers in the number field $E(\Pi)$ and generalize the eigenvalues of the classical *Hecke operators* acting on elliptic modular forms. We return to these operators below.

The highest weight of $L(\Pi_\infty)$ can be expressed as a map from complex embeddings of F to n-tuples of integers, $v \mapsto (a_1(v), a_2(v), \ldots, a_n(v))$, where we may assume the a_i's are ordered so that

$$a_i(v) \geq a_{i+1}(v) \text{ for all } v.$$

The weak regularity condition, originally observed by Blasius and Rogawski in their work on 2- dimensional automorphic Galois representations, is that, for some v and at least one *odd* index i, the above inequality is strict. Under the weak regularity condition, $\rho_{\Pi,\lambda}$ can be constructed as the representation of Γ_F on a direct summand of the cohomology of an appropriate Shimura variety.[5] The remaining representations are constructed by means of ℓ-adic congruences, using results of Chenevier and Bellaïche-Chenevier on *eigenvarieties* parametrizing ℓ-adic automorphic forms on the unitary groups G introduced below. However, one can replace the weak regularity by the condition that Π be of discrete series type at some finite prime. This was the approach used in [CHT] and [T], following [HT]; it suffices for the theorems stated in §1.

The condition (a), that ρ be HT-regular, was invoked in the previous section. All the $\rho_{\Pi,\lambda}$ of Theorem 4.2 have this property, as do the representations $\rho^n_{E,\ell}$ that are our main object of interest. However, the Hodge-Tate weights of $\rho^n_{E,\ell}$ are incompatible with the weak regularity condition. Thus it does not suffice to work in the setting of weakly regular automorphic representations and the congruence methods mentioned above are crucial in general.

Let ρ be as in §3, satisfying conditions (a) and (b) of Theorem 4.2; in other words, ρ is of CM type. We say ρ is *automorphic* if it is isomorphic to $\rho_{\Pi,\lambda}$ for some Π and

[5]This is not strictly true; in general one can only realize the restrictions of $\rho_{\Pi,\lambda}$ to certain subgroups of index 2 in this way. A patching argument explained in [HT] and generalized by C. Sorensen reconstructs the entire representation $\rho_{\Pi,\lambda}$ from this information.

λ as above. More generally, we say $\bar\rho$ is automorphic, or ρ is *residually automorphic*, if $\bar\rho \equiv \bar\rho_{\Pi,\lambda}$ for some Π and λ. For the remainder of this section, we assume $\bar\rho$ to be automorphic, and we let Π be the corresponding automorphic representation.

Let $Def^{aut}_{\bar\rho,\mathcal{S}}$ be the functor on the category of Artinian local $\mathcal{O}$-algebras that to A associates the set of equivalence classes of deformations of $\bar\rho$ to A (unramified outside S) which are *automorphic*. At least when Π satisfies an additional local hypothesis at some finite prime, it is known that this subfunctor of $Def_{\bar\rho,\mathcal{S}}$ is representable by a quotient $\mathbb{T}_{\bar\rho,\mathcal{S}}$ of $R_{\bar\rho,\mathcal{S}}$. When $\rho = \rho^n_{E,\ell}$, the local hypothesis corresponds to the requirement that E have multiplicative reduction at some finite place. This is the setting of [CHT] and [T], where it is assumed that Π_v is a discrete series representation at some finite v prime to ℓ.[6]

The interest of this notion is that the quotient $\mathbb{T}_{\bar\rho,\mathcal{S}}$ of $R_{\bar\rho,\mathcal{S}}$ can be defined purely in terms of automorphic forms. These forms are not functions on the adèles of $GL(n)$ but are rather automorphic forms on a unitary group G. Before introducing this group, I return briefly to the case originally considered by Wiles. Recall that the ultimate result of the techniques he introduced was the existence of a modular form $f_E = \sum a_n(E)q^n$ attached to the elliptic curve E (cf. Theorem 1.3). The f_E belong to a space $S_2(N,\mathbb{C})$ of modular forms of weight 2 and some level N, an integer divisible only by the primes in the set S of bad reduction. More precisely, $S_2(N,\mathbb{C})$ is the space spanned by *newforms* of level N, including f_E. The newforms $f = \sum a_n(f)q^n$ have the property that the $\mathbb{Z}$-algebra $\mathbb{T}_N$ generated by operators T_p, one for each prime p not dividing N, acting on the newform f by $a_p(f)$, diagonalize the space $S_2(N,\mathbb{C})$. Of course the operators T_p have an independent group-theoretic definition, from which one can prove that the $\mathbb{Z}$-submodule $S_2(N,\mathbb{C})$ of $S_2(N,\mathbb{C})$ consisting of series $\sum a_n q^n$ with all $a_n \in \mathbb{Z}$ is stable under $\mathbb{T}_N$. In particular, the eigenvalues $a_p(f)$ are all **algebraic integers**, for each newform f, an obvious consequence of the definition for $f = f_E$.

More importantly, the algebra $\mathbb{T}_N$, although it is semisimple, does not in general diagonalize $S_2(N,\mathbb{Z})$. In other words, some elements of $S_2(N,\mathbb{Z})$ cannot be written as integral linear combinations of $\mathbb{T}_N$-eigenvectors, though they can be written as linear combinations with denominators. It is easy to see that if $f_1, f_2 \in S_2(N,\mathbb{Z})$ have the property that $\frac{1}{m}(f_1 - f_2) \in S_2(N,\mathbb{Z})$ for some integer m, then the Fourier coefficients of f_1 and f_2 are congruent modulo m. But if f_1 and f_2 are also (normalized) newforms, then among their Fourier coefficients are the traces $a_p(f_i)$ of a set of Frobenius operators $Frob_p, p \nmid N$ that suffice to determine the Galois representations ρ_{f_i} attached to $f_i, i = 1, 2$. In other words,

$$\rho_{f_1} \equiv \rho_{f_2} \pmod{m},$$

or again, for any prime ℓ dividing m, ρ_{f_2} is a deformation of the reduction mod ℓ $\bar\rho_{f_1}$ of ρ_{f_1}. It is in this way that congruences among automorphic forms translate into deformations of residual Galois representations.

Early work on congruences of modular forms, for example [M1] and [Hida], found it useful to view newforms as *functions* on the Hecke algebra $\mathbb{T}_N$ and to use geometric

[6]It is likely that work in progress of L. Guerberoff will permit the removal of this hypothesis.

properties of modular forms to derive algebraic properties of $\mathbb{T}_N$, and thus of the corresponding Galois representations. This was also the viewpoint of [W] and [TW]. For other kinds of automorphic forms there is no convenient theory of newforms, and thus no convenient way basis for functions on the relevant Hecke algebras. Diamond and Fujiwara [D,F] independently realized that, with the help of some basic constructions in commutative algebra, the module structures of automorphic forms over varying Hecke algebras could serve as a substitute for the theory of newforms.

To simplify the exposition, I will assume for the remainder of this section that $L(\Pi_\infty)$ is the *trivial* representation[7] and F is a field of the form $\mathbb{Q}(\sqrt{-d})$ for some positive integer d. The group G is easiest to understand as the unitary group of a positive-definite n-dimensional hermitian vector space over F.[8] The automorphic forms in question are functions on the adèles of G that are trivial on the group of real points $G(\mathbb{R})$ which, in our situation, is a *compact* group. Being automorphic, they are also trivial on the rational points $G(\mathbb{Q})$, and moreover right-invariant under an open compact subgroup K_f of the finite adèles $G(\mathbf{A}^f)$ of G. In other words, the automorphic forms belong to

$$S_{K_f}(G,\mathbb{C}) = \{f : Sh_{K_f}(G) = G(\mathbb{Q})\backslash G(\mathbf{A})/G(\mathbb{R})\cdot K_f \to \mathbb{C}\}.$$

By reduction theory, the set $Sh_{K_f}(G)$ is *finite*. However, the finite-dimensional vector space $S_{K_f}(G,\mathbb{C})$ is endowed with a rich structure by the action of the *Hecke operators*, which are $\mathbb{Z}$-valued functions on the discrete set $K_f\backslash G(\mathbf{A}^f)/K_f$ with compact (finite) support. These Hecke operators are attached to the group G rather than to $GL(n)$, but the theory of stable base change, due in this setting to Labesse, establishes close relations between the Hecke operators acting on $S_{K_f}(G,\mathbb{C})$ and the coefficients $a_{i,p}(\Pi)$ introduced in (4.4):

Theorem 4.5.

(1) *There is a commuting set of hermitian operators $\{T_{p,i}\}$ on the finite-dimensional space $S_{K_f}(G,\mathbb{C})$, where p runs through (half of the)*[9] *prime numbers and $1 \le i \le n$, whose simultaneous eigenspaces are in bijection with a subset of the set of Π satisfying (1) and (2) of Theorem 4.2; the eigenvector of $T_{p,i}$ on the eigenspace $V(\Pi)$ corresponding to Π is the coefficient $a_{i,p}(\Pi)$ of (4.4).*[10]

(2) *As K_f varies, every such Π occurs.*

(3) *The $T_{p,i}$ stabilize the $\mathbb{Z}$-submodule $S_{K_f}(G,\mathbb{Z}) \subset S_{K_f}(G,\mathbb{C})$ of $\mathbb{Z}$-valued functions, and in particular $a_{i,p}(\Pi)$ is an algebraic integer for all i, p, and Π.*

[7] In particular, it is not even weakly regular. Fortunately this condition plays no role in [CHT,T].

[8] the algebraic group G used in [CHT] and [T] is a twisted unitary group with a more complex description.

[9] Namely, the prime numbers that split in the imaginary quadratic field F. One can include all but finitely many of the remaining primes but their indexing is more complicated and their presence is superfluous.

[10] At present, this has only been established in this form for the twisted unitary groups used in [CHT] and [T]. A version of this theorem for a more general field F is due to Labesse [L].

Let $\mathbb{T}_{K_f}$ be the $\mathbb{Z}$-subalgebra of $End(S_{K_f}(G,\mathbb{Z}))$ generated by the $T_{p,i}$. For any Π as above, let $V^+(\Pi) \subset S_{K_f}(G,\mathbb{C})$ be the sum of the $V(\Pi')$ such that

$$a_{i,p}(\Pi') \equiv a_{i,p}(\Pi) \pmod{\ell}\ \forall i,p. \tag{4.6}$$

(Here we are implicitly assuming all the $a_{i,p}(\Pi') \in \mathbb{Z}$, as is the case for the $\Pi_{E,n}$ of Theorem 2.5. In general, (4.6) has to be modified to allow congruences in other coefficient fields.) Consider the ℓ-adic completion $\mathbb{T}_{\Pi,K_f}$ of the projection of $\mathbb{T}_{K_f}$ on $End(V^+(\Pi) \cap S_{K_f}(G,\mathbb{Z}))$.

Consider the mod ℓ representation $\bar{\rho} = \bar{\rho}_\Pi$ and the quotient $\mathbb{T}_{\bar{\rho},\mathcal{S}}$ of $R_{\bar{\rho},\mathcal{S}}$ introduced above. Assume $\bar{\rho}$ is absolutely irreducible. The following theorem is an application of a result of Carayol, together with the theory summarized in Theorem 4.5:

Theorem 4.7. *For an appropriate choice of $K_f = K_f(\mathcal{S})$, there is a natural isomorphism*

$$\mathbb{T}_{\bar{\rho}_\Pi,\mathcal{S}} \xrightarrow{\sim} \mathbb{T}_{\Pi,K_f(\mathcal{S})}$$

We thus obtain surjective maps

$$R_{\bar{\rho}_\Pi,\mathcal{S}} \to \mathbb{T}_{\Pi,K_f(\mathcal{S})} \tag{4.8}$$

The main theorem of [CHT] is roughly

Modularity Lifting Theorem 4.8 [CHT]. *Assume the deformation condition $\mathcal{S}$ is minimal, ℓ is split in the imaginary quadratic field F, and Π is unramified at primes dividing ℓ. Suppose further that the image of $\bar{\rho}_\Pi$ is not too small. Then (4.8) is an isomorphism.*

In other words, every (minimal) deformation of the residually automorphic Galois representation $\bar{\rho}_\Pi$ comes from automorphic forms. This theorem is valid when the quadratic imaginary field F is replaced by any CM field. For the applications to the Sato-Tate conjecture, minimal deformation conditions are unfortunately insufficient. Adapting a method of Kisin to handle non-minimal deformation conditions, Taylor proved

Modularity Lifting Theorem 4.9 [T]. *. Assume, ℓ is unramified in F and split in F/F^+, and Π is unramified at primes dividing ℓ. Suppose further that the image of $\bar{\rho}_\Pi$ is not too small. Then – in sufficient generality for the applications – (4.8) induces an isomorphism*

$$R^{red}_{\bar{\rho}_\Pi,\mathcal{S}} \to \mathbb{T}_{\Pi,K_f(\mathcal{S})},$$

where the superscript ^{red} denotes the quotient by the ideal of nilpotent elements.

This theorem implies in particular that any deformation of $\bar{\rho}_\Pi$ of type $\mathcal{S}$ with values in a ring without nilpotents – a representation with values in the integers in an ℓ-adic field, for example – is a $\rho_{\Pi',\ell}$ for some Π' as in Theorem 4.2. In the following section, we show how these results are used to prove Theorem 2.5.

The key to the proofs of Theorems 4.8 and 4.9 is the fact that the isomorphism (4.7) is valid more generally:

$$R_{\bar{\rho}_\Pi, \mathcal{S}(Q)} \to \mathbb{T}_{\Pi, K_f(\mathcal{S}(Q)} \tag{4.10}$$

whenever Q is a set of Taylor-Wiles primes. The estimate (3.7) gives an upper bound on the size of the left-hand side. On the other hand, purely group-theoretic methods give a lower bound on the size of the right-hand side. The elements of Q are all primes q split in F such that

$$q \equiv 1 \pmod{\ell^N}$$

for varying N. Letting Q vary appropriately so that N tends to infinity, an argument from commutative algebra, systematized in [D] and [F], yields the isomorphism in the limit.

In the approach of [T], Taylor follows Kisin [Ki] in replacing the rings in (4.10) by algebras over the moduli spaces of liftings of the local Galois representations at places v for which the deformation condition $\mathcal{D}_v$ is not minimal. Much of [T] is devoted to a careful comparison of these local moduli spaces with rigidified moduli spaces of nilpotent conjugacy classes in $GL(n)$ in characteristic ℓ and their deformations to quasi-nilpotent classes in characteristic zero. These structures are carried along, and the deformation rings $R_{\bar{\rho}_\Pi, \mathcal{S}(Q)}$ are replaced by rings classifying *framed deformations* in Kisin's sense. This considerably complicates the proof, but the Taylor-Wiles counting argument is still the basis of the method.

5. Moduli spaces of Calabi-Yau varieties and potential modularity

Consider the equation

$$f_t(X_0, X_1, \dots, X_n) = (X_0^{n+1} + \dots + X_n^{n+1}) - (n+1)tX_0 \dots X_n = 0, \tag{f_t}$$

where t is a free parameter. This equation defines an $n-1$-dimensional hypersurface $Y_t \in \mathbb{P}^n$ and, as t varies, a family:

$$\begin{array}{ccc} Y & \subset & \mathbb{P}^n \times \mathbb{P}^1 \\ & \searrow & \downarrow \\ & & \mathbb{P}^1_t \end{array}$$

Let μ_{n+1} denote the group of $n+1$'st roots of unity,

$$H = \mu_{n+1}^{n+1} / \Delta(\mu_{n+1}),$$

where Δ is the diagonal map, and let

$$H_0 = \{(\zeta_0, \dots, \zeta_n) \mid \prod_i \zeta_i = 1\} / \Delta(\mu_{n+1}) \subset H.$$

The group H_0 acts on each Y_t and defines an action on the fibration $Y/\mathbb{P}^1$. We examine the H_0-invariant part of the primitive cohomology $PH^{n-1}(Y_t)$ in the middle

dimension. The family Y was studied extensively by Dwork, who published articles about the p-adic variation of its cohomology when $n = 2$ (a family of elliptic curves) and $n = 3$ (a family of $K3$ surfaces). In what follows, we assume n even, so that $PH^{n-1}(Y_t) = H^{n-1}(Y_t)$.

Because f_t is of degree $n + 1$, Y_t, provided it is non-singular, is a Calabi-Yau hypersurface, which means that its canonical bundle is trivial (Y_t has a nowhere vanishing $(n-1)$-form, unique up to scalar multiples). This follows from standard calculations of cohomology of hypersurfaces. When $n = 4$, Y is a family of quintic threefolds in $\mathbb{P}^4$. The virtual number n_d of rational curves (Gromov-Witten invariants) on Y_t is determined by certain solutions of Picard-Fuchs equations describing monodromy on $H^3(Y_t)^{H_0}$. This is the phenomenon of mirror symmetry, predicted by the physicists Candelas, de la Ossa, Green, and Parkes, relating the Gromov-Witten invariants of Y_t with the Picard-Fuchs equation on $H^3((Y_t/H_0)^\sim)$, where $(Y_t/H_0)^\sim$ is a desingularization of (Y_t/H_0), and proved mathematically in a number of situations, including this one.

When $t = \infty$ Y_t is the union of coordinate hyperplanes; this is the totally degenerate case. The fiber Y_0 is the Fermat hypersurface

$$X_0^{n+1} + \cdots + X_n^{n+1} = 0.$$

This point is of great importance in the applications. The singular fibers Y_t are determined by an elementary calculation. In any characteristic prime to $n + 1$, we find the map f_t is smooth over $\mathbb{P}^*\mathbb{P}^1 \setminus \{\infty, {\mu_{\mathbf{n+1}}}^{n+1}\}$. The singularities at $t \in \mu_{n+1}$ are ordinary quadratic singularities and can be analyzed by Picard-Lefschetz theory. For any integer N prime to $n + 1$, the family $R^{n-1}f_{t,*}(\mathbb{Z}/N\mathbb{Z})^{H_0}$ is a local system $V[N]$ in free rank n $\mathbb{Z}/N\mathbb{Z}$-modules over $\mathbb{P}^*$. One verifies that it descends via the map $t \mapsto t^{n+1}$ to a local system over $\mathbb{P}^+ = \mathbb{P}^1 \setminus \{0, 1, \infty\}$, with a new singularity at 0 of finite order. This is a **rigid local system** and can be studied by the methods of [K].

For our purposes, we are interested in the fact, highlighted by the mirror symmetry conjectures, that $H^{n-1}(Y_t)^{H_0}$ has Hodge numbers $H^{p,n-1-p}$ all equal to one, $p = 0, 1, \ldots, n-1$, provided Y_t is nonsingular. This is calculated analytically, over $\mathbb{C}$, using Griffiths' theory of variation of Hodge structure of hypersurfaces, which also determines the Picard-Fuchs equation as an explicit ordinary differential equation of hypergeometric type. The calculation of the Hodge numbers shows that, when $t \in F^+$, the natural representation $\rho_{t,\ell}$ of Γ_F on $V_{t,\ell} := H^{n-1}(Y_t, \mathbb{Q}_\ell)^{H_0}$ is Hodge-Tate regular, in the sense of sections 3 and 4 above. In particular, the Fontaine-Mazur conjectures predict that this representation is obtained by the arrow of Theorem 4.2 from a cuspidal automorphic representation of $GL(n, F^+)$. This is in fact proved in many cases in [HST], using the results of [CHT,T].

The identification of the Picard-Fuchs equation as a hypergeometric equation allows us to apply the results of Beukers-Heckman [BH] to determine the Zariski closure of $\pi_1(\mathbb{P}^*(\mathbb{C}), t_0)$ (any base point t_0) in $Aut(V_{t_0,\ell})$. This is in turn applied by Guralnick and Katz to calculate the monodromy at finite level:

Theorem 5.1 [Guralnick-Katz, GHK, §4]. *Suppose N is a positive integer*

prime to $2(n+1)$. *Then the image of* $\pi_1(\mathbb{P}^*(\mathbb{C}), t_0)$ *in* $Aut(V[N]_{t_0})$ *equals*

$$Sp(V[N]_{t_0}) \xrightarrow{\sim} Sp(n, \mathbb{Z}/N\mathbb{Z}).$$

In other words, the minimal covering space $\mathcal{M}_N$ *of* $\mathbb{P}^*$ *trivializing the local system* $V[N]$ *together with its natural symplectic (Poincaré duality) pairing is an irreducible Galois covering with Galois group* $Sp(n, \mathbb{Z}/N\mathbb{Z})$.

In [HST] a weaker version of this theorem, proved following suggestions of Katz, was used to extend Taylor's *potential modularity* technique to even dimensional representations of dimension greater than 2. The idea is the following. Suppose you want to prove that an ℓ-adic representation, for example ρ_E^n, is automorphic. The Modularity Lifting Theorems of §4 show that, provided the image of the residual representation $\bar{\rho}_E^n$ is sufficiently large and a few other technical conditions are satisfied, this can be done whenever there is an automorphic representation Π of the type considered in Theorem 4.2 such that

$$\rho_E^n \equiv \rho_{\Pi,\ell} \pmod{\mathfrak{m}_{\mathcal{O}}}, \tag{5.2}$$

where $\mathcal{O}$ is the coefficient ring of $\rho_{\Pi,\ell}$. This places the burden of the method on proving *residual modularity* of ρ_E^n, in the sense of (5.2).

The main theorem of [HST] is motivated by the following ideal theorem:

Ideal Theorem 5.3. *There is a point* $t \in \mathbb{P}^*(\mathbb{Q})$, *a pair of rational primes* ℓ, ℓ', *and a cuspidal automorphic representation* Π' *of the type considered in Theorem 4.2, such that*

$$V_{t,\ell} \xrightarrow{\sim} \rho_E^n \pmod{\ell} \tag{5.3.1}$$

$$V_{t,\ell'} \equiv \rho_{\Pi',\ell'} \pmod{\mathfrak{m}_{\mathcal{O}'}} \tag{5.3.2}$$

as representations of $\Gamma_{\mathbb{Q}}$, *where* $\mathcal{O}'$ *is the coefficient ring of* $\rho_{\Pi',\ell'}$ *and* $\mathfrak{m}_{\mathcal{O}'}$ *is its maximal ideal.*

5.4. The basic strategy. Although it is unrealistic to expect to be able to prove such a theorem, it is worth taking a moment to show how it can be used to deduce (5.2). Assume $V_{t,\ell'}$ and ρ_E^n both satisfy the hypotheses of Theorems 4.8 and 4.9; in other words, that they are deformations of an appropriate type $\mathcal{S}$ Condition (5.3.2) shows that $V_{t,\ell'}$ is residually automorphic. We thus conclude that $V_{t,\ell'}$ is of the form $\rho_{\Pi_t,\ell'}$ for some automorphic representation Π_t. Since the ℓ-adic Galois representations on the cohomology of Y_t form a compatible system, it follows that $V_{t,\ell} \xrightarrow{\sim} \rho_{\Pi_t,\ell}$ is also automorphic. Now (5.3.1) implies (5.2), with $\Pi = \Pi_t$, This allows us to apply the Modularity Lifting Theorems to ρ_E^n and to conclude that ρ_E^n is itself automorphic.

Imitating the method of [TFM], the main theorem of [HST] is roughly

Optimal Theorem 5.5. *There is a totally real Galois extension $F_n/\mathbb{Q}$, a point $t \in \mathbb{P}^*(F_n)$, a pair of rational primes ℓ, ℓ' unramified in F_n, and a cuspidal automorphic representation Π' of $GL(n)_{F_n}$ of the type considered in Theorem 4.2, such that*

$$V[\ell]_t \xrightarrow{\sim} \rho_E^n \ |_{\Gamma_{F_n}} \pmod{\ell} \tag{5.4.1}$$

$$V_{t,\ell'} \equiv \rho_{\Pi',\ell'} \pmod{\mathfrak{m}_{\mathcal{O}'}} \tag{5.4.2}$$

as representations of $\Gamma_{\mathbb{Q}}$, where $\mathcal{O}'$ is the coefficient ring of $\rho_{\Pi',\ell'}$ and $\mathfrak{m}_{\mathcal{O}'}$ is its maximal ideal.

Arguing as in 5.4, we find the following potential version of (5.2):

$$\rho_E^n \ |_{\Gamma_{F_n}} \equiv \rho_{\Pi_t,\ell} \pmod{\mathfrak{m}_{\mathcal{O}}}, \tag{5.2(potential)}$$

and we again conclude that ρ_E^n is *potentially automorphic*, in other words it becomes automorphic only after restriction to Γ_{F_n}. This is Theorem 2.5 with $m = 1$, $F_n = F_{1,n}$. The (unconditional) proof of this theorem in [CHT,T,HST] under the hypothesis that $j(E) \notin \mathbb{Z}$, is essentially what has just been described; I have omitted discussion of an additional intermediate step, needed to accomodate the possible incompatibility of the restrictions of $V[\ell]_t$ and ρ_E^n to inertia groups at primes dividing ℓ.

The proof of a result similar to Optimal Theorem 5.5 is based on a diophantine approximation argument known in the literature as "Rumely's local-global principle" that roughly states that, if an irreducible algebraic variety over a number field has points locally at a finite set of places S, then it has points over a number field split at all places in S. In the applications, the number field is $\mathbb{Q}$, the set S consists of the real prime and the primes ℓ and ℓ' (in fact F_n is only assumed unramified at ℓ and ℓ'), in order to apply the modularity lifting theorems 4.7 and 4.9. The crucial irreducibility condition is guaranteed by Theorem 5.1. As in [TFM], [HST] uses a version of this principle, due to Moret-Bailly [MB], that is sufficiently flexible for our applications.

5.6. Remark. Theorem 5.1 and the local-global principle can be applied, as in 5.4, to a rather general class of symplectic representations ρ of Galois groups of totally real fields with values in totally ramified extensions of $\mathbb{Z}_\ell$. For the moment, the method is limited to ρ for which one can find local points over an unramified extension of $\mathbb{Q}_\ell$. An analogous argument has recently been found by T. Barnet-Lamb, using non H_0-invariant pieces of the cohomology of the Dwork family, to treat representations of Galois groups of CM fields satisfying condition (b) of Theorem 4.2 [B-L] but are not necessarily symplectic.

6. Applications of Brauer's theorem

To explain how Theorem 2.3 can be derived from a result like 5.5, I will step back and recall the theory of Artin L-functions, which are L-functions of complex representations of $Gal(\overline{\mathbb{Q}}/F)$, for any number field F.

Let

$$\rho : Gal(\overline{\mathbb{Q}}/F) \to GL(V) \simeq GL(n, \mathbb{C})$$

be a continuous representation on an n-dimensional complex vector space V. Thus the image of ρ is necessarily finite, hence factors through $Gal(E/F)$ for some finite extension E of F; in particular ρ is unramified outside the finite set of primes of F that ramify in E. For any prime ideal v of F that is unramified in E, we can define a (geometric) Frobenius element $Frob_v \in Gal(E/F)$ as before. Again, $Frob_v$ is only well defined up to conjugacy, but

$$L_v(s, \rho) = det(I - \rho(Frob_v)Nv^{-s})^{-1}$$

depends only on v. If v is ramified, we let $I_v \subset \Gamma_v$ be the intertia group. Then Γ_v/I_v acts on V^{I_v}, and we define

$$L_v(s, \rho) = det(I - \rho(Frob_v, V^{I_v})Nv^{-s})^{-1};$$

$$L(s, \rho) = \prod_v L_v(s, \rho).$$

This product converges absolutely for $Re(s) > 1$. Perhaps the most important conjecture in algebraic number theory is

Artin Conjecture. *If ρ is irreducible and non-trivial, then $L(s, \rho)$ is entire and satisfies a certain (explicit) functional equation.*

One has known for some time that

Theorem 6.1. *The function $L(s, \rho)$ is* meromorphic*, satisfies the expected functional equation, and is continous and non-vanishing for $Re(s) \geq 1$.*

This is essentially a consequence of Brauer's theorem on characters. I need to explain a few

6.2. Facts about Galois representations and their L-functions..

6.2.1. Semisimplification. *The representations ρ and ρ' have the same Jordan-Hölder constituents if and only if $Tr(\rho) = Tr(\rho')$, and the latter is true if and only if $L(s, \rho) = L(s, \rho')$ as Euler products.*

In particular, we can always replace ρ by its semisimplification (the direct sum of its Jordan-Hölder constituents.

6.2.2. Additivity. $L(s, \rho \oplus \rho') = L(s, \rho)L(s, \rho')$.

6.2.3. Inductivity. *Let F'/F be a finite extension, ρ' a continuous representation of $Gal(\overline{\mathbb{Q}}/F')$, $\rho = Ind_{F'/F}\rho'$ the induced representation of $Gal(\overline{\mathbb{Q}}/F)$. Then*

$$L(s, \rho') = L(s, \rho).$$

If ρ is the trivial representation of $Gal(\overline{\mathbb{Q}}/F)$, then $L(s, \rho) = \zeta_F(s)$ is the Dedekind ζ-function of F. More generally, if ρ is one-dimensional then it factors through $Gal(\overline{\mathbb{Q}}/F)^a b$.

6.2.4. Abelian L-functions. *Suppose* $\dim \rho = 1$ *and ρ is non-trivial. Then $L(s, \rho)$ is entire and satisfies the expected functional equation. Moreover, $L(s, \rho)$ is continuous and non-vanishing for $Re(s) \geq 1$.*

This is due to Hecke (Dirichlet when $F = \mathbb{Q}$) and follows from class field theory.

In particular, in the inductivity situation, if ρ' is abelian and non-trivial. then $L(s, Ind_{F'/F}\rho')$ satisfies the Artin conjecture.

Theorem 6.3 (Brauer). *Let H be a finite group and $\rho : H \to GL(n, \mathbb{C})$ be any finite-dimensional representation. Then there are solvable subgroups $H_i \subset H$, characters $\chi_i : H_i \to \mathbb{C}^\times$, and integers a_i such that*

$$\rho \equiv \oplus_i a_i Ind_{H_i}^H \chi_i.$$

The decomposition above is not unique, and the integers a_i are certainly not assumed positive. Applied to $\rho : H = Gal(E/F) \to GL(n, \mathbb{C})$, this and additivity implies

$$L(s, \rho) = \prod_i L(s, Ind_{F_i/F}\chi_i)^{a_i},$$

where F_i is the fixed field of H_i and χ_i is the character of $H_i = Gal(E/F_i)$; and again this is

$$\prod_i L(s, \chi_i)^{a_i}.$$

Since each of the $L(s, \chi_i)$ is entire and invertible for $Re(s) \geq 1$, the product is meromorphic and invertible for $Re(s) \geq 1$. The functional equation also follows from this product expression. We have not yet used that the H_i are solvable.

Now we return to the situation of an elliptic curve $E/\mathbb{Q}$ without complex multiplication, and assume $F_n/\mathbb{Q}$ is a finite Galois extension. Let 1_{F_n} be the *trivial* representation of $H = Gal(F_n/\mathbb{Q})$. Brauer's theorem applies to 1:

$$1_{F_n} = \oplus a_i Ind_{H_i}^H \chi_i.$$

Let L_i be the fixed field of H_i in F_n, ρ^n_{E,L_i} the restriction of ρ^n_E to $Gal(\overline{\mathbb{Q}}/L_i)$.

In general, if ρ is a representation of H, ρ' a representation of the subgroup $H' \subset H$, then

$$(Ind_{H'}^{H}\rho') \otimes \rho = Ind_{H'}^{H}(\rho' \otimes Res_{H'}^{H}\rho),$$

where $Res_{H'}^{H}\rho$ is the restriction of ρ to H. Applying this to $\rho = \rho_E^n$, with (H', ρ') varying amond the pairs (H_i, χ_i), it follows that

$$\rho_E^n = \oplus a_i (Ind_{H_i}^{H}\chi_i) \otimes \rho_E^n = \oplus a_i Ind_{H_i}^{H}\chi_i \otimes \rho_E^n)$$

which implies

$$L(s, \rho_E^n) = \prod L(s, \rho_{E,L_i}^n \otimes \chi_i)^{a_i}. \tag{6.4}$$

The following fact follows from a strengthened version of Arthur-Clozel base-change for certain kinds of automorphic representations of totally real fields.

Theorem 6.5. *Suppose F_n is totally real and Galois over $\mathbb{Q}$, and ρ_{E,F_n}^n is automorphic (of a certain type to be made precise below). Then for any solvable subgroup $H_i \subset H$ with fixed field L_i, and any character χ_i of $Gal(\overline{\mathbb{Q}}/L_i)$, $L(s, \rho_{E,L_i}^n \otimes \chi_i)$ is entire, and is invertible for $Re(s) \geq 1$.*

The invertibility statement in this theorem is due to Jacquet and Shalika and, in a more general setting, to Shahidi [JS,Sh].

Thus Theorems 5.5 and (the much older) 6.5 imply that the right-hand side of (6.4) is an alternating product of invertible Euler products, hence is itself invertible. This suffices to imply Theorem 2.3 for *even* n. The case of odd n is deduced in [HST] by a tensor product trick and Shahidi's theorem applied to Rankin-Selberg L-functions. This completes the proof of Theorem 2.3, which in turn implies the Sato-Tate conjecture for an elliptic curve with non-integral j-invariant.

7. Prospects

Since the appearance of Serre's book [S1] it has been understood that the Sato-Tate Conjecture for the elliptic curve E follows immediately once one has established certain analytic properties of L-functions of the ℓ-adic Galois representations ρ_E^n, summarized in Theorem 1.4. The techniques reviewed in sections 3-5 of this paper derive the desired analytic properties by proving the potential automorphy of ρ_E^n. These techniques can in principle be extended to more general ℓ-adic Galois representations ρ of CM type, but there are several obstacles. The most immediate obstacle is the one mentioned in Remark 5.6: one needs to know that the restriction of the residual representation $\bar{\rho}$ to an ℓ-adic decomposition group becomes isomorphic over an unramified extension to the local representation attached to some point t on the moduli space $\mathbb{P}^*$. In [HST] this is used to prove potential automorphy of certain representations of the form $V_{t,\ell}$; this argument has been generalized by Barnet-Lamb in [B-L].

The inequality $\ell > n$ mentioned briefly in §3 in connection with the Taylor-Wiles method is specific to ρ_E^n. For general ρ one obtains formula (3.7) only when $\ell > n(\rho)$, where $n(\rho) \geq n$ is determined by ℓ-adic Hodge theory and is in general much larger than n. This inequality is only compatible with the requirement of Remark 5.6 if the Hodge-Tate weights of ρ are exactly $0, -1, \ldots, 1-n$, each with multiplicity one; in other words, the same as those of ρ_E^n for an elliptic curve E. If such a ρ is automorphic then the corresponding Π must have $L(\Pi_\infty) = \mathbb{C}$ with the trivial representation. This is a serious restriction. Methods are known for relaxing this restriction, especially when $n = 2$ (see [TMC]) or when ρ is an ordinary representation at ℓ, but the application of the potential modularity methods of [HST] to more general ℓ-adic representations of CM type seems to require substantial progress in the p-adic Langlands program (with $p = \ell$), which for the moment is only complete for the group $GL(2, \mathbb{Q}_p)$.

Automorphy of the Galois representations attached to the ℓ-adic cohomology of curves of genus > 1 should in principle provide information on the asymptotics of points of general algebraic varieties over number fields, in the style of the Sato-Tate conjecture. Such representations satisfy condition (b) of Theorem 4.2 and are geometric in the sense of Fontaine-Mazur, but they are never Hodge-Tate regular. One expects these Galois representations to be automorphic, but they cannot only rarely occur directly in the cohomology of Shimura varieties. Existing methods in automorphic forms therefore provide no insight whatsoever into such representations, which is another way of saying that an entirely new approach is needed. For general representations not of CM type, I know one (extremely modest) positive result, a simple consequence of the results of [GHK]:

Theorem 7.1 [GHK]. *Let F^+ be totally real and let ρ be a (finite-dimensional) ℓ-adic representation of Γ_{F^+}. Then there is an ℓ-adic representation ρ' such that $\rho \oplus \rho'$ is residually potentially automorphic*

Here potential automorphy is intended in the following strong sense: for any finite extension M/F^+, there is a totally real Galois extension L/F^+ linearly disjoint from M such that $\rho \oplus \rho' \mid_{\Gamma_L}$ is residually automorphic over L. One even knows that

$$\rho \oplus \rho' \mid_{\Gamma_L} \equiv \rho_{\Pi,\ell} \pmod{\mathfrak{m}_{\mathcal{O}}^r} \tag{7.2}$$

where $\mathfrak{m}_{\mathcal{O}}$ is as in (5.2), Π is an automorphic representation of CM type, and r is any integer, though L may depend on r. And one has strong control of ρ'.

I want to insist, though, that the gap between this statement and genuine automorphy, or even potential automorphy, remains inconceivably vast. The automorphic methods outlined in §4 are simply not well adapted to the questions that arise naturally in arithmetic geometry. Calabi-Yau varieties are distinguished by the fact that a certain Hodge component of their middle-dimensional cohomology is of dimension one. The Hodge-Tate regularity condition imposed by existing methods in automorphic forms, applied to an algebraic variety, amounts to requiring that *all* Hodge components of their middle-dimensional cohomology are of dimension one. This requirement is relaxed slightly in the case of the Dwork family, where the action of the symmetric group is used to single out a part of the cohomology that

does satisfy the regularity condition. For most varieties the regularity condition is too stringent to be applied in any way. The fact that automorphic methods can be applied to elliptic curves, in some cases, has to be seen as a fortunate accident. No other such accident is apparent on the immediate horizon, but see [LBE].

Appendix: Properties of automorphic representations of $GL(n)$

For the purposes of this article, the formal definition of automorphic representation is not at all enlightening. What matters for our purposes is that the family of automorphic representations satisfies a list of axiomatic properties, some of which are recalled below. We start with a number field F. A cuspidal automorphic representation Π of $GL(n, F)$ can be defined as the representation of a certain locally compact group, the *adèle group* $GL(n, \mathbf{A}_F)$ on an irreducible constituent of $L_2^0(GL(n, F) \cdot Z^0 \backslash GL(n, \mathbf{A}_F))$, where $Z^0 \subset GL(n, F \otimes_{\mathbb{Q}} \mathbb{R})$ is a maximal subgroup of the center of $GL(n, F \otimes_{\mathbb{Q}} \mathbb{R})$ isomorphic to a product of copies of $\mathbb{R}$, and $L_2^0 \subset L_2$ is the subspace of cusp forms, whose definition is here omitted. This definition is convenient for defining the L-function and determining its analytic properties but sheds little light on the relation to number theory. What we need to know about Π is contained in the following list of properties.

A.1 (Factorization). For each place (prime ideal or archimedean valuation) v of F, Π has a local factor Π_v, which is an irreducible representation of the locally compact group $GL(n, F_v)$.

A.2 (Strong multiplicity one). Suppose Π and Π' are two cuspidal automorphic representations and S is a finite set of places such that $\Pi_v \xrightarrow{\sim} \Pi'_v$ for all $v \notin S$. Then $\Pi = \Pi'$; they are not only isomorphic but equal as subspaces of $L_2^0(GL(n, F) \cdot Z^0 \backslash GL(n, \mathbf{A}_F))$.

A.3 (Hecke eigenvalues). For all but finitely many prime ideals v, the local factor Π_v is an unramified principal series representation. It is characterized by an (unordered) n-tuple $\{\alpha_{1,v}, \ldots, \alpha_{n,v}\}$ of non-zero complex numbers, the *Satake parameters*, or equivalently by the *Hecke polynomial*

$$P_{\Pi_v}(T) = \prod_{i=1}^{n}(1 - \alpha_{i,v}T) = 1 - a_{1,v}(\Pi)T + a_{2,v}(\Pi)T^2 - \cdots + (-1)^n a_{n,v}(\Pi)T^n.$$

The (ordered) set of $a_{i,v}(\Pi)$ are the *local Hecke eigenvalues* of Π at v.

A.4 (L-function). For each v there is a local L-factor $L(s, \Pi_v)$ – an Euler factor if v is a prime ideal, a normalized product of Gamma-functions if v is archimedean – such that the product

$$L(s, \Pi) = \prod_v L(s, \Pi_v)$$

converges absolutely for $Re(s) > 1$ and extends to an entire function that satisfies a functional equation. For unramified places v,

$$L(s, \Pi_v) = P_{\Pi_v}(Nv^{-s})^{-1}$$

where Nv is the cardinality of the residue field of F at v.

A.5. Each Π_v is determined by its *local Langlands parameter* which is Galois-theoretic if v is a prime ideal.

A.6 (Non-vanishing). The function $L(s, \Pi)$ has no zeroes along the line $Re(s) = 1$.

A.7 (Arthur-Clozel base change). Let F'/F be a cyclic extension of prime degree. Then there is an automorphic representation (usually but not always cuspidal) $BC_{F'/F}(\Pi)$ of $GL(n, F')$. If w is a place of F' over the place v of F, the local factor $BC_{F'/F}(\Pi)_w$ depends only on Π_v. If Π_v is unramified then so is $BC_{F'/F}(\Pi)_w$, and its Satake parameters are given by an explicit formula in terms of the Satake parameters of Π_v. The Galois-theoretic local Langlands parameter of $BC_{F'/F}(\Pi)_w$ is the restriction of that of Π_v to the Galois group of the completion of F' at the prime ideal w.

Acknowledgments. I thank my coauthors, Laurent Clozel, Nick Shepherd-Barron, and Richard Taylor, for their collaboration over many years. Thanks also to Jim Carlson and Ariane Mézard pour their careful reading of the text and many helpful comments.

References

[B-L] T. Barnet-Lamb, Potential automorphy for certain Galois representations to $GL(2n)$, manuscript (2008)

[BH] F. Beukers, G. Heckman, Monodromy for the hypergeometric function ${}_nF_{n-1}$, *Invent. Math.*, **95** (1989), 325-354.

[Book 1] L. Clozel, M. Harris, J.-P. Labesse, and B. C. Ngô, eds., *The stable trace formula, Shimura varieties, and arithmetic applications, Book 1* (in preparation).

[BCDT] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises , *J. Amer. Math. Soc.*, **14** (2001), 843939.

[CH] G. Chenevier and M. Harris, and J.-P. Labesse, Construction of automorphic Galois representations, II, in preparation.

[C] L. Clozel, Motifs et formes automorphes: applications du principe de fonctorialité, in L. Clozel and J. S. Milne, eds., *Automorphic Forms, Shimura Varieties, and L-functions*, New York: Academic Press (1990) Vol. I, 77-160.

[CHL1] L. Clozel, M. Harris, and J.-P. Labesse, Endoscopic transfer for unitary groups, in [Book 1].

[CHL2] L. Clozel, M. Harris, and J.-P. Labesse, Construction of automorphic Galois representations, I, in [Book 1].

[CHT] L. Clozel, M. Harris, and R. Taylor, Automorphy for some ℓ-adic lifts of automorphic mod ℓ Galois representations, *Publ. Math. IHES*, to appear.

[D] F. Diamond, The Taylor-Wiles construction and multiplicity one, *Invent. Math.*, **128**, (1997), 379–391.

[FL] J.-M. Fontaine and G. Laffaille, Construction de représentations p-adiques, *Ann. Sci. E.N.S.*, **15** (1982) 547-608.

[FM] J.-M. Fontaine and B. Mazur, Geometric Galois Representations, in *Elliptic curves, modular forms, and Fermats last theorem (Hong Kong 1993)*, Internat. Press, Cambridge MA, (1995) 41-78.

[F] K. Fujiwara, *Deformation rings and Hecke algebras in the totally real case*, version 2.0, preprint 1999.

[G] A. Granville, Analytic Number Theory, in T. Gowers, J. Barrow-Green, and I. Leader, eds., *The Princeton Companion to Mathematics*, Princeton: Princeton University Press (2008) 332-348.

[GHK] R. Guralnick, M. Harris, and N. Katz, Automorphic realization of residual Galois representations, manuscript 2008.

[H] M. Harris, Potential automorphy of odd-dimensional symmetric powers of elliptic curves, and applications. to appear in Y. Tschinkel and Yu. Zarhin, eds., *Algebra, Arithmetic, and Geometry: Volume II: In Honor Of Y.I. Manin*, Boston: Birkhäuser.

[HST] M. Harris, N. Shepherd-Barron, and R. Taylor, A family of Calabi-Yau varieties and potential automorphy, *Annals of Math.*, to appear.

[Hida] H. Hida, Iwasawa modules attached to congruences of cusp forms. *Ann. Sci. ENS.* **19** (1986) 231-273.

[JS] H. Jacquet and J. A. Shalika, A non-vanishing theorem for zeta functions of $GL(n)$, *Invent. Math.*, **38** (1976/77), 1–16.

[K] N. Katz, *Rigid local systems*, *Annals of Mathematics Studies*, **139**, Princeton University Press (1996).

[Ki] M. Kisin, Moduli of nite at group schemes and modularity, *Annals of Math.*, to appear.

[L] J.-P. Labesse, Changement de base CM et séries discrètes, in [Book 1].

[LBE] R. P. Langlands, Beyond Endoscopy, in H. Hida, D. Ramakrishnan, F. Shahidi, eds., *Contributions to Automorphic Forms, Geometry, and Number Theory: A Volume in Honor of Joseph Shalika*, Johns Hopkins University Press (2004) 611-698.

[M1] B. Mazur, Modular curves and the Eisenstein ideal, *Publ. Math. IHES*, **47** (1977) 33-186.

[M2] B. Mazur, Finding meaning in error terms, *Bull. AMS*, **45** (2008) 185-228.

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

[S1] J.-P. Serre, *Abelian ℓ-adic representations and elliptic curves*, New York: Benjamin (1968).

[S2] J.-P. Serre (article in Motives) *Proc. Symp. Pure Math.*, **55**, Part 1 (1994) 377-400.

[Sh] F. Shahidi, On certain L-functions, *Am. J. Math.* **103** (1981), 297–355.

[Shin] S.-W. Shin, Galois representations arising from some compact Shimura varieties, manuscript (2008).

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

[TMC] R. Taylor, On the meromorphic continuation of degree two L-functions. *Doc. Math.* (2006) Extra Vol., 729-779.

[T] R. Taylor, Automorphy for some ℓ-adic lifts of automorphic mod ℓ Galois representations, II, *Publ. Math. IHES*, to appear.