From the Taniyama-Shimura Conjecture to Fermat's Last Theorem*

Kenneth A. Ribet†

1 Introduction

In this article I outline a proof of the theorem (proved in [25]):

$$\text{Conjecture of Taniyama-Shimura} \implies \text{Fermat's Last Theorem.}$$

My aim is to summarize the main ideas of [25] for a relatively wide audience and to communicate the structure of the proof to non-specialists. The discussion is inevitably technical at points, however, since a large amount of machinery from arithmetical algebraic geometry is required. The reader interested in a genuinely non-technical overview may prefer to begin with Mazur's delightful introduction [19] to the Taniyama-Shimura conjecture, and to relations with Fermat's Last Theorem and similar problems. Another excellent alternative source is the Bourbaki seminar of Oesterlé [21]. This seminar discusses the relation between elliptic curves and Fermat's Last Theorem from several points of view, but gives fewer details about the argument of [25] than the present summary. See also [11].

The proof sketched here differs from that of [25] in two ways. First of all, we exploit a suggestion of B. Edixhoven which allows us to prove the theorem without introducing an auxiliary prime in the proof. Such a prime is necessary to prove the general result of [25], but turns out to be superfluous in the case of Galois representations which are "semistable," but not finite,

*To appear in the *Annales de la Faculté des Sciences de Toulouse*.
†Supported by the N.S.F.

at some prime. (In the application to Fermat, the relevant representations are semistable and non-finite at the prime 2.) Secondly, we give a uniform argument to lower the level, avoiding a preliminary use of Mazur's Theorem ([25], §6), but appealing to the result of [20] instead.

This article is the written version of a June, 1989 lecture given at the Université Paul Sabatier (Toulouse) in connection with the *Prix Fermat*. I wish to thank Matra Espace, the sponsor of the prize, as well as the mathematical community of Toulouse, for their warm welcome. Special thanks are due Professor J.B. Hiriart-Urruty, who devised the prize, arranged its sponsorship, and organized its administration. The author also thanks B. Edixhoven, S. Ling and R. Taylor for helpful conversations, and the I.H.E.S. for hospitality during the preparation of this article.

2 Elliptic curves over Q

An elliptic curve over the rational field $\mathbf{Q}$ is a curve of genus 1 over $\mathbf{Q}$ which is furnished with a distinguished rational point O. (For background on elliptic curves, the reader is invited to consult [13], [34], or [28], Chapter IV.) The curves which will interest us are those arising from an affine equation

$$y^2 = x(x - A)(x + B), \tag{1}$$

where A and B are non-zero relatively prime integers with $A + B$ non-zero. The curve E given by this affine equation is understood to be the curve in the projective plane $\mathbf{P}^2$ given by the homogenized form of (1), namely

$$Y^2Z = X^3 - (A - B)X^2Z - ABXZ^2;$$

the point O is then the unique "point at infinity," which has homogeneous coordinates $(0, 1, 0)$

It is frequently essential to view E as a commutative algebraic group over $\mathbf{Q}$. In particular, given two points P and Q on E with coordinates in a field $K \supseteq \mathbf{Q}$ (we write $P, Q \in E(K)$), a sum $P + Q$ is defined, and is rational over K. The coordinates of $P + Q$ are rational functions in the coordinates of P and Q, with coefficients in $\mathbf{Q}$; these coefficients depend on the defining equation of E. The identity element of this group is the distinguished point O. Further, three distinct points sum to O in the group if and only if they are collinear.

According to the Weierstrass theory ([13], Ch. 9), we may describe E over $\mathbf{C}$ as the quotient $\mathbf{C}/L$, where L is a suitable period lattice for E. To do this, we fix a non-zero holomorphic differential ω of E over $\mathbf{C}$ and construct L as

$$\left\{ \int_\gamma \omega \;\Big|\; \gamma \in H_1\left(E(\mathbf{C}), \mathbf{Z}\right) \right\}.$$

From this description, we see easily for each $n \geq 1$ that the group

$$E[n] = \{\, P \in E(\mathbf{C}) \,|\, n \cdot P = O \,\}$$

is a free $\mathbf{Z}/n\mathbf{Z}$-module of rank 2, and in particular has order n^2. Indeed, this follows from the isomorphism $E[n] \approx \frac{1}{n}L/L$ and the fact that L is free of rank 2 over $\mathbf{Z}$. The points of $E[n]$ are those whose coordinates satisfy certain algebraic equations with rational coefficients, depending again on the defining equation of E. In particular, the finite group $E[n]$ consists of points of $E(\overline{\mathbf{Q}})$ and is stable under conjugation by elements of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Thus $E[n]$ may be viewed as a free rank-2 $\mathbf{Z}/n\mathbf{Z}$-module which is furnished with an action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. This action amounts to a homomorphism

$$\rho_n : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(E[n]),$$

where $\mathrm{Aut}(E[n])$ is the group of automorphisms of $E[n]$ as a $\mathbf{Z}/n\mathbf{Z}$-module, so that $\mathrm{Aut}(E[n]) \approx \mathbf{GL}(2, \mathbf{Z}/n\mathbf{Z})$. Notice that $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts on $E[n]$ through group automorphisms because conjugation is compatible with the addition law on E. To see this, we note again that the addition law is given by formulas with coefficients in $\mathbf{Q}$.

The kernel of ρ_n is an open subgroup H_n of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Let $K_n = \overline{\mathbf{Q}}^{H_n}$. Then K_n is concretely the finite extension of $\mathbf{Q}$ obtained by adjoining to $\mathbf{Q}$ the coordinates of all points in $E[n]$. This field is a Galois extension of $\mathbf{Q}$, whose Galois group $G_n = \mathrm{Gal}(K_n/\mathbf{Q})$ is isomorphic to the image of ρ_n, and especially is a subgroup of $\mathrm{Aut}(E[n])$. A great deal is known about the family of G_n as E/K remains fixed and n varies. For example, J-P. Serre proved in [29] that the index of G_n in $\mathrm{Aut}(E[n])$ remains bounded as $n \to \infty$, provided that E is not an elliptic curve with "complex multiplication."

The curves excluded by this theorem of Serre, i.e., those with complex multiplication, form an extremely interesting class of elliptic curves, for which a variety of problems can be treated. For example, Shimura [33] proved the Taniyama-Shimura Conjecture (described below) for such elliptic curves.

However, this class is essentially disjoint from the class of curves that we are going to consider.

Indeed, suppose that E is given by (1), where the integers A and B satisfy the mild conditions $32|B$, $4|(A+1)$. Then as Serre has indicated ([30], §4.1), E is a *semistable* elliptic curve, having bad reduction precisely at the primes dividing ABC, where $C = -(A+B)$. By a well known theorem of Serre-Tate [31], E does not have complex multiplication.

The concepts "semistability" and "bad reduction" can be explained in a rough way as follows (cf., e.g., [13], Ch. 5). Given an equation for E with integer coefficients, plus a prime number p, we reduce the equation mod p, thus obtaining an equation over the finite field $\mathbf{F}_p$. If this equation defines an elliptic curve $\tilde{E}$ over $\mathbf{F}_p$ (as opposed to a singular cubic curve), then E has *good reduction* at p, and the elliptic curve $\tilde{E}$ over $\mathbf{F}_p$ is the reduction of E mod p. For example, we get an elliptic curve mod p from the equation (1) if and only if the three numbers A, B, $A+B$ remain non-zero mod p. An elliptic curve E has *bad reduction* mod p if no equation for E can be found which reduces to an elliptic curve over $\mathbf{F}_p$. There is, in fact, a "best possible equation" for E: its minimal Weierstrass model ([13], Ch. 5, §2). The curve E has good reduction at p if and only if this model defines an elliptic curve mod p. In the case of bad reduction, the minimal Weierstrass model is singular mod p; one says that the bad reduction is *multiplicative* if the only singularity of the minimal Weierstrass model mod p is an ordinary double point. Finally, E is said to be semistable if, for every prime number p, E has either good or multiplicative reduction at p.

The *conductor* of an elliptic curve E over $\mathbf{Q}$ is a positive integer N which is divisible precisely by the primes p at which E has bad reduction ([13], Ch. 14). To say that E is semistable is equivalent to say that N is square free. The conductor N is then the product $\prod_{p\in\mathcal{B}} p^1$, where $\mathcal{B}$ is the set of primes of bad reduction. For instance, assume again that E is given by (1) and that the conditions

$$32|B, \quad 4|(A+1) \tag{2}$$

are satisfied. Then

$$N = \prod_{p|(ABC)} p. \tag{3}$$

Serre has suggested that N be termed the "radical" of ABC in this case.

A second integer associated with a curve E over $\mathbf{Q}$ is its *minimal discriminant*, a non-zero integer $\Delta = \Delta_E$ which is roughly the discriminant of the minimal Weierstrass equation for E. It is given in terms of the coefficients of this equation by a well known formula (see for example [13], p. 68). If E is given by (1) and the conditions (2) are satisfied, then we have

$$\Delta = \frac{1}{2^8} \cdot (ABC)^2. \tag{4}$$

Next, let E be an elliptic curve over $\mathbf{Q}$, and let p be a prime of good reduction for E. Write again $\tilde{E}$ for the reduction of E mod p, so that $\tilde{E}$ is an elliptic curve over $\mathbf{F}_p$. Let $\tilde{E}(\mathbf{F}_p)$ denote the group of rational points of $\tilde{E}$, i.e., the group of points on $\tilde{E}$ with coordinates in the base field $\mathbf{F}_p$. We define $a_p = a_p(E)$ to be the difference

$$p + 1 \; - \; \#\tilde{E}(\mathbf{F}_p).$$

To motivate consideration of this integer, we recall that $p+1$ represents the number of points on the projective line $\mathbf{P}^1$ over $\mathbf{F}_p$.

3 ℓ-division points

In this §, we suppose that E is the elliptic curve given by an equation (1), where A and B are as usual non-zero relatively prime integers such that $A+B$ is non-zero. We suppose in addition that the divisibilities (2) hold. We fix a prime number $\ell \geq 5$, and let $\rho = \rho_\ell$ be the representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ giving the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the group $E[\ell]$ of ℓ-division points of E. We let $K = K_\ell$ be the Galois extension of $\mathbf{Q}$ gotten by adjoining to $\mathbf{Q}$ the (coordinates of the) points in $E[\ell]$.

Proposition **3.1 (Mazur)** *The representation ρ is an irreducible two-dimensional representation of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

This proposition is proved as Proposition 6 in [30], §4.1. The proof given by Serre in [30] is based on results of Mazur [18], and relies on the fact that the 2-division points of E are rational over $\mathbf{Q}$ (i.e., $K_2 = \mathbf{Q}$). This latter fact is evident from the fact that the right-hand side of (1) factors over $\mathbf{Q}$.

The global property of ρ furnished by Proposition 3.1 is complemented by a second, much more elementary, global property. Let $\chi = \chi_\ell$ be the mod ℓ cyclotomic character $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{F}_\ell^*$ giving the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the group μ_ℓ of ℓ^{th} roots of unity in $\overline{\mathbf{Q}}$. Then we have

PROPOSITION **3.2** *The determinant of ρ is the character χ.*

The proposition follows from the isomorphism $\overset{2}{\Lambda} E[\ell] \xrightarrow{\sim} \mu_\ell$ given by the e_ℓ-pairing of Weil.

We now discuss some local properties of ρ. Let p be a prime number. Choose a place v of $\overline{\mathbf{Q}}$ lying over p, and let $D_v \subset \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ be the corresponding decomposition group. This group is isomorphic to the Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$, where $\overline{\mathbf{Q}}_p$ is the algebraic closure of $\mathbf{Q}_p$ in the completion of $\overline{\mathbf{Q}}$ at v. Let I_v be the inertia subgroup of D_v. The quotient D_v/I_v may be identified canonically with the Galois group of the residue field of v, which is an algebraic closure $\overline{\mathbf{F}}_p$ of $\mathbf{F}_p$. Hence D_v/I_v is isomorphic to the pro-cyclic group $\hat{\mathbf{Z}}$, with the chosen "generator" of the group being the Frobenius substitution $x \mapsto x^p$. Lifting this generator back to D_v, we obtain an element Frob_v of D_v which is well defined modulo I_v. Any other place v' of $\overline{\mathbf{Q}}$ over p is of the form $g \cdot v$ for some $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Replacing v by v' has the effect of conjugating the triple $(D_v, I_v, \mathsf{Frob}_v)$ by g.

We say that ρ is *unramified* at p if $\rho(I_v) = \{1\}$. This condition is visibly independent of the choice of v as a place lying over p. It amounts to the statement that the extension $K/\mathbf{Q}$ is unramified at the prime p, or equivalently to the condition that p be prime to the discriminant of K. If ρ is unramified at p, then $\rho(\mathsf{Frob}_v)$ is an element of the image of ρ which is well defined, i.e., independent of the choice of Frob_v as a lifting of the Frobenius substitution. Further, the *conjugacy class* of $\rho(\mathsf{Frob}_v)$ in the image depends only on p. Note that the representation ρ is ramified (i.e., not unramified) at the prime $p = \ell$. This follows from Proposition 3.2, since χ is ramified at ℓ. Equivalently, Proposition 3.2 guarantees that K contains the cyclotomic field $\mathbf{Q}(\mu_\ell)$, and this field is already ramified at ℓ.

The following result is standard (cf. [28], Ch. 4, §1.3).

PROPOSITION **3.3** *Suppose that p is prime to ℓN, where N is the conductor of E. Then ρ is unramified at p. Moreover, for each v dividing p, we have the congruence*

$$\mathrm{trace}\Big(\rho(\mathsf{Frob}_v)\Big) \equiv a_p \pmod{\ell}. \tag{5}$$

Let $\Delta = \Delta_E$ again be the minimal discriminant of E. For each prime number p, let $v_p(\Delta)$ be the *valuation* of Δ at p, i.e., the exponent of the highest power of p which divides Δ.

PROPOSITION **3.4** *Suppose that $p \neq \ell$ and that p divides N. Then ρ is unramified at p if and only if $v_p(\Delta) \equiv 0 \bmod \ell$.*

This result, noted by Frey [10] and Serre [30], §4, is an easy consequence of the theory of the Tate curve, as exposed in [28]. In concrete terms, one checks the ramification of ρ at p by viewing E over the maximal unramified extension $\mathbf{Q}_p^{\mathrm{nr}}$ of $\mathbf{Q}_p$ in $\overline{\mathbf{Q}}_p$. The representation ρ is unramified at p if and only if all points of $E[\ell]$ are defined over $\mathbf{Q}_p^{\mathrm{nr}}$. By the theory of the Tate curve, we can construct the extension $\mathbf{Q}_p^{\mathrm{nr}}(E[\ell])$ of $\mathbf{Q}_p^{\mathrm{nr}}$ by adjoining to $\mathbf{Q}_p^{\mathrm{nr}}$ the ℓ^{th} roots of the Tate parameter q attached to E over $\mathbf{Q}_p$. Since $\ell \neq p$, q is an ℓ^{th} power in $\mathbf{Q}_p^{\mathrm{nr}}$ if and only if the valuation of q is divisible by ℓ. This valuation coincides with the valuation of Δ.

Proposition 3.4 remains true if the hypothesis $p|N$ is suppressed. Indeed, if p is prime to N, then $v_p(\Delta) = 0$, so the congruence $v_p(\Delta) \equiv 0 \bmod \ell$ is satisfied *a fortiori*. At the same time, the representation ρ is unramified at p, as stated in Proposition 3.3.

In order to remove the hypothesis $p \neq \ell$ in the two previous Propositions, we need to introduce the technical notion of *finiteness* ([30], p. 189). Let p be a prime number, and choose a place $v|p$. The decomposition group D_v is then the Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$, as noted above. By restricting to D_v the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $E[\ell]$, we may view $E[\ell]$ as a $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$-module. We say that ρ is *finite* at p if there is a finite flat group scheme $\mathcal{V}$ of type (ℓ, ℓ) over $\mathbf{Z}_p$ such that the $\mathrm{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$-modules $\mathcal{V}(\overline{\mathbf{Q}}_p)$ and $E[\ell]$ are isomorphic. If $p \neq \ell$, this means simply that ρ is unramified at p. In general, we may paraphrase this property of ρ by saying that $E[\ell]$ has "good reduction at p."

PROPOSITION **3.5** *The representation ρ is finite at a prime number p if and only if $v_p(\Delta) \equiv 0 \bmod \ell$.*

The only new information provided by this Proposition is the criterion for finiteness in the case $p = \ell$. For the proof, see [30], p. 201.

4 Modular elliptic curves

Let E be an elliptic curve over $\mathbf{Q}$. The *Taniyama-Shimura Conjecture* (also known as the Weil-Taniyama Conjecture) states that E is *modular*. One can define this concept in several equivalent ways, either by connecting up the integers a_p of §2 with the coefficients of modular forms, or else by considering the geometry of modular curves.

We begin with the latter. (See [32], [22], or [34], Ch. 11, §3 for background on modular curves.) Let N be a positive integer, and consider the group

$$\Gamma_o(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}(2, \mathbf{Z}) \;\Big|\; c \equiv 0 \bmod N \right\}.$$

Recall that the group $\mathbf{SL}(2, \mathbf{Z})$ acts on the Poincaré upper half plane

$$\mathcal{H} = \{ \tau \in \mathbf{C} \mid \Im(\tau) > 0 \}$$

by fractional linear transformations

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : \tau \mapsto \frac{a\tau + b}{c\tau + d}.$$

The quotient $Y_o(N) = \Gamma_o(N) \backslash \mathcal{H}$ is an open Riemann surface, which has a standard compactification $X_o(N)$. This *modular curve* has a canonical model over $\mathbf{Q}$, which we denote simply $X_o(N)$ to lighten notation.

Although there is no question here of explaining the origin of the canonical model, we should mention at least that this model is connected intimately with the interpretation of $X_o(N)$ and $Y_o(N)$ as *modular varieties*. The idea is that to each $\tau \in \mathcal{H}$ we can associate the lattice $L_\tau = \mathbf{Z} + \mathbf{Z}\tau$ in $\mathbf{C}$, and then the elliptic curve $E_\tau = \mathbf{C}/L_\tau$. Two elliptic curves E_τ and $E_{\tau'}$ are isomorphic if and only if the elements τ and τ' of $\mathcal{H}$ have the same image in $\mathbf{SL}(2, \mathbf{Z}) \backslash \mathcal{H}$. More generally, given $N \geq 1$, we associate to $\tau \in \mathcal{H}$ the pair (E_τ, C_N) consisting of the elliptic curve E_τ and the cyclic subgroup C_N of E generated by the image of $\frac{1}{N} \in \mathbf{C}$ on E_τ. Then two τ's give isomorphic pairs if and only if they map to the same element of $Y_o(N)$. Further, every pair consisting of an elliptic curve E over $\mathbf{C}$, together with a cyclic subgroup of order N arises in this way from some τ. Hence the curve $Y_o(N)$ parametrizes isomorphism classes of data "elliptic curves with cyclic subgroups of order N." It is via this interpretation that one defines $Y_o(N)$ and $X_o(N)$ over $\mathbf{Q}$.

In fact, with a more sophisticated definition of the data to be classified, one arrives at a good definition over $\mathbf{Z}$ [16].

The Jacobian of $X_o(N)$ is the abelian variety $J_o(N)$; this is a complete group variety whose dimension is the genus $g(N)$ of the curve $X_o(N)$. This genus is given by a well known formula (see, e.g., [23], p. 455). The integer $g(N)$ is 0 for all $N \leq 10$, and we have for example $g(11) = 1$, $g(23) = 2$.

We let $S(N)$ be the complex vector space of weight-2 holomorphic cusp forms for $\Gamma_o(N)$. Such a form is a holomorphic function $f(\tau)$ on $\mathcal{H}$ for which the differential $f(\tau)\,d\tau$ is invariant under the group $\Gamma_o(N)$. Thus $f(\tau)\,d\tau$ may be viewed as a regular differential on the open curve $Y_o(N)$, and the condition that $f(\tau)$ belong to $S(N)$ is the condition that $f(\tau)\,d\tau$ extend to a holomorphic differential on $X_o(N)$ over $\mathbf{C}$. The space $S(N)$ may thereby be identified with the space of holomorphic differentials on the curve $X_o(N)$, a space of dimension $g(N)$.

Each $f(\tau) \in S(N)$ may be expanded as a Fourier series $\sum\limits_{n\geq 1} a_n q^n$, where $q = e^{2\pi i \tau}$. The sum $\sum a_n q^n$ is viewed often as a formal series in q, and is known as the "q-expansion" of the cusp form f.

The space $S(N)$ comes equipped with a commuting family of endomorphisms, the *Hecke operators* T_n ($n \geq 1$). (See [32], Chapter 3.) These operators are traditionally written on the right, so that $f|T_n$ denotes the image of f under T_n. The T_n may be expressed as polynomials in the operators T_p where p is prime. (For example, they are multiplicative: $T_{nm} = T_n \circ T_m$ when n and m are relatively prime.) The operators T_p are given by the following well known formula, expressing the q-expansion of $f|T_p$ in terms of that of f:

$$T_p : f = \sum a_n q^n \;\mapsto\; \begin{cases} \sum a_{pn} q^n + p \sum a_n q^{pn} & \text{if } p \text{ is prime to } N; \\ \sum a_{pn} q^n & \text{if } p \text{ divides } N. \end{cases}$$

An *eigenform* in $S(N)$ is a non-zero $f \in S(N)$ which is an eigenvector for each of the operators T_n. The eigenvalues (λ_n) coming from a given eigenform f are known to be algebraic integers which all lie in a finite extension of $\mathbf{Q}$ which is independent of n.

Taniyama-Shimura Conjecture. *Let E be an elliptic curve over $\mathbf{Q}$, and let N be its conductor. Then there is an eigenform $f \in S(N)$ with the following property: for each prime p not dividing N, the integer a_p of §2 is the eigenvalue of f for the operator T_p.*

If f exists for a given E, one says that E is *modular*. Because of our detailed knowledge of the arithmetic of modular forms and abelian varieties (the author is especially thinking of [2, 9]), the condition that E be modular can be reformulated in a variety of ways. For example, B. Mazur [19] has recently proved

PROPOSITION **4.1** *Suppose that E is an elliptic curve over $\mathbf{Q}$ with the following property: there is a non-zero homomorphism of abelian varieties $h\colon E \to J_o(M)$, defined over $\mathbf{C}$, for some $M \geq 1$. Then E is modular.*

The novelty of this statement is that h is not assumed to be defined over $\mathbf{Q}$.

Another remark which may serve to orient the reader is that the form f in the Taniyama-Shimura Conjecture is (up to multiplication by a constant) the unique non-zero element of $S(N)$ having the property

$$f|T_p = a_p \cdot f \text{ for all but finitely many } p.$$

More precisely, f is a constant multiple of a newform (in the sense of [1]) of level N. All coefficients of f are rational integers.

Finally, A. Weil proved in [35] that an elliptic curve E over $\mathbf{Q}$ is modular provided that its L-function $L(E,s)$ has the analytic properties one expects of it. This L-function is defined as an infinite product $\prod_p L_p(E,s)$ ($s \in \mathbf{C}$, $\Re s > 3/2$), over the set of prime numbers p (Euler product). The factor $L_p(E,s)$, for p prime to N, is

$$(1 - a_p p^{-s} + p^{1-2s})^{-1}.$$

5 Modular representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$

Let N be a positive integer, and let $\mathbf{T} = \mathbf{T}_N$ be the subring (i.e., $\mathbf{Z}$-algebra) of $\mathrm{End}_{\mathbf{C}}(S(N))$ generated by the operators T_n for $n \geq 1$. As is well known ([32], Ch. 3), this ring is a free $\mathbf{Z}$-module of rank equal to the dimension $g(N)$ of $S(N)$. In particular, if m is a maximal ideal of $\mathbf{T}$, then the residue field $k_{\mathsf{m}} = \mathbf{T}/\mathsf{m}$ is a finite field, say of residue characteristic ℓ. One can show (see [5], Th. 6.7 or [25], §5) that there is a semisimple continuous homomorphism

$$\rho_{\mathsf{m}} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, k_{\mathsf{m}})$$

having the properties:

1. $\det \rho_{\mathsf{m}} = \chi_\ell : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{F}_\ell^* \subseteq k_{\mathsf{m}}^*$
2. ρ_{m} is unramified at all primes p not dividing ℓN
3. $\mathrm{trace}\, \rho_{\mathsf{m}}(\mathsf{Frob}_v) = T_p \bmod \mathsf{m}$ for all $p \not| \ell N$.

(Here, χ_ℓ is again the mod ℓ cyclotomic character of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ and, in the last property, Frob_v is a Frobenius element in a decomposition group D_v for a prime of $\overline{\mathbf{Q}}$ dividing p.)

The representation ρ_{m} is unique up to isomorphism. This follows from the Cebotarev Density Theory, which states that all elements of the image of ρ_{m} are conjugate to Frobenius elements $\rho_{\mathsf{m}}(\mathsf{Frob}_v)$, plus the fact that a two-dimensional semisimple representation is determined by its trace and determinant.

Example. Let E be a modular elliptic curve of conductor N, and let $f \in S(N)$ be an eigenform whose eigenvalue under T_p is a_p for each p prime to N. The action of $\mathbf{T}$ on f is given by the homomorphism $\varphi : \mathbf{T} \to \mathbf{C}$ which takes each $T \in \mathbf{T}$ to the eigenvalue of f under T. This homomorphism is in fact $\mathbf{Z}$-valued, as remarked above. If ℓ is a prime number, and $\mathsf{m} = \varphi^{-1}\Big((\ell)\Big)$, then $\rho_{\mathsf{m}} : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F}_\ell)$ is the *semisimplification* ρ^{ss} of the representation $\rho : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(E[\ell])$. This semisimplification is defined to be the direct sum of the Jordan-Hölder factors of ρ. In other words, ρ^{ss} is the direct sum of two one-dimensional representations in case ρ is not simple, while ρ^{ss} is ρ if ρ is already semisimple. Hence, we have $\rho_{\mathsf{m}} \approx \rho$ in this latter case.

Now let $\mathbf{F}$ be a finite field and suppose that

$$\sigma : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$$

is a continuous semisimple representation. We say that σ is *modular of level* N if there is a maximal ideal m of $\mathbf{T}$ and an embedding $\iota : \mathbf{T}/\mathsf{m} \subset \overline{\mathbf{F}}$ such that the $\overline{\mathbf{F}}$-representations

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \xrightarrow{\sigma} \mathbf{GL}(2, \mathbf{F}) \subset \mathbf{GL}(2, \overline{\mathbf{F}})$$

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \xrightarrow{\rho_{\mathsf{m}}} \mathbf{GL}(2, \mathbf{T}/\mathsf{m}) \xhookrightarrow{\iota} \mathbf{GL}(2, \overline{\mathbf{F}})$$

are isomorphic. Equivalently, one requires a homomorphism

$$\omega : \mathbf{T} \to \overline{\mathbf{F}}$$

such that

$$\mathrm{trace}\Big(\sigma(\mathsf{Frob}_p)\Big) = \omega(T_p), \qquad \det\Big(\sigma(\mathsf{Frob}_p)\Big) = \bar{p}$$

for all but finitely many primes p. Here, Frob_p denotes Frob_v for some $v|p$ and $\bar{p}$ is the image of p in $\mathbf{F}$. (Thus $\bar{p}$ is p mod ℓ, if ℓ is the characteristic of $\mathbf{F}$.) The term "modular of level N" is a shorthand phrase used to indicate that the representation σ arises from the space $S(N)$; this space could be described more verbosely as the space of weight-2 cusp forms on $\Gamma_o(N)$ with trivial character.

Assuming that ρ is modular of level N, we say that N is *minimal* for ρ if there is no divisor M of N, with $M < N$, such that ρ is modular of level M. Regarding ρ as fixed, and possible levels N as varying, we note that if ρ is modular of some level N, then ρ is modular of some minimal level $N_o|N$. (The question of the uniqueness of N_o is more subtle, but not an issue here.) Also, it is essentially evident that once ρ is modular of level N, it is modular of all levels N' which are multiples of N.

The following statement is a variant of the main theorem of [25].

THEOREM **5.1** *Let σ be an irreducible two-dimensional representation of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ *over a finite field of characteristic $\ell > 2$. Assume that σ is modular of square free level N, and that there is a prime $q|N$, $q \neq \ell$ at which σ is not finite. Suppose further that p is a divisor of N at which σ is finite. Then σ is modular of level N/p.*

This theorem immediately gives:

$$\text{Conjecture of Taniyama-Shimura} \implies \text{Fermat's Last Theorem.}$$

Indeed, suppose that the Taniyama-Shimura Conjecture is true. To prove Fermat's Last Theorem under this assumption, it suffices to show that there is no triple (a, b, c) of non-zero integers which satisfies the equation

$$a^\ell + b^\ell + c^\ell = 0,$$

where $\ell \geq 5$ is a prime. Arguing by contradiction, we assume that there is such a triple. Dividing out by any common factor, we may suppose that the three integers a, b, and c are relatively prime. Further, after permuting these

integers, we may suppose that b is even and that $a \equiv 1 \bmod 4$. Following Frey [10], we consider the elliptic curve E with Weierstrass equation

$$y^2 = x(x - a^\ell)(x + b^\ell).$$

According to our discussion above, the conductor of E is the radical $N = \mathrm{rad}(abc)$ of abc, and its minimal discriminant is $\Delta = (abc)^{2\ell}/2^8$. The group $V = E[\ell]$ provides an irreducible representation σ of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ which is finite at every odd prime but not finite at the prime 2. Applying Theorem 5.1 inductively, we deduce that σ is modular of level 2. This is impossible, since the ring $\mathbf{T}_2 \subseteq \mathrm{End}(S(2))$ is the 0-ring, in view of the fact that the dimension $g(2)$ of $S(2)$ is 0.

To prove the Theorem, we will work with two related abelian varieties over $\mathbf{Q}$. First, we have $J_o(N)$, the Jacobian $\mathrm{Pic}^o(X_o(N))$ of the modular curve $X_o(N)$. We make use of the fact that the ring $\mathbf{T} = \mathbf{T}_N$ operates faithfully on $J_o(N)$ as a ring of endomorphisms which are defined over $\mathbf{Q}$. This action is determined uniquely by the relation between it and the tautological action of $\mathbf{T}$ on $S(N)$. Namely, $S(N)$ is canonically the space of differentials on the Albanese variety $\mathrm{Alb}(J_o(N))$, which in turn is the abelian variety dual to $J_o(N)$. Thus $S(N)$ is functorially the cotangent space to the dual of $J_o(N)$. This functorial relation translates the action of $\mathbf{T}$ on $J_o(N)$ as a ring of endomorphisms into the classical action of $\mathbf{T}$ on $S(N)$.

Suppose now that σ is given as in the Theorem, and let $\mathsf{m} \subset \mathbf{T}$ be a maximal ideal corresponding to σ as in the definition of "modular of level N." One checks immediately that we can replace σ by ρ_{m} in the Theorem. From now on, let us regard m as fixed and let us assume that σ is ρ_{m}. We will write simply k for the residue field k_{m} of m and ρ for ρ_{m}.

For each endomorphism $\alpha \in \mathsf{m}$, we let $J_o(N)[\alpha]$ be the kernel of α on the group $J_o(N)(\overline{\mathbf{Q}})$ of points on $J_o(N)$ with coordinates in $\overline{\mathbf{Q}}$. Let $W = J_o(N)[\mathsf{m}]$ be the intersection

$$\bigcap_{\alpha \in \mathsf{m}} J_o(N)[\alpha].$$

This group is a subgroup of the finite group $J_o(N)[\ell]$ of ℓ-division points of $J_o(N)$, and carries natural commuting actions of k and of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

Here is the first key result needed to prove Theorem 5.1:

PROPOSITION **5.2** *Assume, if ℓ divides N, that ρ is not modular of level N/ℓ. Then ρ is equivalent to the k-representation W of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

Let V be the representation ρ, regarded as a 2-dimensional k-vector space with an irreducible action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Using the Eichler-Shimura relations for $J_o(N)$, one shows immediately ([18], Chapter II, Proposition 14.2, or [25], Theorem 5.2) that the semisimplification of W as a $k[\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$-module is isomorphic to a non-empty product of copies of V. Thus the proposition states, in effect, that the k-dimension of W is 2. This is again easy to prove in case ℓ is prime to N *(loc. cit.)*. When $\ell|N$, [20] proves the desired statement under the hypothesis that ρ is not modular of level N/ℓ. The reader is invited to consult these articles for details of the proofs.

With this preliminary proposition recorded, it is time to discuss the strategy of the proof of Theorem 5.1. As noted above, the representation ρ occurring in Theorem 5.1 will be modular of some minimal level $D|N$. This level is divisible by q, because ρ is assumed to have "bad reduction" at this prime (i.e., assumed not to be finite at q). Further, to prove Theorem 5.1, it suffices to show that D is prime to p. Indeed, if D is prime to p, then N/p is a multiple of D, and ρ is certainly modular of level N/p, as desired.

To prove that D is prime to p, we can argue by contradiction, supposing instead that D is divisible by p. In this situation, D is of the form pqL, with L square free and prime to pq. Thus the pair (ρ, D) has all the properties of the pair (ρ, N), and the additional property that D is a minimal level for ρ. We are to deduce a contradiction from these data.

Replacing D by N, we see that Theorem 5.1 will follow if we can deduce a contradiction from the following assumptions about the irreducible representation ρ:

1. ρ is modular of level N, where N is square free and divisible by the distinct primes p and q, with $q \neq \ell$.

2. ρ is finite at p, but not at q.

3. The level N is a minimal level for ρ.

Note that the last assumption implies in particular the hypothesis of Proposition 5.2. Hence we will be able to use in our argument that the Galois module $W = J_o(N)[\mathsf{m}]$ defines ρ.

Fix now primes of $\overline{\mathbf{Q}}$ dividing p and q, let D_p and D_q be the corresponding decomposition groups in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, and let $\overline{\mathbf{F}}_p$ and $\overline{\mathbf{F}}_q$ be the corresponding

residue fields. For the moment, we focus attention on the action of D_q on $J_o(N)[\ell]$ and its subgroup W.

We require some results about this action which follow from the fact that $J_o(N)$ has semistable reduction at the prime q. These results are certainly not elementary. Indeed, the semistable reduction that we need follows from the work of Deligne-Rapoport [4] on the reduction of modular curves (see also [16] and [7] for recent, related work), together with results of Raynaud [24] relating the reductions of curves with the reductions of their Jacobians. The information that is to be deduced about $J_o(N)[\ell]$ is contained in Grothendieck's article on Néron models and monodromy [12].

We need first to consider the modular curve $X_o(N/q)$ over the base fields $\mathbf{F}_q$ and $\overline{\mathbf{F}}_q$. For this, it is necessary to invoke the "modular interpretation" of the curves $X_o(N)$ which was discussed above. We view $X_o(N/q)$ over $\mathbf{F}_q$ as classifying pairs (F, C), where F is an elliptic curve and C a cyclic subgroup of order N/q on F. The points of the curve $X_o(N/q)$ over $\overline{\mathbf{F}}_q$ (i.e., with coordinates in $\overline{\mathbf{F}}_q$) consist of a finite number of "cusps" (points at infinity), together with points representing the isomorphism classes of pairs (F, C) over $\overline{\mathbf{F}}_q$. Among these latter points are the *supersingular points*, those represented by pairs (F, C) for which F is supersingular in the sense that the group $F(\overline{\mathbf{F}}_q)[q]$ of points on F of order dividing q is reduced to the trivial group. The number of supersingular points on $X_o(N/q)$ over $\overline{\mathbf{F}}_q$ is finite, and can be calculated quite easily. For example, this number is $1 + g(q)$ when $N = q$ and so is 2 when $N = 11$ and 3 when $N = 23$.

Let $\mathcal{S}$ be the set of these supersingular points. Then $\mathcal{S}$ carries a natural action of $\mathrm{Gal}(\overline{\mathbf{F}}_q/\mathbf{F}_q)$, coming from the action of this Galois group, by conjugation, on pairs (F, C) over $\overline{\mathbf{F}}_q$. We may view this action as an action of D_q which is unramified, i.e., which factors through the quotient of D_q by its inertia subgroup. We will interested in the free abelian group $\mathbf{Z}^{\mathcal{S}}$ on $\mathcal{S}$. This group has a commuting family of Hecke operators $T_n \in \mathrm{End}(\mathcal{S})$ $(n \geq 1)$. These can be defined directly in terms of pairs (F, C) for which F is supersingular. In particular, suppose that $\ell \neq q$ is a prime, and let $s \in \mathcal{S}$ be the class of (F, C). For each cyclic subgroup D of F which has order ℓ and which is not contained in C, consider the quotient elliptic curve F/D and its subgroup $(C \oplus D)/D$, which has order N. The image of s under T_ℓ is the sum of the classes of the various quotients $(F/D, (C \oplus D)/D)$. (There are $\ell + 1$ possible subgroups D when ℓ is prime to N, and ℓ such subgroups otherwise.) On the other hand, the operator T_q is defined on $\mathcal{S}$ as the nega-

tive of the operator $(F, C) \mapsto (F, C)^{(q)}$ which takes each pair to its conjugate under the generator $x \mapsto x^q$ of the Galois group $\mathrm{Gal}(\overline{\mathbf{F}}_q/\mathbf{F}_q)$. The operators T_n with n not necessarily prime can be expressed in terms of the T_ℓ by the same formulas which link together the operators labelled T_n on $S(N)$.

Let L now be the group of elements of $\mathbf{Z}^{\mathcal{S}}$ which have degree 0, the degree of an element being the sum of its coefficients. It is immediate that the operators T_n we have just defined map L into itself. We get in this way an action of the formal polynomial ring $\mathbf{Z}[\ldots, T_n, \ldots]$ on L. By transposition, we obtain an action of this ring on $\mathrm{Hom}(L, \mu_m)$ for every $m \geq 1$, where μ_m is the group of m^{th} roots of unity in $\overline{\mathbf{Q}}_q$. We have a compatible action of the decomposition group $D_q = \mathrm{Gal}(\overline{\mathbf{Q}}_q/\mathbf{Q}_q)$ on $\mathrm{Hom}(L, \mu_m)$ in which D_q acts both on L (in the manner just explained) and on μ_m (in the standard way).

It follows from the work of Grothendieck, Raynaud, and Deligne-Rapoport [12, 24, 4] that there is a natural inclusion

$$\mathrm{Hom}(L, \mu_m) \hookrightarrow J_o(N)(\overline{\mathbf{Q}}_q)[m]$$

which is compatible with the actions of D_q and the T_n ($n \geq 1$) on both sides. The image of this inclusion is known as the *toric part* of $J_o(N)(\overline{\mathbf{Q}}_q)[m]$, and will be denoted $J_o(N)[m]^{\mathsf{t}}$.

This terminology requires some explanation. Consider the abelian variety $J_o(N)$ over $\mathbf{Q}_q$ and imagine the problem of "reducing $J_o(N)$ mod q." In general, the variety $J_o(N)$ has bad reduction at q, which is to say that its "best reduction" (Néron model) over $\mathbf{F}_q$ is not necessarily an abelian variety over $\mathbf{F}_q$. This reduction has, however, a maximal toric subgroup $\mathcal{T}$, and it is the character group $\mathrm{Hom}_{\overline{\mathbf{F}}_q}(\mathcal{T}, \mathbf{G}_{\mathrm{m}})$ of $\mathcal{T}$ which we have called L. The group $\mathrm{Hom}(L, \mu_m)$ naturally extends to a finite flat group scheme over $\mathbf{Z}_q$ whose reduction mod q is the kernel $\mathcal{T}[m]$ of multiplication by m on the $\mathbf{F}_q$-torus $\mathcal{T}$.

The action of $\mathbf{Z}[\ldots, T_n, \ldots]$ which we have defined in an *ad hoc* way is in fact the functorial action of $\mathbf{T} \subseteq \mathrm{End}\Big(J_o(N)\Big)$ on $\mathrm{Hom}_{\overline{\mathbf{F}}_q}(\mathcal{T}, \mathbf{G}_{\mathrm{m}})$. In other words, $\mathbf{Z}[\ldots, T_n, \ldots]$ acts on L through its quotient $\mathbf{T} = \mathbf{T}_N$, and the action is just the functorial one coming from the interpretation of L in terms of the Néron model.

We now set $m = \ell$ and return to consideration of $J_o(N)[\ell]$ and its submodule W. We set $W^{\mathsf{t}} = W \cap J_o(N)[\ell]^{\mathsf{t}}$. Equivalently, we have

$$W^{\mathsf{t}} = \mathrm{Hom}(L/\mathsf{m}L, \mathbf{G}_{\mathrm{m}}),$$

since W is the kernel of m on $J_o(N)[\ell]$. The quotient W/W^{t} is naturally a submodule of $J_o(N)[\ell]/J_o(N)[\ell]^{\mathsf{t}}$. We recall that $\ell \neq q$.

PROPOSITION **5.3** *The D_q-modules W^{t} and W/W^{t} are unramified.*

Proof. Since $\ell \neq q$, the D_q-module μ_ℓ is unramified. The D_q-module L is certainly unramified, since D_q acts on L, by construction, through its quotient $\mathrm{Gal}(\overline{\mathbf{F}}_q/\mathbf{F}_q)$. Hence $J_o(N)[\ell]^{\mathsf{t}}$ is unramified, and we may conclude that its submodule W^{t} is unramified.

Similarly, to prove that W/W^{t} is unramified, it suffices to show that $J_o(N)[\ell]/J_o(N)[\ell]^{\mathsf{t}}$ is unramified. This latter fact follows from Grothendieck's theory of semistable reduction ([12], §2.3–2.4). Indeed, let A be a semistable abelian variety over $\mathbf{Q}_q$ (to fix ideas). Identify the dual $\mathrm{Hom}(A[\ell], \mu_\ell)$ of the group of ℓ-division points on A with the kernel $\check{A}[\ell]$ of multiplication by ℓ on the abelian variety $\check{A}$ dual to A. Then $A[\ell]/A[\ell]^{\mathsf{t}}$ is dual to a subgroup of the largest subgroup $\check{A}[\ell]^{\mathsf{f}}$ of $\check{A}[\ell]$ which extends to a finite flat group scheme over $\mathbf{Z}_q$ (i.e., which is unramified). ∎

COROLLARY *The k-vector space $L/\mathsf{m}L$ has dimension 1.*

Proof. The statement to be proved is equivalent to the statement that W^{t} has k-dimension 1. Since W has dimension 2, we must show that W is isomorphic neither to W^{t} nor to W/W^{t}. By the Proposition, each of these modules is unramified. By hypothesis, however, W is *ramified* at q. ∎

To prove Theorem 5.1, we will derive the inequality

$$\dim_k(L/\mathsf{m}L) \geq 2,$$

which contradicts the above Corollary. To do this, we need to find a second interpretation of the lattice L, or rather of a "primitive" subgroup of it. To define this subgroup, we recall that L is associated to the mod q reduction of the abelian variety $J_o(N)$, where N is a square-free integer divisible by the distinct primes p and q. Write M for the quotient $N/(pq)$, and consider the Jacobian $J_o(Mq)$ of the modular curve $X_o(Mq)$. Let X be the analogue of L for $J_o(Mq)$, i.e., the character group of the largest torus in the mod q reduction of $J_o(Mq)$. There are two natural degeneracy maps

$$X_o(N) \rightrightarrows X_o(Mq),$$

corresponding to the two obvious ways of associating to a pair (F, C), with C cyclic of order Mpq, a pair (F', C') with $C' \subset F'$ cyclic of order Mq. (We can take $F' = F$ and let C' be the unique subgroup of C having order Mq, or else take F' to be the quotient of F by the subgroup C'' of C with order p. In the latter case, we take $C' = C/C''$.) These lead to two maps $L \rightrightarrows X$, or equivalently to a map

$$\pi : L \to X \oplus X.$$

This map is *surjective*, as one sees by applying theorems of Eichler on the arithmetic of quaternion algebras over $\mathbf{Q}$. The kernel of π is the *p-primitive* subgroup of L, and it is this subgroup that we wish to reinterpret.

The reinterpretation comes from the second type of modular curve, a *Shimura curve* coming from a rational quaternion division algebra. Choose a quaternion division algebra of discriminant pq over $\mathbf{Q}$, and let $\mathcal{O}$ be a maximal order in this algebra.

Let C be the algebraic curve over $\mathbf{Q}$ associated to the problem of classifying abelian varieties A of dimension 2, furnished with an action of $\mathcal{O}$, together with a "$\Gamma_o(M)$-structure." The latter object is an $\mathcal{O}$-stable subgroup of A which is isomorphic to $(\mathbf{Z}/M\mathbf{Z})^2$. The curve C is the Shimura curve which we need. It is, in some sense, a variant of the curve $X_o(M)$, since the replacement $\mathcal{O} \mapsto \mathrm{M}(2, \mathbf{Z})$ returns us to the theory of modular curves which we have already met. Indeed, to give an operation of $\mathrm{M}(2, \mathbf{Z})$ on a 2-dimensional abelian surface is essentially the same as to give an elliptic curve: if E is an elliptic curve, then $E \times E$ has an evident action of $\mathrm{M}(2, \mathbf{Z})$, and one sees easily that any A with an action of this ring is isomorphic to such a product. On the other hand, the fact that p and q appear in the discriminant of $\mathcal{O}$ makes $X_o(N) = X_o(pqM)$ a better analogue of C.

Before we can reinterpret the p-primitive subgroup of L, we have two minor choices to make. Namely, the ring $\mathcal{O}$ has residue fields at p and at q which are finite fields of cardinality p^2 and q^2, respectively. Choose isomorphisms κ_p and κ_q between these fields and the quadratic subfields of $\overline{\mathbf{F}}_p$ and $\overline{\mathbf{F}}_q$, respectively. (There are two possible choices for each of p and q.) These maps may be viewed as homomorphisms

$$\kappa_p : \mathcal{O} \to \overline{\mathbf{F}}_p, \qquad \kappa_q : \mathcal{O} \to \overline{\mathbf{F}}_q.$$

Let $\mathcal{T}'$ be the toric part of the mod p reduction of the Néron model of the Jacobian $\mathrm{Pic}^o(C)$. One can show that this torus coincides with the

connected component of 0 in the reduction. This means that $\mathrm{Pic}^o(C)$ has "purely multiplicative reduction" at the prime p. To see that this is the case, one uses results of Grothendieck and Raynaud, cited above, to relate the reduction of $\mathrm{Pic}^o(C)$ to the mod p reduction of C. The latter is studied in well known articles of Cerednik and Drinfeld [3, 6]. (Further discussion of the mod p reduction of $\mathrm{Pic}^o(C)$ is found in Kurihara [17], articles of Jordan-Livné [14, 15], and [26].)

Let $Y = \mathrm{Hom}_{\overline{\mathbf{F}}_p}(\mathcal{T}', \mathbf{G}_{\mathrm{m}})$, so that Y is a free abelian group whose rank is the dimension of $\mathrm{Pic}^o(C)$. Then we have

THEOREM **5.4 ([25, 26])** *The group Y is naturally isomorphic to the p-primitive part of L. Equivalently, we have an exact sequence*

$$0 \to Y \xrightarrow{\theta} L \xrightarrow{\pi} X \oplus X \to 0. \tag{6}$$

The words "naturally isomorphic" are intended to convey the information that the map θ introduced in (6) is compatible with the Hecke operators $T_n \in \mathrm{End}(J_o(N))$ for $n \geq 1$ and their analogues $T_n \in \mathrm{End}(J)$ which were introduced by Shimura. In connection with these operators, it might be worth remarking that the operators T_p and T_q are each *involutions* on the subgroup Y of L. (The operator T_q is already an involution on L.) The map θ is canonical in the sense that it depends only on the choices of κ_p and κ_q. The effect of a change in κ_p or κ_q is to compose θ (on either the left or right) with the corresponding Hecke operator T_p or T_q.

The Hecke operators T_n on J form a commutative family, and thus amount *a priori* to an action of the formal polynomial ring $\mathbf{Z}[\ldots, T_n, \ldots]$ on J. One sees from Theorem 5.4 that this ring acts on Y, at least, through its quotient $\mathbf{T}_N$, since $\mathbf{Z}[\ldots, T_n, \ldots]$ acts on L through this quotient. On the other hand, the fact that J has purely multiplicative reduction at p ensures that $\mathrm{End}(J)$ acts faithfully on Y. Putting these two facts together, we learn that the T_n may be viewed as an action of $\mathbf{T} = \mathbf{T}_N$ on J.

Next, let us view $X \oplus X$ as a $\mathbf{T}$-module via (6). In other words, we use the degeneracy maps introduced above to consider $X \oplus X$ as a quotient of L. Write $(X \oplus X)_{\mathsf{m}}$ for the localization of this $\mathbf{T}$-module at the maximal ideal m of $\mathbf{T}$.

PROPOSITION **5.5** *The module $(X \oplus X)_{\mathsf{m}}$ vanishes.*

To prove the proposition, we use the third assumption made about ρ, namely the supposed minimality of N as a level for ρ. Let $\overline{\mathbf{T}}_N$ be the image of $\mathbf{T}_N$ in $\mathrm{End}(X \oplus X)$, and let $\overline{\mathsf{m}} \subseteq \overline{\mathbf{T}}_N$ be the image of m in this quotient. Clearly, the non-vanishing of $(X \oplus X)_{\mathsf{m}}$ would imply that $\overline{\mathsf{m}}$ is a maximal ideal of $\overline{\mathbf{T}}_N$, rather than the unit ideal (1) of this ring. From this, we would be able to deduce the existence of a maximal ideal $\mathsf{m}_{N/p} \subset \mathbf{T}_{N/p}$ which gives the representation $\rho = \rho_{\mathsf{m}}$, contrary to the supposed minimality. ∎

Here are some details concerning the construction of $\mathsf{m}_{N/p}$: Consider the functorial action of $\mathbf{T}_{N/p}$ on X and the resulting diagonal action of this ring on $X \oplus X$. Let $\overline{\mathbf{T}}_{N/p}$ be the image of $\mathbf{T}_{N/p}$ in $\mathrm{End}(X \oplus X)$. The idea is that the rings $\overline{\mathbf{T}}_N$ and $\overline{\mathbf{T}}_{N/p}$ are essentially identical. More precisely, they share a large common subring R: the subring of $\mathrm{End}(X \oplus X)$ generated by the T_n with n prime to p. (The operator labelled T_n in $\mathbf{T}_N$ and the operator labelled T_n in $\mathbf{T}_{N/p}$ act in the same way on $X \oplus X$, provided that n is prime to p.) Moreover, they both lie in the commutative subring S of $\mathrm{End}(X \oplus X)$ which is generated by R and the two operators labelled T_p: one coming from $\mathbf{T}_N$ and one coming from $\mathbf{T}_{N/p}$. Suppose now that $\overline{\mathsf{m}}$ is a maximal ideal of $\overline{\mathbf{T}}_N$. Then $\overline{\mathsf{m}} \cap R$ is a maximal ideal of R. By using Nakayama's lemma (or the going-up theorem of Cohen-Seidenberg), we may find a maximal ideal $\overline{\mathsf{m}}_{N/p}$ of $\overline{\mathbf{T}}_{N/p}$ whose intersection with R is $\overline{\mathsf{m}} \cap R$ of R. Let $\mathsf{m}_{N/p}$ be the inverse image of $\overline{\mathsf{m}}_{N/p}$ in $\mathbf{T}_{N/p}$. Then the representations ρ and $\rho_{\mathsf{m}_{N/p}}$ are easily seen to coincide in the following strong sense: the residue fields of m, $\overline{\mathsf{m}} \cap R$, and $\mathsf{m}_{N/p}$ are all identical, and the representations ρ and $\rho_{\mathsf{m}_{N/p}}$ are equivalent representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ over this finite field.

COROLLARY *The k-vector space $Y/\mathsf{m}Y$ has dimension 1.*

This assertion is an immediate consequence of the corollary to Proposition 5.3, the exact sequence (6), and Proposition 5.5.

PROPOSITION **5.6** *The kernel $J[\mathsf{m}]$ of m on $J(\overline{\mathbf{Q}})$ is non-zero.*

Proof. Indeed, this kernel, viewed "locally" as the kernel of m on $J(\overline{\mathbf{Q}}_p)$, contains the toric subgroup

$$J[\mathsf{m}]^{\mathsf{t}} = \mathrm{Hom}(Y/\mathsf{m}Y, \mu_\ell).$$

The above corollary shows that this subgroup is of dimension 1, and in particular that it is non-zero. ∎

PROPOSITION **5.7** *The group $J[\mathsf{m}]$, viewed as a $k[\mathrm{Gal}]$-module, contains a submodule isomorphic to the module V.*

Proof. We recall that V is by definition a k-vector space of dimension 2 with an action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ which is equivalent to ρ. By using the Eichler-Shimura relations for J, and the argument referred to in the proof of Proposition 5.2, we may deduce that the semisimplification of $J[\mathsf{m}]$ is isomorphic to a direct sum of copies of V. The desired result follows from this fact, plus the non-vanishing of $J[\mathsf{m}]$. ∎

We again take the "local" point of view which amounts to viewing $J[\ell]$ and its submodules V, $J[\mathsf{m}]$, ... as D-modules, where $D = D_p$ is a decomposition group for p in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. As in [12], we let $J[\ell]^{\mathsf{f}}$ be the largest subgroup of $J[\ell]$ which extends to a finite flat group scheme over $\mathbf{Z}_p$. (Note that the case $p = \ell$ is not excluded in our situation, so that "finite" cannot automatically be equated with "unramified.") We have $J[\ell]^{\mathsf{t}} \subseteq J[\ell]^{\mathsf{f}}$, where $J[\ell]^{\mathsf{t}}$ is a subgroup of $J[\ell]$ isomorphic to $\mathrm{Hom}(Y/\ell Y, \mu_\ell)$. The quotient $J[\ell]^{\mathsf{f}}/J[\ell]^{\mathsf{t}}$ may be identified with the kernel $\Psi[\ell]$ of multiplication by ℓ on the group Ψ of components in the mod p reduction of J, since J has purely multiplicative reduction at p (cf. [12], 11.6.12). By hypothesis, V is finite at p, so that we have $V \subseteq J[\ell]^{\mathsf{f}}$. Equivalently, the k-vector space $V^{\mathsf{f}} = V \cap J[\ell]^{\mathsf{f}}$ is of dimension 2. (It coincides with V.) On the other hand, the intersection $V^{\mathsf{t}} = V \cap J[\ell]^{\mathsf{t}}$ is contained in the group $\mathrm{Hom}(Y/\mathsf{m}Y, \mu_\ell)$, which is of dimension 1 over k by the Corollary to Proposition 5.5. Therefore, $V^{\mathsf{f}}/V^{\mathsf{t}}$ is non-zero. This gives in particular the statement:

$$\left(J[\ell]^{\mathsf{f}}/J[\ell]^{\mathsf{t}}\right)[\mathsf{m}] \neq 0.$$

Equivalently, we have

PROPOSITION **5.8** *The group $\Psi[\mathsf{m}]$ is non-zero.*

This Proposition will lead to the final contradiction, showing the incompatibility of the three assumptions made above. As explained earlier, this incompatibility implies Theorem 5.1. The principal point is that the assertion of Proposition 5.8 implies that $\rho = \rho_{\mathsf{m}}$ is modular of level N/p, contrary to the assumption that N is a minimal level for ρ. Indeed, the author shows in [25] that there is a map

$$F : X \oplus X \to \Psi$$

which is, for practical purposes, surjective. Thus, if $\Psi[\mathsf{m}]$ is non-zero, then $(X \oplus X)_{\mathsf{m}}$ is non-zero, which contradicts Proposition 5.5.

The existence of F is implicit in [15], which exhibits a quantitative relation between Ψ and a quotient of $X \oplus X$. To construct F, the author uses the exact sequence (6), plus a description of Ψ in terms of Y which is due to Grothendieck [12]. Namely, there is a bilinear pairing

$$(,) : Y \times Y \to \mathbf{Z},$$

the *monodromy pairing* for J mod p, such that Ψ is canonically the cokernel of the map

$$Y \to \mathrm{Hom}(Y, \mathbf{Z}), \qquad y \mapsto (y, \cdot).$$

This pairing is defined in terms of the mod p reduction of the curve C. An analysis similar to that which establishes (6) shows that the monodromy pairing on Y is the restriction to Y of an analogous pairing on L, which in turn is the restriction to L of an explicit diagonal pairing on $\mathbf{Z}^{\mathcal{S}}$, where $\mathcal{S}$ is again the set of supersingular points of $X_o(N/q)$ over $\overline{\mathbf{F}}_q$. This means, in particular, that Ψ may be calculated in concrete terms.

In [25], the author inserts Ψ into a commutative diagram of exact sequences and constructs F by appealing to the Snake Lemma. The resulting long exact sequence gives a description of the cokernel of F, which turns out to be the a quotient of Φ, the group analogous to Ψ which is associated to the mod q reduction of $J_o(N)$. One should consider F to be "essentially surjective" because Φ is a group which can be ignored when studying irreducible two-dimensional representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. The point is that Φ is *Eisenstein* in the sense that the relation $T_r = (r+1)$ holds on Φ, for all primes r which do not divide N ([25, 27, 8]). This implies easily that the support of Φ as a $\mathbf{T}$-module is contained in the set of maximal ideals for which the corresponding two-dimensional representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ are reducible. In particular, $\Phi_{\mathsf{m}} = 0$.

References

[1] Atkin, A.O.L. and Lehner, J. Hecke operators on $\Gamma_o(m)$. Math. Ann. **185**, 134–160 (1970)

[2] Carayol, H. Sur la mauvaise réduction des courbes de Shimura. Compositio Math. **59**, 151–230 (1986)

[3] Cerednik, I.V. Uniformization of algebraic curves by discrete arithmetic subgroups of $\mathrm{PGL}_2(k_w)$ with compact quotients (in Russian). Mat. Sb. **100** , 59–88 (1976). Translation in Math USSR Sb. **29**, 55–78 (1976)

[4] Deligne, P. and Rapoport, M. Les schémas de modules de courbes elliptiques. Lecture Notes in Math. **349**, 143–316 (1973)

[5] Deligne, P. and Serre, J-P. Formes modulaires de poids 1. Ann. Sci. Ec. Norm. Sup. **7**, 507–530 (1974)

[6] Drinfeld, V.G. Coverings of p-adic symmetric regions (in Russian). Functional. Anal. i Prilozen. **10**, 29–40 (1976). Translation in Funct. Anal. Appl. **10**, 107–115 (1976)

[7] Edixhoven, S. J. Minimal resolution and stable reduction of $X_o(N)$. To appear

[8] Edixhoven, S. J. L'action de l'algèbre de Hecke sur les groupes de composantes des jacobiennes des courbes modulaires est "Eisenstein". To appear

[9] Faltings, G. Endlichkeitssätze für abelsche Varietäten über Zahlkörpern. Invent. Math **73**, 349–366 (1983). English translation in: Arithmetic Geometry, G. Cornell and J. H. Silverman, eds. New York: Springer-Verlag (1986)

[10] Frey, G. Links between stable elliptic curves and certain diophantine equations. Annales Universitatis Saraviensis **1**, 1–40 (1986)

[11] Frey, G. Links between solutions of $A - B = C$ and elliptic curves. Lecture Notes in Math. **1380**, 31–62 (1989)

[12] Grothendieck, A. SGA7 I, Exposé IX. Lecture Notes in Math. **288**, 313–523 (1972)

[13] Husemöller, D. Elliptic Curves. Graduate Texts in Mathematics **111**. New York: Springer-Verlag (1987)

[14] Jordan, B. and Livné, R. Local diophantine properties of Shimura curves. Math. Ann. **270**, 235–248 (1985)

[15] Jordan, B. and Livné, R. On the Néron model of Jacobians of Shimura curves. Compositio Math. **60**, 227–236 (1986)

[16] Katz, N. M. and Mazur, B. Arithmetic Moduli of Elliptic Curves. Annals of Math. Studies **108**. Princeton: Princeton University Press, 1985

[17] Kurihara, A. On some examples of equations defining Shimura curves and the Mumford uniformization. J. Fac. Sci. Univ. Tokyo, Sec. IA, **25**, 277–300 (1979)

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

[19] Mazur, B. Number theory as gadfly. American Math. Monthly. To appear

[20] Mazur, B. and Ribet, K. Two-dimensional representations in the arithmetic of modular curves. To appear

[21] Oesterlé, J. Nouvelles approches du "théorème" de Fermat. Séminaire Bourbaki n° 694 (1987–88). Astérisque **161–162**, 165–186 (1988)

[22] Ogg, A. Rational points on certain elliptic modular curves. Proc. Symp. Pure Math. **24**, 221–231 (1973)

[23] Ogg, A. Hyperelliptic modular curves. Bull. SMF **102**, 449–462 (1974)

[24] Raynaud, M. Spécialisation du foncteur de Picard. Publ. Math. IHES **38**, 27–76 (1970)

[25] Ribet, K. On modular representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ arising from modular forms. Preprint

[26] Ribet, K. Bimodules and abelian surfaces. Advanced Studies in Pure Mathematics. In press

[27] Ribet, K. On the component groups and the Shimura subgroup of $J_o(N)$. Sém. Th. Nombres, Université Bordeaux, 1987–88

[28] Serre, J-P. Abelian ℓ-adic Representations and Elliptic Curves. New York: W.A. Benjamin 1968. (Republished 1989 by Addison-Wesley)

[29] Serre, J-P. Propriétés galoisiennes des points d'ordre fini des courbes elliptiques. Invent. Math. **15**, 259–331 (1972)

[30] Serre, J-P. Sur les représentations modulaires de degré 2 de Gal($\overline{\mathbf{Q}}/\mathbf{Q}$). Duke Math. J. **54**, 179–230 (1987)

[31] Serre, J-P. and Tate, J. Good reduction of abelian varieties. Ann. of Math. **88**, 492–517 (1968)

[32] Shimura, G. Introduction to the Arithmetic Theory of Automorphic Functions. Princeton: Princeton University Press 1971

[33] Shimura, G. On elliptic curves with complex multiplication as factors of the Jacobians of modular function fields. Nagoya Math. J. **43**, 199–208 (1971)

[34] Silverman, J. The Arithmetic of Elliptic Curves. Graduate Texts in Mathematics **106**. New York: Springer-Verlag (1986)

[35] Weil, A. Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen. Math. Ann. **168**, 165–172 (1967)

K. A. Ribet
Mathematics Department
University of California
Berkeley CA 94720
USA