REPORT ON MOD ℓ REPRESENTATIONS OF $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$

Kenneth A. Ribet

University of California, Berkeley

In memory of Kenneth F. Ireland

1. Introduction.

Let $N \geq 1$ and $k \geq 2$ be integers. Let $\Gamma_1(N)$ be the group

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{SL}(2, \mathbf{Z}) \,\middle|\, c \equiv 0,\ a \equiv d \equiv 1 \bmod N \right\}$$

and let $S = S_k(\Gamma_1(N))$ be the complex vector space of cusp forms of weight k on $\Gamma_1(N)$. There is a standard action $d \mapsto \langle d \rangle \in \operatorname{Aut} S$ of the group $(\mathbf{Z}/N\mathbf{Z})^*$ on S. In addition, S comes equipped with a family of Hecke operators T_n $(n \geq 1)$; the T_n commute with each other and with the "diamond bracket" operators $\langle d \rangle$. These operators are traditionally written on the right; for example, the n^{th} Hecke operator is normally written $f \mapsto f|T_n$.

An *eigenform* in S is a non-zero cusp form $f \in S$ which is an eigenvector for each of the operators T_n and $\langle d \rangle$. In particular, if f is an eigenform, then there is a Dirichlet character ϵ defined mod N such that $f|\langle d \rangle = \epsilon(d) f$ for $d \in (\mathbf{Z}/N\mathbf{Z})^*$. The functional equation satisfied by a weight-k cusp form implies that this "Nebentypus" character has the same parity as k: $\epsilon(-1) = (-1)^k$. The eigenvalues a_n of the T_n acting on f are algebraic integers, and generate together a finite extension of $\mathbf{Q}$.

Let $\overline{\mathbf{Q}}$ be an algebraic closure of the field of rational numbers, and consider the Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Suppose that $f \in S$ is an eigenform, and let E be a number field containing the associated eigenvalues a_n and $\epsilon(d)$. A construction of Deligne [10] attaches to f a family of continuous representations

$$\rho_\lambda\colon\ \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, E_\lambda);$$

here, λ runs over the set of prime ideals of the integer ring of E, and E_λ denotes the completion of E at λ. Each representation ρ_λ is irreducible; it is characterized up to isomorphism by the identities

$$\operatorname{trace}(\rho_\lambda(\mathrm{Frob}_p)) = a_p, \qquad \det(\rho_\lambda(\mathrm{Frob}_p)) = \epsilon(p) p^{k-1}$$

1991 *Mathematics Subject Classification.* Primary 11F11, 11R32; Secondary 11F33, 11G18.

This paper is in final form and no version of it will be submitted for publication elsewhere. The author was supported in part by NSF Grant #DMS 88-06815. He wishes to thank both the I.H.E.S. and the Université de Paris 7 for gracious hospitality during the period when this article was begun. Thanks are due to Fred Diamond, Bas Edixhoven and Barry Mazur for helpful comments. The referee deserves special thanks for a thorough reading of the manuscript and for many useful comments and suggestions.

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

for prime numbers p which are prime both to N and to the norm of λ. The symbol Frob_p denotes an arithmetic Frobenius element for p in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. (For a survey of this material, the reader may wish to consult [37].)

The reduction of ρ_λ mod λ is a certain semisimple representation

$$\bar{\rho}_\lambda\colon\ \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F}_\lambda),$$

where $\mathbf{F}_\lambda$ is the residue field of λ. Its construction involves a choice: one chooses a model of ρ_λ which has values in $\mathbf{GL}(2, \mathcal{O}_\lambda)$, where $\mathcal{O}_\lambda$ is the ring of integers of E_λ, and then forms the "naive reduction" of this model, i.e., the composite of ρ_λ and the canonical map $\mathbf{GL}(2, \mathcal{O}_\lambda) \to \mathbf{GL}(2, \mathbf{F}_\lambda)$. The *semisimplification* of this naive reduction depends only on ρ_λ. It is the desired representation $\bar{\rho}_\lambda$.

Consider the representations $\bar{\rho}_\lambda$ obtained for varying N, k, f, but with $\mathbf{F}_\lambda$ having a fixed characteristic. To assemble them, fix a prime number ℓ, and choose a place v dividing ℓ of the field of algebraic numbers in $\mathbf{C}$. Let $\mathbf{F}$ be the residue field of v, so that $\mathbf{F}$ is an algebraic closure of its prime field $\mathbf{F}_\ell$. For each algebraic number field E contained in $\mathbf{C}$, v induces a prime λ on E, together with a canonical inclusion $\mathbf{F}_\lambda \hookrightarrow \mathbf{F}$. In particular, if E is the field generated by the eigenvalues attached to an eigenform f, and λ is the prime of E induced by v, the representation $\bar{\rho}_\lambda$ may be viewed as taking values in $\mathbf{GL}(2, \mathbf{F})$.

Suppose that $\rho\colon\ \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ is a continuous semisimple representation. When can we expect that ρ is isomorphic to one of the $\bar{\rho}_\lambda$? Identify ϵ with a character defined on $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, and let $\tilde{\chi}$ be the ℓ-adic cyclotomic character. Recall that the identity $\det(\rho_\lambda(\mathrm{Frob}_p)) = \epsilon(p)p^{k-1}$ on Frobenius elements implies the formula

$$\det \rho_\lambda = \epsilon \tilde{\chi}^{k-1} \tag{1.1}$$

for $\det \rho_\lambda$. On reducing this identity, one obtains $\det \bar{\rho}_\lambda = \bar{\epsilon}\chi^{k-1}$, where χ is the mod ℓ cyclotomic character, and $\bar{\epsilon}$ is the reduction of ϵ mod λ (i.e., mod v). In particular, let $c \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ be a complex conjugation. Then $\det \bar{\rho}_\lambda(c) = \bar{\epsilon}(c)\chi(c)^{k-1}$. Now $\chi(c) = -1$, and furthermore $\epsilon(c)$ is another name for $\epsilon(-1)$. Using the formula $\epsilon(-1) = (-1)^k$, we obtain $\det \bar{\rho}_\lambda(c) = -1$. In other words, $\bar{\rho}$ must be an *odd* representation if it is to be isomorphic to some $\bar{\rho}_\lambda$.

Serre [49] has proposed that this parity condition represents a sufficient, as well as a necessary, condition for a representation $\rho\colon\ \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ to arise from some eigenform. The *reducible* odd representations ρ arise from Eisenstein series, and will be neglected in the following discussion. For *irreducible* representations, we have the

(1.2) Serre Conjecture [49, (3.2.3?)]. *Let $\rho\colon\ \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ be an irreducible, odd representation. Then ρ is isomorphic to the representation $\bar{\rho}_\lambda$ associated to some eigenform in one of the spaces $S_k(\Gamma_1(N))$.*

The associated **Refined Conjecture** [49, (3.2.4?)] asserts that ρ is a representation $\bar{\rho}_\lambda$ associated to an eigenform in a *specific* space $S_{k(\rho)}(\Gamma_1(N(\rho)))$. The invariants $k(\rho)$ and $N(\rho)$ are defined by local properties of ρ. More precisely, as we recall in §2, $N(\rho)$ is an integer prime to ℓ which depends only on the restrictions of ρ to decomposition groups in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ for each prime number $p \neq \ell$. Similarly, $k(\rho)$ is an integer ≥ 2 which depends only on the restriction of ρ to a decomposition group for ℓ in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. (See [49, §2].)

The formulation of [49, (3.2.4?)] specifies that ρ arise from an eigenform $f \in S_{k(\rho)}(\Gamma_1(N(\rho)))$ whose character ϵ is also predicted in advance. This extra requirement means simply that ϵ can be chosen to be of order prime to ℓ. Subsequently, however, Serre found examples for $\ell \leq 3$ showing that the requirement cannot always be satisfied [50]; his examples concern the two-dimensional space $S_2(\Gamma_1(13))$. On the other hand, Carayol [7, Prop. 3] and Serre have each shown that the situation for $\ell \geq 5$ is much more favorable: the representations $\bar{\rho}_\lambda$ arising from a given space $S_k(\Gamma_1(N))$ all arise from eigenforms whose associated character has prime-to-ℓ order.

More precisely, Carayol proves the following result [7, Proposition 3]:

(1.3) Theorem. *Assume that $\ell \geq 5$ and that $\rho\colon$ Gal$(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ is an odd irreducible representation. Suppose that ρ arises from an eigenform $f \in S_k(\Gamma_1(N))$ and that ϵ is the Dirichlet character which is associated to f. Let ϵ' be a character on $(\mathbf{Z}/N\mathbf{Z})^*$ which is congruent to ϵ mod v. Then ρ arises from an eigenform $f' \in S_k(\Gamma_1(N))$ whose Nebentypus character is ϵ'.*

In view of this result, and the counterexamples for $\ell = 2$ and $\ell = 3$, we will restrict attention to the invariants $N(\rho)$ and $k(\rho)$ in discussing Serre's conjectures.

The Refined Conjecture has a number of striking consequences, which are catalogued in §4 of [49]. Among these are Fermat's Last "Theorem," the conjecture of Taniyama that all elliptic curves over $\mathbf{Q}$ are modular, and variants of Taniyama's conjecture concerning other motives of rank 2.

To bridge the gap between the Serre Conjecture and its refinement, we formulate the following

(1.4) Motivating Problem. *Suppose that $\rho\colon$ Gal$(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ is an irreducible representation which arises from some space of cusp forms $S_k(\Gamma_1(N))$. Does ρ arise from an eigenform in $S_{k(\rho)}(\Gamma_1(N(\rho)))$, where $k(\rho)$ and $N(\rho)$ are the invariants assigned to ρ by Serre?*

An affirmative solution to (1.4) would imply an equivalence between Serre's conjecture and its refinement. At the time of this writing, one is very close to a complete solution of (1.4), at least when ℓ is an odd prime. The solution involves the work of a large number of mathematicians, including N. Boston, H. Carayol, F. Diamond, B. Edixhoven, G. Faltings, B. H. Gross, B. Jordan, H. W. Lenstra, Jr., R. Livné, B. Mazur, J-P. Serre, and the author.

The goal of this article is twofold. On the one hand, we shall describe what is known about (1.4). On the other, we shall present the following new result:

(1.5) Theorem. *Let $\ell \geq 3$ be a prime, and let $\rho\colon$ Gal$(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ be irreducible. Suppose that ρ arises from an eigenform f of weight two and trivial character on $\Gamma_1(M) \cap \Gamma_o(p)$, where p is prime to ℓM. Assume that ρ is unramified at p. Then ρ arises from a weight-two eigenform with trivial character on $\Gamma_1(M)$.*

The new feature of this theorem is that the level M is not required to be prime to ℓ. In fact, M can be divisible by an arbitrarily high power of ℓ. As it stands, (1.5) applies only to forms with trivial character on $\Gamma_1(M) \cap \Gamma_o(p)$, i.e., to forms on $\Gamma_o(Mp)$. However, it is very likely that the proof we give for (1.5) will generalize without difficulty to cover eigenforms on $\Gamma_1(M) \cap \Gamma_o(p)$ whose Nebentypus characters are *arbitrary* characters mod M. The discussion below shows that (1.4) will be solved for primes $\ell \geq 5$ as soon as Theorem 1.5 is generalized to this case.

One can also contemplate a variant of (1.5) which applies to forms of weight k on $\Gamma_1(M) \cap \Gamma_o(p)$, where k satisfies $2 \leq k \leq \ell{+}1$ and M is required to be prime to ℓ. Such a variant, which accepts arbitrary characters mod M, will again lead to a solution of (1.4). It seems very likely that the methods of Faltings, Jordan and Livné [16], [24] will lead to a variant of this type.

In light of the substantial progress made in relating Serre's conjecture to its refinement, it is striking that there is little to report concerning the conjecture itself [49, 3.2.3$_?$], aside from the numerical evidence presented in §5 of [49].

2. Stripping powers of ℓ from the level.

Let ρ: $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ be an irreducible odd representation, as above, and let $N(\rho)$ and $k(\rho)$ be the invariants assigned to ρ by Serre. Recall [49, §1] that $N(\rho)$ is a product $\prod p^{n(p,\rho)}$ extended over the set of prime numbers $p \neq \ell$. The integer $n(p, \rho)$ is defined by restricting ρ to a decomposition group D_p for p in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Consider the ramification subgroup G_0 of D_p, along with the higher ramification groups G_i $(i > 0)$. Let V be a two-dimensional $\mathbf{F}$-vector space affording the representation ρ. For each $i \geq 0$, let V_i be the subspace of V consisting of those $v \in V$ which are fixed by all elements of G_i. Then

$$n(p, \rho) = \sum_{i \geq 0} \frac{1}{(G_0 : G_i)} \dim V/V_i.$$

In particular, $N(\rho)$ is prime to ℓ, so that Serre's refined conjecture implies the weaker statement that ρ arises from some space of cusp forms $S_k(\Gamma_1(N))$, with N prime to ℓ. Let us say simply that ρ arises from a subgroup Γ of $\mathbf{SL}(2, \mathbf{Z})$ if it arises from the space of weight-k cusp forms on Γ, for *some* $k \geq 2$. Then the weak statement implied by Serre's conjecture is that ρ arises from the group $\Gamma_1(N)$ for some N which is prime to ℓ. We shall now show (at least when $\ell \geq 3$) that any ρ which is modular in the sense that it arises from some $\Gamma_1(N)$ with N not necessarily prime to ℓ is automatically modular in the apparently stronger sense that it arises from $\Gamma_1(N)$ for some N which is prime to ℓ. This is certainly a "well known" fact, which is close to results which are already in the literature, cf. [49, *Remarque*, p. 195].

(2.1) Theorem. *Assume $\ell \geq 3$. Suppose that ρ arises from $\Gamma_1(M)$, where M is the product $N\ell^\alpha$ with $(N, \ell) = 1$. Then ρ arises from $\Gamma_1(N)$.*

We prove (2.1) using the concrete techniques of Serre [46, §3], [47, Theorem 5.4] and Queen [35, §3]. An alternative method would be based on results of N. Katz: see the appendices of [25] and the discussions in [20, §1] and [21, §1]. (Thanks are due to Fred Diamond for pointing out these references.) Note that the assumption $\ell \geq 3$ is made principally for convenience; it should be possible to analyze the case $\ell = 2$ without great difficulty.

Our proof consists of several independent steps. Before beginning it, we shall normalize the eigenforms which occur in the proof. Recall that if f satisfies $f|T_n = a_n f$ for all $n \geq 1$, then we may scale f by a constant so that its Fourier coefficients are the eigenvalues a_n. We will assume that all eigenforms are scaled in this way, i.e., that they are "normalized eigenforms." We represent these eigenforms as series $\sum_{n \geq 1} a_n q^n$; the variable q initially represents $e^{2\pi i \tau}$, where τ is an element of the complex upper half plane, but rapidly becomes a formal variable. In particular, we say that two forms are congruent mod ℓ or mod v if their Fourier series are formally congruent. Similarly, the mod v reduction of an eigenform f is meant to be the formal power series $\sum \bar{a}_n q^n$, where "$^-$" denotes the reduction map mod v.

Step 1. *The representation ρ arises from $\Gamma_o(\ell^r) \cap \Gamma_1(\ell N)$ for some $r \geq 0$.*

When $\ell \geq 5$, the assertion follows from (1.3). In that case, we can take $r = \alpha$. It can be proved in an elementary manner, for all $\ell \geq 3$, by the following argument.

Suppose that ρ arises from an eigenform f in $S_k(\Gamma_1(M))$ and that κ is the associated Dirichlet character mod M. We assume that $k \geq 2$ and that α is positive (i.e., that M is divisible by ℓ). Decompose κ as a product $\epsilon\eta\omega^i$, where: ϵ has conductor dividing N, η has ℓ-power order and ℓ-power conductor, and ω is the "Teichmüller" character, i.e., that character of conductor ℓ and order $\ell{-}1$ which is congruent to the identity function mod v. The exponent i is naturally an integer mod $\ell{-}1$. Because η has, in particular, odd order, we may write it in the form ξ^{-2}, where ξ is a character of ℓ-power order. The cusp form $f \otimes \xi$ is a form of type $(k, \epsilon\omega^i)$ (i.e., weight k and character $\epsilon\omega^i$) which gives rise to ρ. If $r \geq \alpha$ is large enough so that ℓ^r is divisible by the square of the conductor of ξ, then $f \otimes \xi$ is again of level $N\ell^r$.

In what follows, we fix an $r > 0$ so that ρ arises from $\Gamma_o(\ell^r) \cap \Gamma_1(\ell N)$.

Step 2. *ρ arises from $\Gamma_o(\ell^r) \cap \Gamma_1(N)$.*

We assume as above that ρ arises from an eigenform of type $(k, \epsilon\omega^i)$ on the group $\Gamma_o(\ell^r) \cap \Gamma_1(\ell N)$. Take the exponent i to be a positive integer. The Eisenstein series

$$G := L(1{-}i, \omega^{-i})/2 + \sum_{n=1}^{\infty} \left(\sum_{d|n} \omega^{-i}(d) d^{i-1} \right) q^n$$

is then of type (i, ω^{-i}) on $\Gamma_o(\ell)$, as proved (for example) in [46, Lemme 10]. Thus fG is on the group $\Gamma_o(\ell^r) \cap \Gamma_1(N)$, if r is positive. On the other hand, it is well known that the order at v of $L(1{-}i, \omega^{-i})$ is negative: this follows from the Kummer congruences and the Von Staudt theorem. (See, e.g., [22, §3.4].) In other words, if c is the constant term of G, then $E := c^{-1}G$ is a form with v-integral coefficients such that $E \equiv 1 \bmod v$. Consequently, the product fE, viewed mod v, is a non-zero eigenform with the same eigenvalues as f. At the same time, fE is on the group $\Gamma_o(\ell^r) \cap \Gamma_1(N)$. By a well known lemma [12, Lemme 6.11], we may find an eigenform on $\Gamma_o(\ell^r) \cap \Gamma_1(N)$ whose eigenvalues are congruent to those of f.

Step 3. *ρ arises from $\Gamma_o(\ell) \cap \Gamma_1(N)$.*

Assume that ρ arises from an eigenform $f = \sum a_n q^n$ on $\Gamma_o(\ell^r) \cap \Gamma_1(N)$, with $r > 1$. Let K be a finite Galois extension of $\mathbf{Q}$ containing the a_n, and let $\sigma \in \mathrm{Gal}(K/\mathbf{Q})$ be a Frobenius element for v. Thus $\sigma a \equiv a^\ell \bmod v$ for all elements a of the ring of integers of K. The series $\sum \sigma^{-1} a_n q^n$ is the Fourier expansion of a normalized eigenform $\sigma^{-1} f$ of the same weight as f. We wish to show that f is congruent mod v to a cusp form of some weight on $\Gamma_o(\ell^{r-1}) \cap \Gamma_1(N)$. Consider $g := (\sigma^{-1} f)^\ell | U$, where U is the ℓ^{th} Hecke operator T_ℓ. On the one hand, g is a form on $\Gamma_o(\ell^{r-1}) \cap \Gamma_1(N)$ (see, e.g., [29, Lemma 1]). On the other hand, the Fourier expansion of g is congruent mod v to $\sum (\sigma^{-1} a_n)^\ell q^n$, which in turn is congruent to the Fourier expansion of f.

Step 4. *ρ arises from $\Gamma_1(N)$.*

To prove this, we use the argument given by Serre in [46, §3.2], replacing the operators W, U, V, ... by their analogues for forms on $\Gamma_1(N) \cap \Gamma_o(\ell)$. More precisely, we let U be the ℓ^{th} Hecke operator, as above, and let V be the operator $\sum a_n q^n \mapsto \sum a_n q^{\ell n}$, which takes forms on $\Gamma_1(N)$ to forms of the same weight on $\Gamma_1(N) \cap \Gamma_o(\ell)$. Further, we let W be the operator $V_\ell^{N\ell}$ as defined in §1 of [29]. Thus W is given by the matrix $\begin{pmatrix} \ell x & y \\ N\ell z & \ell \end{pmatrix}$, where x, y, and z are integers and we have $\ell x - Nyz = 1$. This operator is the inverse of the operator denoted W_Q in [1]. Indeed, by [1, Proposition 1.1], one has $W = \epsilon(\ell) W_Q$ on the space of forms with character ϵ on $\Gamma_1(N) \cap \Gamma_o(\ell)$. (We consider ϵ as a Dirichlet character mod N, which permits us to evaluate it at ℓ.) At the same time, W^2 is multiplication by $\epsilon(\ell)$, according to Lemma 2 of [29].

Next, suppose that F is a form of weight w and character ϵ on $\Gamma_1(N) \cap \Gamma_o(\ell)$. Then [29, Lemma 3] implies that the form

$$\operatorname{Tr}(F) := F + \epsilon^{-1}(\ell)\ell^{1-w/2} F|W|U$$

is a form of weight w on $\Gamma_1(N)$. (Apply [29, Lemma 3] to $F|W$.) Further, if G is a form of weight w and character ϵ on $\Gamma_1(N)$, one has $G|W = \ell^{w/2}\epsilon(\ell) G|V$ [1, Prop. 1.5].

If $\ell > 3$, let E be the normalized Eisenstein series of weight $\ell - 1$ on $\mathbf{SL}(2, \mathbf{Z})$, so that we have $E \equiv 1 \bmod \ell$. Similarly, if $\ell = 3$ let E be the normalized Eisenstein series E_4 of weight four. Let a denote the weight of E. As in [46], we introduce the form

$$g := E - \ell^{a/2} E|W = E - \ell^a E|V.$$

One has $g \equiv 1 \bmod \ell$; furthermore, $g|W$ is divisible by a positive power of ℓ (in fact, by $\ell^{1+a/2}$).

Let f be an eigenform giving rise to ρ, as above, and consider $\operatorname{Tr}(fg^i)$, where i is a positive integer. The modular form $\operatorname{Tr}(fg^i)$ is a form on $\Gamma_1(N)$. On the other hand, a calculation similar to that of §3.2 of [46] shows that $\operatorname{Tr}(fg^i) \equiv f \bmod \ell$ for sufficiently large i. ■

A partial converse to (2.1) is the following result:

(2.2) Theorem. *Suppose that ρ arises from $S_k(\Gamma_1(N))$, with N prime to ℓ and $2 \le k \le \ell + 1$. Assume that $\ell > 3$ or that $N > 3$. Then ρ arises from $S_2(\Gamma_1(N\ell))$.*

For $\ell \ge 5$, the theorem is proved by Ash and Stevens [2, Theorem 3.5a] by methods involving parabolic cohomology. Another approach, which originated with Serre, is worked out in Gross's article [17, Proposition 9.3]. Although Gross's prime p (which corresponds to our ℓ) is arbitrary, he assumes that the level N satisfies $N \ge 4$. According to comments in §10 of [17], Proposition 9.3 of [17] extends to the case where $N \le 2$ and k is even.

3. Adjustment of the weight.

Let $\rho\colon \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ be an irreducible representation, and let $k(\rho)$ be the "weight" attached to ρ by Serre [49, §2]. Edixhoven [15, §4] proved that if ρ arises from a cusp form of some weight, then ρ arises from a cusp form of weight $k(\rho)$. More precisely, [15, Th. 4.3] gives the following result:

(3.1) Theorem. *Let N be prime to ℓ, and assume that ρ arises from an eigenform $f \in S_k(\Gamma_1(N))$ with character ϵ. Then ρ arises from an eigenform $f' \in S_{k(\rho)}(\Gamma_1(N))$ with character ϵ. The integers k and $k(\rho)$ are congruent mod $\ell-1$; moreover, we have $k \geq k(\rho)$ provided that ℓ is odd.*

The proof of (3.1) relies on two points which merit independent discussion. First, let D_ℓ be a decomposition group for ℓ in $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ and let I_ℓ be the inertia subgroup of D_ℓ. Suppose that

$$\rho\colon\ \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}\, V \approx \mathbf{GL}(2, \mathbf{F})$$

arises from an eigenform $f = \sum a_n q^n$ of weight k and character ϵ on $\Gamma_1(N)$, with N prime to ℓ. The proof of (3.1) requires information about the restriction of ρ to D_ℓ, in the case where k satisfies $2 \leq k \leq \ell$.

Specifically, suppose that this inequality is satisfied, and that f is ordinary in the sense that $a_\ell \not\equiv 0 \bmod v$. A theorem of Deligne (proved in [17]) states that V has an unramified quotient V/V_0 on which Frob_ℓ acts by multiplication by a_ℓ (or, more precisely, by the image of a_ℓ in $\mathbf{F}$). The action of D_ℓ on V_0 is determined by this information, since the determinant of ρ may be computed from k and ϵ. On the other hand, suppose that f is supersingular, i.e., not ordinary. Let ψ and ψ' be the two fundamental characters $I_\ell \to \mathbf{F}$ of level 2, as defined in [44, §1.7]. A theorem of Fontaine states that the restriction of ρ to I_ℓ is the direct sum of two one-dimensional representations, on which I_ℓ operates by the $k-1^{\mathrm{st}}$ powers of ψ and ψ'. This theorem was proved by Fontaine in an exchange of letters with Serre in 1979; a new proof was given by Edixhoven in [15].

Secondly, Edixhoven's proof requires B. H. Gross's result about companion forms [17], which conversely may be viewed as a special case of (3.1). Before stating Gross's result, we recall that there is an operator $\theta = q\frac{d}{dq}$ on mod ℓ modular forms for $\Gamma_1(N)$ when N is prime to ℓ. This operator maps forms of weight k to forms of weight $k+\ell+1$. It was considered initially by Serre and Swinnerton-Dyer [45], [53] for forms on $\mathbf{SL}(2, \mathbf{Z})$, and then constructed by Katz [26] for forms of full level N, where $N \geq 3$. It is introduced in [17,§4] in the case of forms on $\Gamma_1(N)$, with $N \geq 4$. According to Edixhoven [15, §2.1], mod ℓ modular forms of weight k on $\Gamma_1(N)$ are $\mathbf{GL}(2, \mathbf{Z}/n\mathbf{Z})$-invariant sections of the line bundle ω^k on the curve X. Here $n \geq 3$ is prime to ℓ, and X is the modular curve parametrizing elliptic curves with a $(\Gamma_1(N), \Gamma(n))$-structure. The integer N is assumed only to be prime to ℓ. Using this fact, Edixhoven constructs θ for modular forms of weight k on $\Gamma_1(N)$.

Consider now an ordinary form f as above, whose weight k satisfies $2 \leq k \leq \ell-1$. Suppose in addition that the exact sequence of $\mathbf{F}[D_\ell]$-modules

$$0 \to V_0 \to V \to V/V_0 \to 0$$

is *split*. The inertia group I_ℓ acts on V_0 via the character χ^{k-1}, since $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts on $\det V$ via the product $\epsilon\chi^{k-1}$ and since I_ℓ acts trivially on V/V_0. Consider the representation $\rho' := \rho \otimes \chi^{1-k}$ of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Locally at ℓ, it is the direct sum of $V_0 \otimes \chi^{1-k}$ and $(V/V_0) \otimes \chi^{1-k}$. In particular, it is an extension of the unramified line $V_0 \otimes \chi^{1-k}$ by a line on which I_ℓ acts via the character $\chi^{1-k} = \chi^{\ell+1-k-1}$. This suggests that ρ' arises from an eigenform of weight $k' := \ell+1-k$ and character ϵ on $\Gamma_1(N)$; and indeed Serre's recipe for weights sets $k(\rho') = k'$. It is clear that ρ' is modular of some weight, because twisting representations by χ corresponds to applying the operator θ on modular forms. Hence (3.1) asserts, in particular, that ρ' arises from an eigenform f' of weight k'. The construction of the "companion" f' to f is the main result of [17]. (If $k = \ell$ and ρ is split locally at ℓ, then Gross still produces a companion form, of weight 1, in certain cases. For ℓ odd, the remaining cases of weight 1 are treated in a recent article by Coleman and Voloch [9].) This completes the discussion of Theorem 3.1.

As above, we shall use the phrase "ρ arises from $\Gamma_1(N)$" to indicate that ρ arises from an eigenform in $S_k(\Gamma_1(N))$ for some $k \geq 2$.

(3.2) Corollary. *Suppose that ρ arises from $\Gamma_1(N)$, with N prime to ℓ. Then there exists a power χ^i of the mod ℓ cyclotomic character χ such that $\rho \otimes \chi^i$ arises from $S_k(\Gamma_1(N))$ for some $k \leq \ell{+}1$.*

Proof. Replacement of ρ by any of its twists $\rho \otimes \chi^i$ $(i = 0, 1, 2, \dots)$ corresponds to applying the operator $\theta = q\frac{d}{dq}$ on modular forms i times. Hence, ρ arises from a given $\Gamma_1(N)$ if and only if $\rho_i := \rho \otimes \chi^i$ arises from the same group. The definition of $k(\rho)$ is such that a suitable twist ρ_i satisfies the inequality

$$2 \leq k(\rho_i) \leq \ell{+}1.$$

(See [15, Theorem 3.4], and the discussion following the statement of that theorem.) By Theorem 3.1, ρ_i arises from $S_{k(\rho_i)}(\Gamma_1(N))$. ■

Let $\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ be an irreducible representation. Consider the following four sets of positive integers:

$$\begin{aligned}
\mathcal{N}_1 &:= \{\, N \text{ prime to } \ell \mid \rho \text{ arises from } S_{k(\rho)}(\Gamma_1(N)) \,\};\\
\mathcal{N}_2 &:= \{\, N \text{ prime to } \ell \mid \rho \text{ arises from } \Gamma_1(N) \,\};\\
\mathcal{N}_3 &:= \{\, N \text{ prime to } \ell \mid \rho \text{ arises from } \Gamma_1(N\ell^\alpha) \text{ for some } \alpha \geq 0 \,\};\\
\mathcal{N}_4 &:= \{\, N \text{ prime to } \ell \mid \rho \text{ arises from } S_2(\Gamma_1(N\ell^2)) \,\}.
\end{aligned}$$

(3.3) Theorem. *If $\ell \geq 5$, then the four sets of integers $\mathcal{N}_i(\rho)$ are equal.*

Proof. The equality of $\mathcal{N}_1$ and $\mathcal{N}_2$ is guaranteed by Theorem 3.1. That $\mathcal{N}_2$ and $\mathcal{N}_3$ are equal follows from Theorem 2.1. Since it is clear that $\mathcal{N}_4 \subseteq \mathcal{N}_3$, it remains only to show that $\mathcal{N}_3 \subseteq \mathcal{N}_4$.

We will show, equivalently, that $\mathcal{N}_2 \subseteq \mathcal{N}_4$. Assume that ρ arises from $\Gamma_1(N)$, and choose $i \geq 0$ such that $\rho \otimes \chi^i$ arises from $S_k(\Gamma_1(N))$, with $2 \leq k \leq \ell{+}1$. By (2.2), $\rho \otimes \chi^i$ arises from $S_2(\Gamma_1(N\ell))$. To see that $N \in \mathcal{N}_4$, we view the twisting operator $\otimes\chi^{-i}$ as a a Dirichlet twist on modular forms. Such a twist changes the level of a modular form, but not its weight. From [1, Proposition 3.1], we find that $\rho = (\rho{\otimes}\chi^i) \otimes \chi^{-i}$ arises from $S_2(\Gamma_1(N\ell^2))$. ■

In case $\ell \geq 5$, we will write simply $\mathcal{N}(\rho)$ for the common value of the four sets $\mathcal{N}_i(\rho)$.

4. The levels of ρ and of f.

Suppose that an irreducible representation $\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ is "modular" in the sense that it arises from an eigenform of some weight and some level. Problem (1.4) requires that we find an eigenform of weight $k(\rho)$ and level $N(\rho)$ which gives rise to ρ. Because of Edixhoven's theorem (Theorem 3.1 above), the weight aspect of (1.4) disappears. In fact, suppose that $\ell \geq 5$. Then Theorem 3.3 translates (1.4) into the folllowing question:

$$N(\rho) \overset{?}{\in} \mathcal{N}(\rho). \tag{4.1}$$

In other words, our goal is to show that ρ is modular of level $N\ell^\alpha$, for some $\alpha \geq 0$. With (4.1) in mind, we are led to examine the set $\mathcal{N}(\rho)$ more closely. The following theorem was proved by Carayol [7] and, independently, by Livné [30, Proposition 0.1]:

(4.2) Theorem. *Suppose that ρ arises from $\Gamma_1(N)$. Then Serre's invariant $N(\rho)$ divides N.*

In attacking problem (4.1), one begins with an eigenform f giving rise to ρ. If N is the level of f, we have $N(\rho)|N$. Since $N(\rho)$ is, by definition, prime to ℓ, we have $N(\rho)|N'$, where N' is the prime-to-ℓ part of N. In case this divisibility is strict, the aim is to "lower" N by replacing it by a divisor. For the most part, the strategy is to do this "locally": we consider a prime $p \neq \ell$ which divides $N/N(\rho)$ and seek to replace N by N/p^i for some positive integer i.

As a preliminary step, we should certainly replace f by the *newform* (or primitive form) associated to f. This is an eigenform whose system of λ-adic representations coincides with the system (ρ_λ) attached to f, and whose level is minimal among all such eigenforms. The level of the newform attached to f is a divisor of N. (For the theory of newforms, see for instance the account in [29], or [34, §4.6].) According to a theorem of Deligne, Langlands [28], and Carayol [6], the conductor of the system (ρ_λ) coincides with N once this replacement is made.

Assume, then, that f is a newform. Theorem 4.2 is actually a consequence of a much more refined comparison of $N(\rho)$ with (the prime-to-ℓ part of) N. This comparison is again due to Carayol [7] and Livné [30]. We will discuss the analysis of Carayol and Livné, under the simplifying assumption $\ell \geq 5$. In view of (1.3), this assumption permits us to assume that the character ϵ which is associated to f has prime-to-ℓ order, and thereby facilitates the discussion.

Consider that representation ρ_λ which is attached to f and the prime λ induced by the chosen place v of $\overline{\mathbf{Q}}$. Then ρ is the reduction of ρ_λ in the sense which is outlined above. Since ρ is irreducible, ρ is in fact the "naive" reduction of any model of ρ_λ over $\mathcal{O}_\lambda$: no semisimplification is required.

Fix a prime $p \neq \ell$, and let $D_p \subset \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ again be a decomposition group for p. Let I_p be the inertia subgroup of D_p, and let $e = \mathrm{ord}_p(N/N(\rho))$. Thus, e is a non-negative integer, according to (4.2).

The inequality $e \geq 0$ can be interpreted conceptually. Consider ρ_λ and ρ as homomorphisms

$$\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}(V_\lambda), \qquad \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{Aut}\, V,$$

where V_λ and V are two dimensional vector spaces over E_λ and $\mathbf{F}$ respectively. Then one has

$$e = \dim_{\mathbf{F}} V^{I_p} - \dim_{E_\lambda} V_\lambda^{I_p}, \tag{4.3}$$

where V^{I_p} is the space of I_p-invariants in V and $V_\lambda^{I_p}$ is space of I_p-invariants in V_λ. In particular, we have $e > 0$ if and only if

$$\dim_{\mathbf{F}} V^{I_p} > \dim_{E_\lambda} V_\lambda^{I_p}.$$

Carayol and Livné have classified the situations in which we have $e > 0$. We will now describe their classification. For each situation which arises in the classification, we will summarize what is known about the question of showing that ρ is modular of level (dividing) N/p, i.e., that ρ arises from $\Gamma_1(N/p)$. Our summary appeals to a number of results which are proved only later in this article (for example, Theorems 5.1, 1.5, and 8.1). The material in this § may thus be viewed as a motivation for the theorems in §§5–8.

The first step in the classification is to distinguish cases according to the behavior of the component at p of the automorphic representation attached to f. Recall that the newform f is associated to an automorphic representation $\pi_{\mathbf{A}}$ of $\mathbf{GL}(2, \mathbf{A})$, where $\mathbf{A}$ is the adèle ring of $\mathbf{Q}$. Let π be the component of $\pi_{\mathbf{A}}$ at p, so that π is an admissible representation of $\mathbf{GL}(2, \mathbf{Q}_p)$. This representation may be classified as a principal series representation, a special representation, or a cuspidal representation of $\mathbf{GL}(2, \mathbf{Q}_p)$.

The Langlands correspondence attaches to π a λ-adic representation $\rho_{\lambda,\pi}$ of the Weil group of $\mathbf{Q}_p$. (Strictly speaking, $\rho_{\lambda,\pi}$ is defined only up to isomorphism. See [54], §4 for a discussion of the relation between λ-adic representations of the Weil group of $\mathbf{Q}_p$ and complex representations of the corresponding Weil-Deligne group.) By the main result of [6], $\rho_{\lambda,\pi}$ is somorphic to the restriction of ρ_λ to the Weil group of $\mathbf{Q}_p$. Therefore, properties of π are mirrored by the local behavior of ρ_λ at p. The restriction of ρ_λ to D_p is: reducible and semisimple when π is a principal series representation; reducible, but not semisimple when π is special; and irreducible when π is cuspidal.

We now discuss the inequality $e > 0$ in each of these three cases.

The principal series case.

In this case, which was treated by Carayol, the representation π arises from an (unordered) pair of Grossencharacters of $\mathbf{Q}_p^*$, say α and β. Identify these characters with characters of D_p, using the reciprocity map of local classfield theory. Then $\rho_\lambda|_{D_p}$ is the direct sum $\alpha \oplus \beta$.

(4.4) Proposition [7]. *Assume that $e > 0$. Then there is a Dirichlet character ϕ of conductor p and ℓ-power order such that the newform associated to $f \otimes \phi$ has level dividing N/p. In particular, ρ is modular of level N/p.*

Proof. One of α and β is ramified, but has unramified reduction mod λ. After possibly permuting α and β, we may assume that the restriction of α to I_p is non-trivial, but has ℓ-power order. On the other hand, because of formula (1.1), and the assumption that the order of ϵ is prime to ℓ, the restriction of $\det \rho_\lambda$ to I_p has prime-to-ℓ order. (Note that $\tilde{\chi}$ is ramified only at ℓ.) Thus $\alpha\beta|_{I_p}$ has prime-to-ℓ order. It follows from this that, on I_p, α is a power of β, since β is the product $\alpha^{-1} \cdot \alpha\beta$ of characters of relatively prime orders. Thus the kernel of $\alpha^{-1}|_{I_p}$ contains that of $\beta|_{I_p}$, and it follows that the conductor of $\alpha^{-1}\beta$ is at most that of β. (Incidentally, we will write the conductor multiplicatively: the conductor of a character of D_p is a power of p, and the conductor of the trivial character of D_p is $1 = p^0$.)

Since $\alpha|_{I_p}$ has prime-to-p order, α coincides on I_p with some power ω^i of the Teichmüller character ω which appeared in the proof of (2.1). Note that ω is a character which is defined naturally on $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. Let $\phi = \omega^{-i}$, and consider the twist $f \otimes \phi$, which corresponds to the λ-adic representation $\rho_\lambda \otimes \phi$. Since ϕ has ℓ-power order, the reductions of $\rho_\lambda \otimes \phi$ and ρ_λ coincide. On the other hand,

$(\rho_\lambda \otimes \phi)|_{D_p}$ is a direct sum of the *unramified* $\alpha\phi$ and the character $\beta\phi$. This latter character coincides with $\beta\alpha^{-1}$ on I_p. Therefore, its conductor is a divisor of the conductor of β. At the same time, the conductor of $\alpha\phi$, which is 1, is a strict divisor of the conductor of α. Therefore, the power of p dividing the conductor of $\rho_\lambda\phi$ is smaller than the power of p dividing the conductor of ρ_λ.

In the language of cusp forms, this means that the level of the newform associated to $f \otimes \phi$ is divisible by a smaller power of p than the level N of f. The two levels differ only at p, since the character ϕ is ramified only at p. ■

The special case.

Here, we have $\pi = \alpha \otimes \mathrm{sp}$, where sp is a "standard" special representation, and α is a Grossencharacter of $\mathbf{Q}_p^*$, which we identify as above with a character of D_p. We consider that sp has been normalized so that $\rho_\lambda|_{D_p}$ is an extension of the one-dimensional representation with character α by the one-dimensional representation with character $\alpha\tilde{\chi}$. (The fact that $\rho_\lambda|_{D_p}$ has this form, when π is special, was proved by Langlands [28].) The semisimplification of $\rho|_{D_p}$ then has the form $\bar{\alpha} \oplus \bar{\alpha}\chi$, where $\bar{\alpha}$ is the reduction of α mod λ. Also, $\epsilon = \alpha^2\tilde{\chi}^{2-k}$. Finally, the conductor of $\alpha \otimes \mathrm{sp}$ is the square of the conductor of α if α is ramified, and p^1 if not.

(4.5) Proposition [7]. *Assume that $e > 0$ and that α is ramified. Then there is a Dirichlet character ϕ of conductor p and ℓ-power order such that the newform associated to $f \otimes \phi$ has level dividing N/p. In particular, ρ is modular of level N/p.*

Proof. Since $e > 0$, we have $\dim_{\mathbf{F}} V^{I_p} > 0$, so that $\bar{\alpha}$ is unramified. Thus $\alpha|_{I_p}$ has ℓ-power order, and coincides with some power ω^i of ω. It follows that the conductor of α is p. As above, we set $\phi = \omega^{-i}$. The representation $\pi \otimes \phi$ then has conductor p, whereas π itself has conductor p^2. Hence replacing f by the newform associated to $f \otimes \phi$ replaces N by N/p. ■

In the situation of Proposition 4.5, $\bar{\epsilon}$ is unramified at p, since the determinant of ρ is unramified at p. Since we have assumed that the order of ϵ is prime to ℓ, this implies that ϵ is unramified at p. Hence f is in fact a newform on the group $\Gamma_1(M) \cap \Gamma_o(p^2)$, where $M = N/p^2$ is prime to p, and the character ϕ is quadratic. In the case where π is special but Proposition 4.5 does *not* apply, α is an unramified character, so that ϵ, in particular, is unramified. The newform f then has level Mp, with M prime to p, and f is a cusp form on the group $\Gamma_1(M) \cap \Gamma_o(p)$. We have $e > 0$ if and only if the representation ρ is unramified. When this is the case, $N(\rho)$ is prime to p, and one must show that ρ is represented by an eigenform of level M. Here is a statement of the problem to be solved (cf. [7, Conjecture B]):

(4.6) Problem. *Suppose that $\ell \geq 5$ and that p is a prime distinct from ℓ. Assume that ρ arises from an eigenform f of weight k on $\Gamma_1(M) \cap \Gamma_o(p)$, where M is prime to p and that ρ is unramified at p. Show that ρ arises from an eigenform on $\Gamma_1(M)$.*

The representation π is automatically special in all non-trivial cases of (4.6). Indeed, let f be an eigenform on $\Gamma_1(M) \cap \Gamma_o(p)$. If the newform associated to f has level dividing M, then in particular ρ arises from an eigenform on $\Gamma_1(M)$. If, to the contrary, the level of the newform associated to f is divisible by p, then π is special.

We now describe the extent to which (4.6) has been treated. A first remark, which is a consequence of the discussion concerning the four sets $\mathcal{N}_i(\rho)$, is that it suffices to solve (4.6) in *either* of the following situations:

a. $k = 2$ and M is of the form $N\ell^\alpha$ with N prime to ℓ;

b. M is prime to ℓ and $2 \leq k \leq \ell+1$.
One can assume $\alpha \leq 2$ in case (a), but this does not seem to be helpful.

(4.7) Mazur's Principle. *Suppose that $\ell \geq 5$ and that $p \neq \ell$ is a prime such that $p \not\equiv 1 \bmod \ell$. Assume that ρ arises from an eigenform f of weight k on $\Gamma_1(M) \cap \Gamma_o(p)$, with M prime to p, and that ρ is unramified at p. Then ρ arises from an eigenform on $\Gamma_1(M)$.*

This theorem was proved by Mazur in a letter to J-F. Mestre [31] in the special case where $k = 2$ and f has trivial character, i.e., is a cusp form on $\Gamma_o(Mp)$. Mazur's proof was presented by the author in [41, §6]. In proving the theorem more generally, we may apply the analysis above to work either in case (a) or in case (b). We choose to work in situation (a), i.e., in the case of forms of weight $k = 2$. This is essentially the case we have already treated in [41, §6]; the main difference is that we now allow the Nebentypus character of f to be non-trivial. (This character is naturally defined mod M, because f is a form on $\Gamma_1(M) \cap \Gamma_o(p)$.) For the convenience of the reader, we establish Mazur's principle, in case (a), in §8 below. (In §8, we require only that ℓ be odd: we do not exclude the case $\ell = 3$. Also, we use slightly different notation: the prime q of §8 plays the role of p in 4.7.)

After Mazur formulated his principle in 1985, problem 4.6 was studied extensively when $k = 2$ and f is a cusp form on $\Gamma_o(Mp)$. Here, one wishes to show that ρ arises from a weight-two eigenform on $\Gamma_o(M)$. In [41], the author proved this result when M is prime to ℓ. Next, in a joint article [32], Mazur and the author established a "multiplicity one" theorem which implies, by the techniques of [41], that Theorem 1.5 holds when ℓ, but not ℓ^2, divides M. Then, in a recent note [43], the author proved Theorem 1.5 when M is "exactly divisible" by a prime q, different from p and ℓ, at which ρ is ramified. The argument given in [43] relies heavily on a recent theorem of N. Boston, H.W. Lenstra, Jr., and the author [4], but avoids appeal to the "multiplicity one" principle which appears in [41]. Theorem 1.5, which will be proved in §7, may be viewed as the most recent link in this chain.

Theorem 1.5 falls short of solving (4.6) because of the requirement that the character of f be trivial. However, as already mentioned in §1, the author expects that the proof of (1.5) will extend without difficulty to the case where the character associated to f is an arbitrary character on $(\mathbf{Z}/M\mathbf{Z})^*$.

Problem 4.6 has been studied from perspective (b) by Jordan and Livné [24], who expect to treat this case (at least) in the case where f has trivial character and the weight k satisfies the inequality $2 \leq k < \ell$. Their arguments are an adaptation to weight k of arguments given in [41]. In particular, the arguments of Jordan and Livné rely on a weight-k "multiplicity one" principle, which will be treated in a forthcoming article of Faltings and Jordan [16]. It is quite possible that the arguments of §7 can be adapted to their situation, thereby obviating the necessity of using a weight-k multiplicity-one principle.

The cuspidal case.

In this case, the restriction of ρ_λ to D_p is irreducible. Assume that $e > 0$. Then, according to Carayol [7, §§1.1–1.2], $\rho_\lambda|_{D_p}$ is the two-dimensional representation induced from an abelian character $\xi\colon H \to \overline{\mathbf{Q}}^*$, where H is the subgroup of index two in D_p which corresponds to the unramified quadratic extension of $\mathbf{Q}_p$. Moreover, the reduction of ξ mod v is unramified. The level N is divisible exactly by p^2; i.e., it is divisible by p^2, but not by p^3. The Serre invariant $N(\rho)$ is divisible at most by p. One wishes to show that ρ arises from an eigenform on $\Gamma_1(N/p)$ of some weight, i.e., that ρ arises from $\Gamma_1(N/p)$.

For this, we can (and will) assume that the weight k of f satisfies $2 \leq k \leq \ell{+}1$.

By Theorem 5.1 below, there are infinitely many prime numbers q, prime to N and congruent to -1 mod ℓ, such that: ρ arises from a weight-k eigenform on the group $\Gamma_1(N) \cap \Gamma_o(q)$ for which the associated newform has level divisible by q. (One takes q to be an "auxiliary prime" as defined in §5.) Fix such a prime, together with an eigenform F satisfying the given condition. Let $\pi_{\mathbf{A}}$ be the representation of $\mathbf{GL}(2, \mathbf{A})$ associated with F. In the language of automorphic representations, the condition on F ensures that the component at q of $\pi_{\mathbf{A}}$ is a special representation.

Consider the twisted form of $\mathbf{GL}(2)$ which arises from the quaternion algebra over $\mathbf{Q}$ which is ramified exactly at p and at q. Since the components at both p and q of $\pi_{\mathbf{A}}$ are discrete series representations, we can conclude that $\pi_{\mathbf{A}}$ arises by the Jacquet-Langlands correspondence from an automorphic representation on this twisted form. Using this correspondence, and a lemma which generalizes (1.3), Carayol proves that ρ arises from an eigenform on $\Gamma_1(N/p) \cap \Gamma_o(q)$. Note that, since ℓ is odd, the congruence $q \equiv -1$ mod ℓ implies that one has $q \not\equiv +1$ mod ℓ. Using this fact, together with Mazur's Principle (4.7), we conclude, as desired, that ρ is modular of level N/p.

5. Diamond's theorem.

This § concerns the result of F. Diamond [13] which was alluded to above. We establish an analogue (Theorem 5.1) of a result proved by the author [38], [42] in the case of weight-two forms on $\Gamma_o(N)$. This analogue is a simple variant of Theorem 1 of [13]. More precisely, we shall focus attention on some intermediate results obtained by Diamond during the course of the proof of [13, Theorem 1] and then derive Theorem 5.1 from these.

Results like Theorem 5.1 have been useful both in "level lowering" (as in§4 above) and in other contexts. For example, [5] applies the theorem of [42] to the study of deformations of mod ℓ representations arising from weight-two modular forms.

Let ρ: Gal($\overline{\mathbf{Q}}/\mathbf{Q}$) $\to$ $\mathbf{GL}(2, \mathbf{F})$ be an irreducible mod ℓ representation arising from $S_k(\Gamma_1(N))$, with N prime to ℓ. Consider the following two conditions on a prime number q, assumed to be prime to $N\ell$:

I. The representation ρ arises from a weight-k eigenform on $\Gamma_1(N) \cap \Gamma_o(q)$ for which the associated newform has level divisible by q.

II. The characteristic polynomial of $\rho(\mathrm{Frob}_q)$ is of the form $(T - a)(T - qa)$, with $a \in \mathbf{F}^*$.

(5.1) Theorem (Diamond). *Assume that k satisfies $2 \leq k \leq \ell{+}1$. Then conditions* I *and* II *are equivalent.*

We first make some comments concerning the two conditions. Let $\mathcal{S}$ be the space $S_k(\Gamma_1(N) \cap \Gamma_o(q))$, and let $\mathcal{S}^{q-\mathrm{new}}$ be the "q-new" subspace of $\mathcal{S}$, defined, for example, as the kernel of the natural trace map from $\mathcal{S}$ to the direct sum of two copies of $S_k(\Gamma_1(N))$. Condition I means simply that ρ arises from an eigenform in the space $\mathcal{S}^{q-\mathrm{new}}$.

Next, we should point out that Condition II is satisfied in the case where q is an *auxiliary prime.* Let σ be the three-dimensional representation $\rho \times \chi$, where the mod ℓ cyclotomic character χ is regarded as a one-dimensional representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ over $\mathbf{F}$. An auxiliary prime is a prime number q, prime to ℓN, such that $\sigma(\mathrm{Frob}_q)$ is conjugate to the matrices $\sigma(c)$, with $c \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ a complex conjugation. If q is such a prime, the characteristic polynomial of $\rho(\mathrm{Frob}_q)$ coincides with that of the $\rho(c)$. Since ρ is an odd representation, $\rho(c)$ is conjugate to $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, and thus has characteristic polynomial $(T-1)(T+1)$. Because $\chi(\mathrm{Frob}_q) = \chi(c) = -1$, we have $q \equiv -1 \pmod{\ell}$. Thus the characteristic polynomial of $\rho(\mathrm{Frob}_q)$ may be written $(T-a)(T-qa)$, where a can be taken to be either $+1$ or -1, so that condition II of §5 is satisfied.

By the Čebotarev Density Theorem, the set of prime numbers for which c and Frob_q map to conjugate elements in the image of $\rho \times \chi$ has positive density. In particular, there are an infinite number of auxiliary primes.

Also, we should point out that the implication "Condition I ⇒ Condition II" is a direct consequence of Langlands's theorem ([28], see also [6]). Indeed, suppose that I is satisfied, and let f be a newform in $\mathcal{S}^{q-\mathrm{new}}$ which gives rise to ρ. Let π be the admissible representation of $\mathbf{GL}(2, \mathbf{Q}_q)$ which is associated to f. The fact that f is associated with $\Gamma_1(D) \cap \Gamma_o(q)$, for some $D|N$, implies that π is a special representation $\alpha \otimes \mathrm{sp}$, with α unramified. As discussed above, a theorem of Langlands states that $\rho_\lambda|_{D_q}$ is an extension of the one-dimensional representation with character α by the one-dimensional representation with character $\alpha\tilde{\chi}$. The semisimplification of $\rho|_{D_q}$ has the form $\bar{\alpha} \oplus \bar{\alpha}\chi$, where $\bar{\alpha}$ is the reduction of α mod λ. Applying $\bar{\alpha} \oplus \bar{\alpha}\chi$ to $\mathrm{Frob}_q \in D_q$, we obtain the matrix $\begin{pmatrix} a & 0 \\ 0 & qa \end{pmatrix}$, with $a = \alpha(\mathrm{Frob}_q)$. The characteristic polynomial of this matrix is $(T-a)(T-qa)$. Hence condition I implies condition II.

We now turn to the implication "Condition II ⇒ Condition I," which is the essential content of Theorem 5.1. We first compare eigenforms in $S_k(\Gamma_1(N))$ and in $\mathcal{S}$. Suppose that ρ arises from a weight-k eigenform $f = \sum c_n e^{2\pi i n\tau}$ on $\Gamma_1(N)$, and let ϵ be the character of $(\mathbf{Z}/N\mathbf{Z})^*$ associated with f. (For obvious reasons, we must refrain from writing $q = e^{2\pi i \tau}$.) The forms f and $f' := \sum c_n (e^{2\pi i n\tau})^q$ may be considered as elements of $\mathcal{S}$; indeed, this construction defines the standard inclusion of $S_k(\Gamma_1(N)) \oplus S_k(\Gamma_1(N))$ into $\mathcal{S}$ as a space of oldforms. Both forms f and f' in $\mathcal{S}$ are eigenvectors for the diamond operators $\langle d \rangle$ with $d \in (\mathbf{Z}/N\mathbf{Z})^*$ and the Hecke operators T_n with $(n, q) = 1$; the eigenvalues are the same as those for f when considered as an element of $S_k(\Gamma_1(N))$. (Some authors would write $\langle d \rangle_N$ for the operator $\langle d \rangle$, to emphasize that only the group $(\mathbf{Z}/N\mathbf{Z})^*$, rather than $(\mathbf{Z}/Nq\mathbf{Z})^*$, is operating. The group $(\mathbf{Z}/Nq\mathbf{Z})^*$ operates through its quotient $(\mathbf{Z}/N\mathbf{Z})^*$ because we are considering cusp forms on $\Gamma_o(q) \cap \Gamma_1(N)$.) However, neither of the forms f, f' is (in general) an eigenvector for the q^{th} Hecke operator $U := T_q$ on $\mathcal{S}$.

In order to obtain a true eigenform in $\mathcal{S}$, we consider the two roots α and β of the polynomial $T^2 - c_q T + q^{k-1}\epsilon(q)$. A computation shows that the two forms $f - \alpha f'$ and $f - \beta f'$ are eigenvectors for U on $\mathcal{S}$, with eigenvalues β and α, respectively.

Consider the Hecke algebra $\mathbf{T} = \mathbf{T}_{\mathcal{S}}$ associated with $\mathcal{S}$, i.e., the subring of $\mathrm{End}\,\mathcal{S}$ generated by the T_n with $n \geq 1$. We have $\langle d \rangle \in \mathbf{T}$ for all $d \in (\mathbf{Z}/N\mathbf{Z})^*$. Indeed, if r is a prime number which is prime to qN, then $T_r^2 - T_{r^2} = \langle r \rangle r^{k-1}$, so that $r^{k-1}\langle r \rangle \in \mathbf{T}$. By taking two distinct r's which map to d mod N, we find $\langle r \rangle \in \mathbf{T}$.

The action of $\mathbf{T}$ on $g := f - \beta f'$ is given by a homomorphism $\varphi\colon \mathbf{T} \to \mathbf{C}$, which is characterized by the formulas $\varphi(\langle d \rangle) = \epsilon(d)$, $\varphi(U) = \alpha$, and $\varphi(T_n) = c_n$, the latter valid for n prime to q. We may view φ as a homomorphism $\mathbf{T} \to \mathcal{O}$, where $\mathcal{O}$ is the integer ring of a suitable number field. Our fixed place $v|\ell$ of $\overline{\mathbf{Q}}$ induces a prime $\lambda|\ell$ of $\mathcal{O}$, together with an inclusion $\mathcal{O}/\lambda \hookrightarrow \mathbf{F}$. On composing φ with the map $\mathcal{O} \to \mathbf{F}$, we obtain a ring homomorphism

$$\bar{\varphi}\colon \mathbf{T} \to \mathbf{F}.$$

This map depends on f, and also on (α, β) as an ordered pair.

Assume now that condition II is satisfied and fix $a \in \mathbf{F}^*$ as in the condition. Then $T^2 - c_q T + q^{k-1}\epsilon(q) \equiv (T-a)(T-qa) \bmod \lambda$. After permuting the roots α and β if necessary, we have $\alpha \equiv a \bmod \lambda$. Setting

$$\eta = U^2 - q^{k-2}\langle q \rangle,$$

we find from condition II that $\bar{\varphi}(\eta) = 0$. Indeed, one has $\bar{\varphi}(q\eta) = qa^2 - \bar{\epsilon}(q)q^{k-1}$, but $\epsilon(q)q^{k-1}$ is congruent to qa^2 in view of condition II. In other words, the kernel $\mathfrak{m}$ of $\bar{\varphi}$ is a maximal ideal of $\mathbf{T}$ which contains η.

The subspace $\mathcal{S}^{q-\text{new}}$ of $\mathcal{S}$ is stable under all Hecke operators T_n and $\langle d \rangle$. Thus $\mathcal{S}^{q-\text{new}}$ cuts out a quotient $\mathbf{T}_{q-\text{new}}$ of $\mathbf{T}$: the image of $\mathbf{T}$ in $\mathrm{End}\,\mathcal{S}^{q-\text{new}}$. One verifies condition I by proving that $\mathfrak{m}$ arises by pullback from a maximal ideal $\overline{\mathfrak{m}}$ of $\mathbf{T}_{q-\text{new}}$, i.e., that $\mathfrak{m}$ lies in the support of the $\mathbf{T}$-module $\mathcal{S}^{q-\text{new}}$. It is well known that this property of $\mathfrak{m}$ implies that ρ arises from an eigenform in $\mathcal{S}$; see [17, Proposition 9.3] and [12, Lemma 6.11], and the proofs of these results. Therefore, to prove that condition II implies condition I, it suffices to verify

(5.2) Proposition. *Suppose that ℓ is prime to N and that k satisfies $2 \le k \le \ell+1$. Let $\mathfrak{m}$ be a maximal ideal of $\mathbf{T}$ associated with $\mathcal{S} = S_k(\Gamma_1(N) \cap \Gamma_o(q))$, where q is a prime number not dividing $N\ell$. Assume that $\mathfrak{m}$ divides ℓ (i.e., $\mathfrak{m}$ contains ℓ) and that $\mathfrak{m}$ contains $\eta = U^2 - q^{k-2}\langle q \rangle$. Then $\mathfrak{m}$ arises by pullback from the q-new quotient of $\mathbf{T}$.*

This proposition is implicit in Diamond's work, although it is not explicitly stated in [13].

To prove the proposition, we recall some results from [13], using somewhat different notation. Let L be the parabolic cohomology group with $\mathbf{Z}_\ell$-coefficients which is associated with $\mathcal{S}$, so that L is a free $\mathbf{Z}_\ell$-module of rank $2\dim_{\mathbf{C}} \mathcal{S}$. To construct L, we let Γ be the image of $\Gamma_1(N) \cap \Gamma_o(q)$ in $\mathbf{PSL}(2, \mathbf{Z})$, and consider the parabolic cohomology group $H^1_{\text{par}}(\Gamma, \mathrm{Sym}^{k-2}(\mathbf{Z}_\ell \oplus \mathbf{Z}_\ell))$. This group is a finitely-generated $\mathbf{Z}_\ell$-module M which is not necessarily torsion free; L is defined as the image of M in $M \otimes \mathbf{Q}_\ell$, i.e., the largest torsion free quotient of M. We let $X = X_1 \oplus X_2$ be the direct sum of two copies of the corresponding parabolic cohomology group made with $\Gamma_1(N)$ in place of $\Gamma_1(N) \cap \Gamma_o(q)$. Diamond considers the degeneracy map $X \hookrightarrow L$ which corresponds to the inclusion of $S_k(\Gamma_1(N)) \oplus S_k(\Gamma_1(N))$ in $\mathcal{S}$ on the level of modular forms. We will denote this map by α. (There is no further need to refer to the roots α and β which occurred in the discussion above.) The map α becomes $\mathbf{T}$-equivariant when one endows X with its natural action of $\mathbf{T}$. (In this action, the T_n with n prime to q act "diagonally" on X, while U acts as the two-by-two matrix $\begin{pmatrix} T_q & q^{k-1} \\ -\langle q \rangle & 0 \end{pmatrix}$.) In particular, the image of α is stable under $\mathbf{T}$.

Using results of Ihara, and the hypotheses on ℓ, Diamond proves that the cokernel Y of α is *torsion free*. The tautological exact sequence

$$0 \to X \to L \to Y \to 0 \tag{5.3}$$

is thus an exact sequence of free $\mathbf{Z}_\ell$-modules. The quotient of $\mathbf{T} \otimes \mathbf{Z}_\ell$ cut out by Y is the ℓ-adic completion of the q-new quotient of $\mathbf{T}$, while the quotient of $\mathbf{T} \otimes \mathbf{Z}_\ell$ cut out by X is the completion of the analogous q-old quotient of $\mathbf{T}$; this is the same quotient which is defined by the action of $\mathbf{T}$ on $S_k(\Gamma_1(N)) \oplus S_k(\Gamma_1(N))$.

In particular, let $\mathfrak{m}$ be a maximal ideal of $\mathbf{T}$ with residue characteristic ℓ. Because $\mathbf{T}$ acts faithfully on L, $\mathfrak{m}$ lies in the support of L. In view of (5.3), it follows that $\mathfrak{m}$ lies either in $\operatorname{Supp} X$ or in $\operatorname{Supp} Y$, and perhaps in both. The statement to be proved may be rephrased as follows: if $\mathfrak{m} \in \operatorname{Supp} X$ and if $\eta \in \mathfrak{m}$, then $\mathfrak{m} \in \operatorname{Supp} Y$. Note that the two hypotheses on $\mathfrak{m}$ may be combined into the statement that $\mathfrak{m}$ belongs to the support of $X/\eta X$. Indeed, if $\mathfrak{m} \in \operatorname{Supp} X$, then $X/\mathfrak{m}X$ is non-zero, by Nakayama's lemma. If furthermore we have $\eta \in \mathfrak{m}$, then $X/\mathfrak{m}X$ is a quotient of $X/\eta X$. Hence $\mathfrak{m}$ belongs to the support of $X/\eta X$, since $\mathfrak{m}$ belongs to the support of $X/\mathfrak{m}X$. Conversely, if $\mathfrak{m}$ lies in the support of $X/\eta X$, it is clear that η is an element of $\mathfrak{m}$ and that $\mathfrak{m} \in \operatorname{Supp} X$. Thus one must prove

(5.4). *If* $\mathfrak{m} \in \operatorname{Supp}(X/\eta X)$*, then* $\mathfrak{m} \in \operatorname{Supp} Y$.

Diamond proves (5.4) by constructing a $\mathbf{T}$-equivariant surjection $Y \to X/\eta X$. To construct this map, we first dualize (5.3), thereby obtaining an exact sequence

$$0 \to \operatorname{Hom}(Y, \mathbf{Z}_\ell) \to \operatorname{Hom}(L, \mathbf{Z}_\ell) \to \operatorname{Hom}(X, \mathbf{Z}_\ell) \to 0.$$

As Diamond recalls, perfect pairings constructed by Hida [19, Theorem 3.2] identify $\operatorname{Hom}(L, \mathbf{Z}_\ell)$ with L and $\operatorname{Hom}(X, \mathbf{Z}_\ell)$ with X. Using these pairings, we obtain an exact sequence

$$0 \to \operatorname{Hom}(Y, \mathbf{Z}_\ell) \to L \xrightarrow{\beta} X \to 0. \tag{5.5}$$

The map β is "almost" $\mathbf{T}$-equivariant: it intertwines the maps labeled T_n on L and on X whenever $(n, q) = 1$, but intertwines the q^{th} Hecke operator U of L with the endomorphism $\begin{pmatrix} 0 & q \\ -\langle q\rangle q^{k-2} & T_q \end{pmatrix}$ of X (cf. [19, Prop. 3.3]).

Consider the automorphism $\omega = \begin{pmatrix} -q^{k-2}\langle q\rangle & T_q \\ 0 & -\langle q\rangle \end{pmatrix}$ of X. A computation shows that the composite $\omega\circ\beta\colon L \to X$ is $\mathbf{T}$-equivariant. Moreover, $(\omega\circ\beta)\circ\alpha$ is the endomorphism η of X. It follows that $\omega\circ\beta$ induces a surjection $L/\alpha X \to X/\eta X$. Since Y is, by definition, $L/\alpha X$, we obtain the desired map.

6. Character groups and component groups.

In this §, we recall some material from [41] concerning the component groups and character groups associated with bad reductions of ordinary modular curves and Shimura curves. In addition, we explore in further detail a relation that was first found by Jordan and Livné [23] and then deepened by [41, Theorem 4.3].

For each $N \geq 1$, we let $X_o(N)$ be the classical modular curve (over $\mathbf{Q}$) which classifies elliptic curves which are endowed with cyclic subgroups of order N. The space of holomorphic differentials on $X_o(N)_{/\mathbf{C}}$ may be identified with $S_2(\Gamma_o(N))$

in a canonical fashion. The curve $X_o(N)$ comes equipped with a family of correspondences T_n $(n \geq 1)$; the correspondence T_n induces on $S_2(\Gamma_o(N))$ the endomorphism T_n. The Jacobian $\mathrm{Pic}^{\mathrm{o}}\, X_o(N)$ will be denoted $J_o(N)$, as usual; we write again simply T_n for the endomorphism of $J_o(N)$ which is induced by the correspondence T_n of $X_o(N)$.

Consider first a prime number q and a positive integer N prime to q. We are interested in the modular curves $X_o(qN)$ and $X_o(N)$, and in their Jacobians $J_o(qN)$ and $J_o(N)$. The two standard degeneracy coverings $X_o(qN) \rightrightarrows X_o(N)$ induce, by pullback, a map $\delta\colon J_o(N) \times J_o(N) \to J_o(qN)$. The kernel of δ is finite; it is the "antidiagonal" image in $J_o(N) \times J_o(N)$ of the kernel of the pullback of the natural covering $X_1(N) \to X_o(N)$ [38]. (The curve $X_1(N)$ classifies elliptic curves which are furnished with a point of order N.)

For each $n \geq 1$, a Hecke operator labeled T_n acts diagonally on the product $J_o(N) \times J_o(N)$, and a similarly-named operator acts on the target $J_o(qN)$ of δ. When n is prime to q, δ is compatible with the two operators T_n. However, when $n = q$, the situation is more subtle. Reserve the symbol T_q for the q^{th} Hecke operator on $J_o(N)$ and let U denote the q^{th} Hecke operator on $J_o(qN)$. Then $U \circ \delta = \delta \circ U$, where the latter operator U denotes the matrix $\begin{pmatrix} T_q & q \\ -1 & 0 \end{pmatrix}$ of endomorphisms of $J_o(N)$. (We view this matrix as an endomorphism of $J_o(N) \times J_o(N)$.) Let $\mathbf{T} = \mathbf{T}_{qN}$ be the subring of $\mathrm{End}\, J_o(qN)$ generated by the T_n for $n \geq 1$, i.e., by the T_n for n prime to q and the operator $U = T_q$. We endow $J_o(N) \times J_o(N)$ with the operation of $\mathbf{T}$ such that the elements U and T_n of $\mathbf{T}$ (with n prime to q) act as the operators of $J_o(N) \times J_o(N)$ with the same names. This is the unique operation for which δ is $\mathbf{T}$-equivariant.

We note in passing that the endomorphisms T_n and U and the homomorphism δ are defined over $\mathbf{Q}$. Also, in analogy with the situation of §5, we let $\eta = U^2 - 1$ in $\mathbf{T}$.

(6.1) Theorem. *There is a unique homomorphism of abelian varieties*

$$\sigma\colon J_o(qN) \to J_o(N) \times J_o(N)$$

such that $\sigma \circ \delta = \eta$. This homomorphism is $\mathbf{T}$-equivariant.

To construct σ, we consider the map $\delta'\colon J_o(qN) \to J_o(N) \times J_o(N)$ which is induced by the degeneracy coverings $X_o(qN) \rightrightarrows X_o(N)$ using Albanese (i.e., covariant) functoriality of the Jacobian. The definition of T_q as a correspondence on $X_o(N)$ shows that $\delta' \circ \delta = \begin{pmatrix} q+1 & T_q \\ T_q & q+1 \end{pmatrix}$. Define $\sigma = \begin{pmatrix} -1 & T_q \\ 0 & -1 \end{pmatrix} \circ \delta'$. A computation shows that $\sigma \circ \delta$ coincides with η as a matrix of endomorphisms of $J_o(N)$.

For the unicity and the $\mathbf{T}$-equivariance, we observe that

$$\mathrm{Hom}(Q, J_o(N) \times J_o(N)) = 0,$$

where $Q = J_o(qN)/\delta(J_o(N) \times J_o(N))$. This fact may be seen arithmetically, by noting that Q has purely toric reduction at the prime q [11], whereas $J_o(N)$ has good reduction at q. It proves the unicity, since the difference between two operators σ which satisfy $\sigma \circ \delta = \eta$ is an operator which vanishes on the image of δ. Similarly, let $T \in \mathbf{T}$, and consider $\sigma \circ T - T \circ \sigma$, which is a priori a homomorphism $h\colon J_o(qN) \to$

$J_o(N)\times J_o(N)$. We have $h\circ\delta = \eta T - T\eta = 0$, so that h factors through Q; as a consequence, $h = 0$. ■

Now specialize to the case where $N = pM$, where p is a prime number which does not divide M. The level qN becomes the product pqM in which the two prime numbers p and q play symmetrical roles. The space $\mathcal{S}$ of weight-two cusp forms on $\Gamma_o(pqM)$ is then analogous to the space denoted $\mathcal{S}$ in §5, in the special case $N = pM$, $k = 2$. The main difference is that we have replaced $\Gamma_1(N)\cap\Gamma_o(q)$ by $\Gamma_o(qN)$.

Let $J_o(pqM)_{/\mathbf{F}_p}$ be the fiber at p of the Néron model of $J_o(pqM)$. This group variety is an extension of a finite "component group" Θ_p by a connected group $J_o(pqM)^o_{/\mathbf{F}_p}$, which is in turn an extension of the product of two copies of $J_o(qM)$ by a torus $\mathcal{T}$ over $\mathbf{F}_p$. Let L_p be the character group of this torus, and let L_q be the character group of the analogous torus for $J_o(pqM)_{/\mathbf{F}_q}$. Similarly, let X_p be the analogue of L_p for $(J_o(pM)\times J_o(pM))_{/\mathbf{F}_p}$ and let X_q be the analogue of X_p with p replaced by q. The groups X_p and X_q are naturally direct sums of two copies of the character groups coming from $J_o(pM)_{/\mathbf{F}_p}$ and $J_o(qM)_{/\mathbf{F}_q}$, respectively; thus, we may represent endomorphisms of X_p, say, as two-by-two matrices of endomorphisms of the character group associated to $J_o(pM)_{/\mathbf{F}_p}$. The endomorphisms T_n (and U) of $J_o(pqM)$, $J_o(pM)\times J_o(pM)$ and $J_o(qM)\times J_o(qM)$ induce maps on the four character groups L_p, X_p, L_q, and L_p. In what the author hopes is an acceptable abuse of notation, these maps will be denoted simply T_n and U. In particular, the endomorphism U of X_p is the matrix of endomorphisms $\begin{pmatrix} T_q & -1 \\ q & 0 \end{pmatrix}$.

Continuing the abuse of notation, we will write simply δ for the map $L_p \to X_p$ induced by the degeneracy map $\delta\colon J_o(Mp)\times J_o(Mp) \to J_o(Mpq)$. The surjectivity of $\delta\colon L_p \to X_p$ was established in [41, Th. 3.15]. (In [41], the author wrote "X" for the character group associated to a *single* copy of $J_o(pM)$ or $J_o(qM)$.) Theorem 4.1 of [41] identifies the kernel of δ with an analogue of L_q in which $X_o(pqM)$ is replaced by an appropriate Shimura curve. Namely, let C be the Shimura curve made with a quaternion algebra of discriminant pq and $\Gamma_o(M)$-type level structure (cf. [41, §4]). Let J be the Jacobian of C, and let Y_p and Y_q be the analogues of L_p and L_q for J. The groups $J_{/\mathbf{F}_p}$ and $J_{/\mathbf{F}_q}$ are extensions of their component groups Ψ_p and Ψ_q by the tori whose character groups are Y_p and Y_q. (As was proved by Cerednik [8] and Drinfeld [14], all components of $C_{/\mathbf{F}_p}$ and $C_{/\mathbf{F}_q}$ have genus zero.) In other words, the reductions of J at p and q are "semiabelian," with trivial abelian variety parts. Then theorem 4.1 of [41] provides an exact sequence

$$0 \to Y_q \xrightarrow{\iota} L_p \xrightarrow{\delta} X_p \to 0 \tag{6.2}$$

and an analogue

$$0 \to Y_p \to L_q \to X_q \to 0 \tag{6.3}$$

in which the roles of p and q have been permuted. These sequences are compatible with the Hecke operators labeled T_n on J and on $J_o(pqM)$; note, incidentally, that the operators T_p and T_q on J are in fact *involutions.* (See [41, Theorem 4.1] and also the results proved in [40].) The sequences (6.2) and (6.3) seem to play a role analogous to that played by (5.3) and its $\mathbf{Z}_\ell$-dual in the discussion of §5.

One deduces from (6.2) or (6.3) that the Hecke ring $\mathbf{Z}[\ldots, T_n, \ldots] \subseteq \operatorname{End} J$ associated to J is naturally a quotient of the ring $\mathbf{T}$ associated to $J_o(pqM)$. Thus, we may speak of the action of $\mathbf{T}$ on J.

Let us introduce the notation $'$ to denote a dual abelian variety; thus $J_o(pqM)'$, for example, will be the abelian variety dual to $J_o(pqM)$. We use T_n' and U' to denote the endomorphisms of the dual objects which are induced functorially by the operators T_n and U. Thus, for instance, there is a natural action of the ring $\mathbf{T}$ on J', in which the element of $\mathbf{T}$ labeled T_n acts on J' as the endomorphism T_n'. We let Y_q', L_p', and so on, represent the character groups of the tori arising from the dual abelian varieties. The operator T_n (say) in $\mathbf{T}$ then induces an endomorphism of L_q' which will be denoted T_n'; this endomorphism is induced by the endomorphism T_n' of $J_o(pqM)'$. Similar remarks apply to the component groups which we have introduced: Θ_p and Θ_q in the case of $J_o(pqM)$, and Ψ_p and Ψ_q in the case of J.

Next, we discuss monodromy pairings. To fix ideas we will first consider L_p. The pairing associated with L_p is a bilinear map $\langle\ ,\ \rangle : L_p \times L_p' \to \mathbf{Z}$, which induces an injection $L_p' \hookrightarrow \mathrm{Hom}(L_p, \mathbf{Z})$ with finite cokernel. According to [18, §11], the cokernel of this injection is canonically isomorphic to Θ_p. The exact sequence

$$0 \to L_p' \to \mathrm{Hom}(L_p, \mathbf{Z}) \to \Theta_p \to 0$$

is $\mathbf{T}$-equivariant when the elements of $\mathbf{T}$ act in the natural way, i.e., when $T_n \in \mathbf{T}$ acts as T_n on Θ_p, as T_n' on L_p', and as $\mathrm{Hom}(T_n, \mathbf{Z})$ on $\mathrm{Hom}(L_p, \mathbf{Z})$. In a similar way, we have monodromy pairings

$$Y_q \times Y_q' \to \mathbf{Z}, \qquad X_p \times X_p' \to \mathbf{Z}.$$

One potential source of confusion is that the abelian varieties under discussion are Jacobians of modular curves, or else products of two Jacobians. It follows that they are naturally self-dual. After using the autoduality of the Jacobian to identify J' and J, for example, we find that the lattices Y_q and Y_q' are *equal*, or at least canonically isomorphic. In point of fact, it is often fruitful to identify both Y_q and Y_q' with $H_1(\Gamma, \mathbf{Z})$, where Γ is the "dual graph" attached to the mod q reduction of C. The group $H_1(\Gamma, \mathbf{Z})$ thereby acquires two actions of $\mathrm{End}\, J$; these differ by the Rosati involution of $\mathrm{End}\, J$ which arises from the theta divisor on C. Similarly, the group X_p has two natural actions of $\mathbf{T}$, which differ by the Rosati involution $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a' & c' \\ b' & d' \end{pmatrix}$ on $\mathrm{End}\, J_o(pM) \times J_o(pM)$, where $'$ is the Rosati involution on $\mathrm{End}\, J_o(pM)$. For example, the operator U, which acts on X_p as $\begin{pmatrix} T_q & -1 \\ q & 0 \end{pmatrix}$ in its standard action, becomes $\begin{pmatrix} T_q' & q \\ -1 & 0 \end{pmatrix}$ in the other action. (It might be worth pointing out that the endomorphisms T_q and T_q' of $J_o(N)$ are equal.) As a mnemonic device, we will attempt whenever possible to use the $'$ notation when the Hecke operators in $\mathbf{T}$ act as T_n' and U'.

These examples show that the natural autoduality $J_o(N) \approx J_o(N)'$ does not necessarily intertwine the actions of $\mathbf{T}$ on these two abelian varieties. Nonetheless, it is a fact that are natural $\mathbf{T}$-equivariant isomorphisms $J \approx J'$ and $J_o(N) \approx J_o(N)'$ for each N. In the case of $J_o(N)$, the isomorphism is provided by the involution $w = w_N$ on $J_o(N)$ which is induced by the Atkin-Lehner involution $(E, C) \mapsto (E/C, E[N]/C)$ on the set of elliptic curves with cyclic subgroups of order N (see [52, Proposition 3.54]). In the case of J, there is an analogous involution w on the set of "fake elliptic curves" which are classified by C. These $\mathbf{T}$-equivariant isomorphisms are occasionally useful.

Once we identify L_p with L'_p, the monodromy pairings associated to L_p and Y_q become maps $L_p \times L_p \to \mathbf{Z}$ and $Y_q \times Y_q \to \mathbf{Z}$, respectively. Both pairings are symmetric; they are given by explicit formulas involving the dual graphs of the mod p reduction of $X_o(pqM)$ and the mod q reduction of C. According to [41, Th. 4.1], the restriction to $Y_q \times Y_q$ of the monodromy pairing $L_p \times L_p \to \mathbf{Z}$ agrees with the monodromy pairing on Y_q. Here, one uses ι to embed Y_q in L_p.

(6.4) Proposition. *The embedding $\iota\colon Y_q \hookrightarrow L_p$ remains $\mathbf{T}$-equivariant when it is regarded as a map $Y'_q \to L'_p$.*

We regard L'_p and Y'_q as the same physical groups as L_p and Y_q, only with different actions of $\mathbf{T}$. Fix $T \in \mathbf{T}$, and use the symbols T_Y and T_L to denote the endomorphisms of Y_q and of L_p induced by T. Write T'_Y and T'_L for the endomorphisms of these groups which T induces in its “dual” actions. The endomorphism T'_L (say) is the adjoint of T_L with respect to the monodromy pairing $\langle\,,\,\rangle_L$ on L_p. Proposition (6.4) states the formula $T'_L \circ \iota = \iota \circ T'_Y$.

Consider the homomorphism $\kappa\colon J_o(Mp)' \times J_o(Mp)' \to J_o(Mpq)'$ which is dual to the map σ of (6.1). It induces a $\mathbf{T}$-equivariant map $\kappa\colon L'_p \to X'_p$; this map is adjoint to the map $X_p \to L_p$ induced by σ. The construction of σ shows that κ differs from δ by an automorphism of X_p. Therefore, the kernel of κ coincides with the kernel of δ, which is the image of ι. It follows that T'_L preserves this image, since we may write $\kappa \circ T'_L \circ \iota = T'_X \circ \kappa \circ \iota = 0$. Hence the map $\iota \circ T'_Y - T'_L \circ \iota$ maps Y_q to $\iota(Y_q)$. Consequently, to show that $\iota \circ T'_Y = T'_L \circ \iota$, it suffices to verify the equality of $\langle T'_Y y, z\rangle_Y$ and $\langle T'_L \circ \iota y, \iota z\rangle_L$ for all $y, z \in Y_q$. Here, we make use of the fact that the monodromy pairing $\langle\,,\,\rangle_Y$ on Y_q is the restriction to Y_q of the pairing $\langle\,,\,\rangle_L$ on L_p.

To verify the desired equation, we note the series of equalities

$$\langle T'_L \iota y, \iota z\rangle_L = \langle \iota y, T_L \iota z\rangle_L = \langle \iota y, \iota T_Y z\rangle_L = \langle y, T_Y z\rangle_Y = \langle T'_Y y, z\rangle_Y.$$

These follow from the adjuction relations and the equivariance of ι with respect to the standard actions of $\mathbf{T}$, plus the compatibility between the two monodromy pairings. ■

(6.5) Corollary. *We have an exact sequence of $\mathbf{T}$-modules*

$$0 \to Y'_q \xrightarrow{\iota} L'_p \xrightarrow{\kappa} X'_p \to 0. \ ■$$

Theorem 4.3 of [41] asserts that there is an exact sequence of $\mathbf{T}$-modules

$$0 \to \mathcal{K} \to X_p/\eta X_p \to \Psi_q \to \mathcal{C} \to 0.$$

Here, $\mathcal{K}$ and $\mathcal{C}$ are each closely related to Θ_p and to the component group of the group variety $(J_o(pM) \times J_o(pM))_{/\mathbf{F}_p}$; in particular, they are Eisenstein $\mathbf{T}$-modules in the sense that they are annihilated by $T_r - (r+1)$ for almost all prime numbers r, cf. [41, Th. 3.12], or [39]. (This allows us to neglect $\mathcal{C}$ and $\mathcal{K}$ in practice.) The author now believes that he was not sufficiently attentive to the distinction between the two actions of $\mathbf{T}$ and that the exact sequence of [41, Th. 4.3] should read

$$0 \to \mathcal{K}' \to X_p/\eta X_p \to \Psi'_q \to \mathcal{C}' \to 0, \tag{6.6}$$

where $\mathcal{C}'$ and $\mathcal{K}'$ are variants of $\mathcal{K}$ and $\mathcal{C}$, which are again Eisenstein. Note, however, that Ψ_q and Ψ'_q are *isomorphic* as $\mathbf{T}$-modules, in view of the Atkin-Lehner automorphism $w\colon J \approx J'$ which intertwines the two actions of $\mathbf{T}$. Hence [41, Th. 4.3] can be used as stated.

To derive a version of (6.6), we consider the monodromy pairings on L_p and X_p as injections

$$X_p \hookrightarrow \operatorname{Hom}(X'_p, \mathbf{Z}) \quad \text{and} \quad L_p \hookrightarrow \operatorname{Hom}(L'_p, \mathbf{Z}),$$

respectively. Their cokernels are the component groups Φ'_p and Θ'_p associated with the mod p reductions of $J_o(pM)\times J_o(pM)$ and $J_o(pqM)$. Using ι, we regard Y_q as a submodule of L_p. We consider the $\mathbf{T}$-equivariant maps $\sigma\colon X_p \to L_p$ and $\kappa\colon L'_p \to X'_p$ induced by the homomorphism $\sigma\colon J_o(pqM) \to J_o(pM)\times J_o(pM)$, and use the abbreviation κ^* to refer to the map $\operatorname{Hom}(\kappa, \mathbf{Z})$ which is the $\mathbf{Z}$-linear dual of κ. We have a commutative square

$$\begin{tikzcd}
L_p \arrow[r] & \operatorname{Hom}(L'_p, \mathbf{Z}) \\
X_p \arrow[u, "\sigma"] \arrow[r] & \operatorname{Hom}(X'_p, \mathbf{Z}) \arrow[u, "\kappa^*"]
\end{tikzcd}$$

in which the horizontal maps are the injections coming from the two monodromy pairings. On dividing L_p by its submodule Y_q, we obtain a commutative diagram

$$\begin{tikzcd}
0 \arrow[r] & L_p/Y_q \arrow[r] & \operatorname{Hom}(L'_p, \mathbf{Z})/Y_q \arrow[r] & \Theta'_p \arrow[r] & 0 \\
0 \arrow[r] & X_p \arrow[u, "\sigma"] \arrow[r] & \operatorname{Hom}(X'_p, \mathbf{Z}) \arrow[u, "\kappa^*"] \arrow[r] & \Phi'_p \arrow[u] \arrow[r] & 0
\end{tikzcd}$$

in which the rows are exact. The map $\delta\colon L_p \to X_p$ identifies L_p/Y_q with X_p. Also, we have on X_p the identity $\delta\circ\sigma = \eta$, in view of the equation $\sigma\circ\delta = \eta$ involving homomorphisms of abelian varieties. Hence the diagram may be rewritten

$$\begin{tikzcd}
0 \arrow[r] & X_p \arrow[r] & \operatorname{Hom}(L'_p, \mathbf{Z})/Y_q \arrow[r] & \Theta'_p \arrow[r] & 0 \\
0 \arrow[r] & X_p \arrow[u, "\eta"] \arrow[r] & \operatorname{Hom}(X'_p, \mathbf{Z}) \arrow[u, "\kappa^*"] \arrow[r] & \Phi'_p \arrow[u] \arrow[r] & 0.
\end{tikzcd}$$

In view of (6.5), the cokernel of κ^* is $\operatorname{Hom}(Y'_q, \mathbf{Z})$. Therefore, the cokernel of the middle vertical map is the finite group $\operatorname{Hom}(Y'_q, \mathbf{Z})/Y_q = \Psi'_q$. By a dimension count, this shows, in particular, that the middle vertical map is injective. From the snake lemma, we deduce a version of (6.6) in which the groups $\mathcal{K}'$ and $\mathcal{C}'$ are the kernel and cokernel of the right-hand vertical map.

(6.7) Proposition. *The* $\mathbf{T}$*-modules* Ψ_q *and* Ψ'_q *are mutually* $\mathbf{Q}/\mathbf{Z}$*-dual.*

Proof. The indicated duality between the two component groups may be read directly from the recipe for making Ψ_q and Ψ'_q from the monodromy pairing on Y_q. For more details, see [18,§11.4], which treats the more general case of an abelian variety with semistable reduction, and also [18,§11.3], which discusses the prime-to-p parts of component groups associated to abelian varieties over a local field of residue characteristic p which do not necessarily have semistable reduction. ■

As O. Gabber and R. Livné explained to the author, Grothendieck conjectured in [18, 1.2.1] that there is a perfect $\mathbf{Q}/\mathbf{Z}$-duality between the component groups associated to two dual abelian varieties over a local field. In [18], Grothendieck did not treat the p-primary parts of component groups associated to arbitrary abelian varieties over local fields of residue characteristic p. This gap was filled, in the case of a perfect residue field, by L. Bégueri as Th. 8.3.3 of [3]. Another proof, for an abelian variety over the fraction field of a discrete valuation ring with finite residue field, is given in [33].

In the statement of the following corollary, we introduce Mazur's notation "$[\mathfrak{m}]$" for a kernel: if $\mathfrak{m}$ is an ideal of $\mathbf{T}$, and M is a $\mathbf{T}$-module, then $M[\mathfrak{m}]$ denotes the submodule of M consisting of those element of M which are annihilated by all elements of $\mathfrak{m}$.

(6.8) Corollary. *Let $\mathfrak{m} \subset \mathbf{T}$ be a maximal ideal. Then $\mathfrak{m}$ lies in the support of Ψ_q if and only if $\mathfrak{m}$ lies in the support of Ψ'_q. Moreover, the $\mathbf{T}/\mathfrak{m}$-dimensions of $\Psi_q/\mathfrak{m}\Psi_q$ and $\Psi'[\mathfrak{m}]$ are equal; similarly, the dimensions of $\Psi'_q/\mathfrak{m}\Psi'_q$ and $\Psi_q[\mathfrak{m}]$ coincide.*

Proof. The two equalities of dimension result directly from (6.7). Suppose now that $\mathfrak{m}$ lies in the support of Ψ_q. By Nakayama's Lemma, $\Psi_q/\mathfrak{m}\Psi_q$ is non-zero. According to (6.7), however, $\Psi_q/\mathfrak{m}\Psi_q$ and $\Psi'_q[\mathfrak{m}]$ are duals of each other. Thus $\Psi'_q[\mathfrak{m}]$ is non-zero. This non-vanishing implies immediately that $\mathfrak{m}$ lies in the support of Ψ'_q. Similarly, the support of Ψ'_q is contained in the support of Ψ_q. ■

(6.9) Theorem. *Let $\mathfrak{m}$ be a maximal ideal of $\mathbf{T}$ for which the associated representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ is irreducible. Then $\mathfrak{m}$ lies in the support of Ψ_q if and only if $\mathfrak{m}$ contains η and lies in the support of X_p. Moreover, in this case we have the equalities of $\mathbf{T}/\mathfrak{m}$-dimensions*

$$\dim X_p/\mathfrak{m}X_p = \dim \Psi'_q/m\Psi'_q = \dim \Psi_q[\mathfrak{m}].$$

Proof. Assume that the representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ associated with $\mathfrak{m}$ is irreducible. By [41, 5.2c], this hypothesis implies that $\mathfrak{m}$ does not lie in the support of either of the modules $\mathcal{K}'$ or $\mathcal{C}'$ which appear in (6.6). Accordingly, $X_p/\eta X_p$ and Ψ'_q are isomorphic locally at $\mathfrak{m}$. Thus $\mathfrak{m}$ lies in the support of Ψ'_q if and only if it lies in the support of $X_p/\eta X_p$. By (6.8), the supports of Ψ_q and Ψ'_q coincide. Also, the support of $X_p/\eta X_p$ clearly consists of those $\mathfrak{m}$ in the support of X_p which contain η. Finally, whenever $\mathfrak{m}$ is prime to the flanking modules in (6.6) and $\eta \in \mathfrak{m}$, we have an isomorphism between $X_p/\mathfrak{m}X_p$ and $\Psi'_q/m\Psi'_q$. This observation, together with (6.8), gives the assertion concerning dimensions. ■

7. Proof of Theorem 1.5.

Let ρ, ℓ, M, and p be as in the statement of Theorem 1.5. In other words, we assume that $\ell \geq 3$ is a prime, that $\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$ is an irreducible representation arising from an eigenform f of weight two on $\Gamma_o(Mp)$, and that p is prime to ℓM. Suppose that ρ is unramified at p. We shall prove that ρ arises from a weight-two eigenform on $\Gamma_o(M)$.

The first step in the proof is the introduction of an "auxiliary prime" into the level (cf. the "cuspidal case" discussion in §4 above and [41, §7]). Choose a prime number q, prime to $p\ell M$, such that $\rho(\mathrm{Frob}_q)$ is conjugate to each matrix $\rho(c)$, with $c \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ a complex conjugation. The characteristic polynomial of $\rho(\mathrm{Frob}_q)$ then coincides with that of the $\rho(c)$; since ρ is an odd representation, $\rho(c)$ is conjugate to $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, and thus has characteristic polynomial

$(T-1)(T+1)$. Since the determinant of ρ is the mod ℓ cyclotomic character χ, we have $q \equiv -1 \pmod{\ell}$. (Because ℓ is odd, this congruence gives $q \not\equiv +1 \bmod \ell$.) Thus $(T-1)(T+1) = (T-a)(T-qa)$, where a can be taken to be either $+1$ or -1, so that condition II of §5 is satisfied.

An antecedent of Theorem 5.1 implies that ρ arises from a weight-two newform on $\Gamma_o(pqM)$ whose level is divisible by q. (See [38], [42], and [41, §7].) Curiously, we will not use this fact in what follows, although it will appear implicitly. Instead, we consider the Hecke algebra $\mathbf{T} = \mathbf{T}_{pqM}$ which appeared in §6 and mimic the construction of the maximal ideal $\mathfrak{m}$ of §5. Namely, let U again be the q^{th} Hecke operator in $\mathbf{T}$, and let $\eta = U^2 - 1$. The construction of §5 enables one to find a maximal ideal $\mathfrak{m}$ of $\mathbf{T}$ such that $\mathfrak{m}$ contains η, together with an embedding $\varphi\colon \mathbf{T}/\mathfrak{m} \hookrightarrow \mathbf{F}$, such that T_r is mapped to $\mathrm{trace}(\rho(\mathrm{Frob}_r))$ for almost all prime numbers r. Fix such a maximal ideal in what follows. (For the purposes of the proof of Theorem 1.5, $\mathbf{T}$ could be defined, more sparsely, as the subring of $\mathrm{End}\, S_2(\Gamma_o(pqM))$ which is generated by the T_n with n prime to pqM, along with the operator $U = T_q$. Since no "multiplicity-one" theorems appear in the following argument, it is possible, and perhaps desirable, to work with this more economical Hecke algebra.)

Our next step is to replace ρ by a model $\rho_{\mathfrak{m}}\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{T}/\mathfrak{m})$ of ρ over the finite field $\mathbf{T}/\mathfrak{m}$. Note that ρ is an irreducible finite-dimensional representation of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ over $\mathbf{F}$ and that the characteristic polynomials of $\rho(g)$ lie in $\mathbf{T}/\mathfrak{m}$ for each $g \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. It is well known that this necessary condition for descending ρ to $\mathbf{T}/\mathfrak{m}$ is also sufficient.

To descend ρ directly, we write K for $\mathbf{T}/\mathfrak{m}$, and observe that ρ descends, in any case, to some finite extension L of K. (Only a finite number of matrices are required to describe ρ.) Let W be an $L[\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$-module which affords a model of ρ over L, and let V be the vector space W, viewed as a module over $K[\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$. (Thus V is the "restriction of scalars" of W from L to K.) Let X be a minimal (non-zero) $K[\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$-submodule of W. The commutant $E := \mathrm{End}_{K[\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]} X$ of X is a division algebra by Schur's lemma; since E has finite cardinality, it is a finite commutative field. It follows that the $L[\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$-module $X \otimes_K L$ has commutant $E \otimes_K L$. Because this commutant is commutative, $X \otimes_K L$ must be multiplicity free.

On the other hand, $X \otimes_K L$ is a submodule of $V \otimes_K L$. This latter module is a direct sum $\oplus_\gamma {}^\gamma W$, where γ runs through $\mathrm{Gal}(L/K)$ and where ${}^\gamma W$ represents the twist of W by γ, i.e., the tensor product $W \otimes_L L$ where L is viewed as a base extension of itself via $\gamma\colon L \to L$. The hypothesis on characteristic polynomials ensures that the various ${}^\gamma W$ are isomorphic to W (Brauer-Nesbitt theorem). Hence $V \otimes_K L$ is isomorphic to a sum of copies of W, and it follows in particular that $X \otimes_K L$ is a sum of copies of W. Since $X \otimes_K L$ is multiplicity free, there can be only one summand. This means that $X \otimes_K L$ is isomorphic to W, so that X affords a model of ρ over K.

Now that $\rho_{\mathfrak{m}}$ has been constructed, we will abuse notation by writing simply ρ for this representation.

We will prove Theorem 1.5 by an indirect reasoning: we will assume that ρ is *not* modular of level qM, and obtain a contradiction from this assumption. This reasoning will show that ρ is modular of level qM, but then Mazur's Principle (4.7) will enable us to deduce the desired conclusion that ρ is modular of level M.

We continue to work with the series of abelian varieties which appeared in §6:

$J_o(pqM)$, $J_o(pM)^2 = J_o(pM)\times J_o(pM)$, $J_o(qM)^2 = J_o(qM)\times J_o(qM)$, and the Jacobian $J = \operatorname{Pic}^{\mathrm{o}} C$ of the Shimura curve C. We attach "multiplicities" λ and μ to $J_o(pqM)$ and J in the following manner. Consider the kernel $J[\mathfrak{m}]$ of $\mathfrak{m}$ on $J(\overline{\mathbf{Q}})$. The main theorem of [4] implies that this $\mathbf{T}/\mathfrak{m}[\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$-module is isomorphic to a direct sum $V \oplus \cdots \oplus V$, where V is a two-dimensional $\mathbf{T}/\mathfrak{m}$-vector space with a $\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$-action equivalent to ρ. We define μ to be the number of factors, and refer to it as the multiplicity of ρ in J. The multiplicity λ is defined analogously, by considering $J_o(pqM)[\mathfrak{m}]$ in place of $J[\mathfrak{m}]$.

Assume from now on that ρ is not modular of level qM. Using this assumption, we shall prove that both multiplicities λ and μ are zero. This will give the desired contradiction, since λ is positive. Indeed, $\mathbf{T}$, by definition, operates faithfully on $J_o(pqM)$, and it is well known that this implies the inequality $\lambda > 0$.

To prove that λ and μ are both zero, we will establish a series of relations among the dimensions of the $\mathbf{T}/\mathfrak{m}$-vector spaces obtained as the mod $\mathfrak{m}$ reductions of the character groups which appear in (6.2) and (6.3). The hypothesis that ρ is not modular of level qM implies that $\mathfrak{m}$ generates the unit ideal in the "p-old" quotient of $\mathbf{T}$. This implies, for example, that $\mathfrak{m}$ does not lie in the support of the $\mathbf{T}$-module X_q, since $\mathbf{T}$ operates on X_q through its p-old quotient.

(7.1) Proposition. *There are exact sequences of* $\mathbf{T}$*-modules*

$$Y_p/\mathfrak{m}Y_p \approx L_q/\mathfrak{m}L_q; \tag{7.2}$$

$$\cdots \to Y_q/\mathfrak{m}Y_q \to L_p/\mathfrak{m}L_p \to X_p/\mathfrak{m}X_p \to 0. \tag{7.3}$$

Proof. The isomorphism (7.2) and the exact sequence (7.3) are obtained from (6.2) and (6.3) by first localizing at $\mathfrak{m}$ and then reducing mod $\mathfrak{m}$. The localization at $\mathfrak{m}$ of X_q vanishes, as remarked above. ■

Proposition 7.1 implies the following relations among $\mathbf{T}/\mathfrak{m}$-dimensions:

$$\dim Y_p/\mathfrak{m}Y_p = \dim L_q/\mathfrak{m}L_q; \tag{7.4}$$

$$\dim L_p/\mathfrak{m}L_p \le \dim Y_q/\mathfrak{m}Y_q + \dim X_p/\mathfrak{m}X_p. \tag{7.5}$$

(7.6) Proposition. *We have*

$$\dim Y_p/\mathfrak{m}Y_p = 2\mu, \quad \dim L_p/\mathfrak{m}L_p = 2\lambda.$$

As in [43], these equalities are obtained by considering actions of a decomposition group D_p for p in $\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. By the main theorem of [4], the kernels of $\mathfrak{m}$ on $J_o(pqM)$, and J are direct sums of copies of ρ, first as representations of $\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, and then (by restriction) as representations of D_p. In particular, they are unramified at p.

By a well known theorem of Serre-Tate [51], the kernel of $\mathfrak{m}$ in the mod p reduction of J may be identified with the group of I_p-invariants in $J[\mathfrak{m}]$, where I_p is the inertia subgroup of D_p. Since $J[\mathfrak{m}]$ is unramified at p, the group of I_p-invariants is all of $J[\mathfrak{m}]$, and therefore $J_{/\mathbf{F}_p}(\overline{\mathbf{F}}_p)[\mathfrak{m}]$ is of dimension 2μ.

Now, as we recalled above, $J_{/\mathbf{F}_p}$ is an extension of its "group of components" Ψ_p by its toric part. Also, $\mathfrak{m}$ does not lie in the support of X_q. By (6.9), or more precisely the variant of (6.9) gotten by interchanging p and q, the $\mathbf{T}$-module Ψ_p is prime to $\mathfrak{m}$. In view of this fact, we may deduce that the toric part of $J_{\overline{\mathbf{F}}_p}(\overline{\mathbf{F}}_p)[\mathfrak{m}]$ has dimension 2μ. This gives $\dim Y_p/\mathfrak{m}Y_p = 2\mu$, which was the first of the two formulas to be proved.

The formula involving L_p is proved by an analogous argument in which we consider the mod p reduction of $J_o(pqM)$. Let A be the fiber at p of the Néron model of $J_o(pqM)$, and let B be the analogue of A for $J_o(qM) \times J_o(qM)$. Thus B is an abelian variety over $\mathbf{F}_p$. As in the discussion involving J, we know that $A(\overline{\mathbf{F}}_p)[\mathfrak{m}]$ is of dimension 2λ. In contrast to the situation for J, however, the group of components of A is "Eisenstein" ([41, Th. 3.12], or [39]), and therefore prime to $\mathfrak{m}$ [41, Th. 5.2c]. Thus, if A^o is the connected component of 0 in A, $A^o(\overline{\mathbf{F}}_p)[\mathfrak{m}]$ is of dimension 2λ.

By results of Deligne-Rapoport [11] and Raynaud, there is an exact sequence

$$0 \to T \to A^o \to B \to 0,$$

where T is the "toric part" of A^o, i.e., that torus whose character group is L_p (cf. [41, p. 446]). Since $\mathfrak{m}$ is not modular of level qM, we have $B(\overline{\mathbf{F}}_p)[\mathfrak{m}] = 0$. (Compare [41, Theorem 3.11], and the discussion in §8 below.) Therefore, $T(\overline{\mathbf{F}}_p)[\mathfrak{m}]$ has $\mathbf{T}/\mathfrak{m}$-dimension 2λ. Consequently, $L_p/\mathfrak{m}L_p$ is of dimension 2λ. ■

(7.7) Proposition. *We have*

$$\dim Y_q/\mathfrak{m}Y_q \le \mu, \quad \dim L_q/\mathfrak{m}L_q \le \lambda.$$

This proposition is proved, analogously, by considering the action of a decomposition group D_q in Gal($\overline{\mathbf{Q}}/\mathbf{Q}$) for the prime q. Pick such a group, together with a Frobenius element Frob_q in it. It will be useful to view D_q as $\text{Gal}(\overline{\mathbf{Q}}_q/\mathbf{Q}_q)$.

By the choice of q, ρ is unramified at q, and $\rho(\text{Frob}_q)$ has the distinct eigenvalues $+1$ and -1. Therefore, on $J[\mathfrak{m}]$, say, Frob_q has the eigenvalues $+1$ and -1, each with multiplicity μ, in view of the fact that $J[\mathfrak{m}]$ is a direct sum of μ copies of ρ.

By [18, §5.1] (or by the results of [51]), the $\mathbf{T}[D_q]$-module $J[\mathfrak{m}]$ contains the submodule $\text{Hom}(Y_q/\mathfrak{m}Y_q, \mu_\ell)$. (Here, μ_ℓ denotes the group of ℓ^{th} roots of unity of $\overline{\mathbf{Q}}_q$; the author expects that there will be no confusion with the multiplicity μ.)

We claim that the Frobenius element Frob_q acts on Y_q as the qth Hecke operator $U = T_q$. This claim implies that Frob_q acts on $\text{Hom}(Y_q/\mathfrak{m}Y_q, \mu_\ell)$ as the product qT_q, and therefore in particular as an element of the field $\mathbf{T}/\mathfrak{m}$. This scalar will be ±1, since U is an involution and since q is congruent to -1 (mod $\mathfrak{m}$). It follows from this that the dimension of $Y_q/\mathfrak{m}Y_q$ is bounded from above by the multiplicity μ, since each eigenvalue ±1 occurs at most μ times on $J[\mathfrak{m}]$. In other words, we obtain the desired inequality $\dim Y_q/\mathfrak{m}Y_q \le \mu$.

To prove the claim, we view Y_q as the first integral homology group of the dual graph associated with the reduced Shimura curve $C_{/\mathbf{F}_q}$. This graph may be described explicitly in terms of certain abelian varieties in characteristic q with "quaternionic multiplication" by an order $\mathcal{O}$ in a rational quaternion algebra [40, §5]. The abelian varieties A which occur in the description of the graph are all "exceptional" in the language of [40, §4]. This means that the kernel of the Frobenius morphism $A \to A^{(q)}$ coincides with the kernel of multiplication by $\mathfrak{q}$ on A, where $\mathfrak{q}$ is the unique two-sided maximal ideal of $\mathcal{O}$ whose residue field has q^2 elements. Since operator T_q has the modular description $A \mapsto A/A[\mathfrak{q}]$, we may deduce that the actions of T_q and of the Frobenius automorphism coincide on the graph.

The inequality $\dim L_q/\mathfrak{m}L_q \le \lambda$ is obtained similarly. ■

(7.8) Proposition. *We have* $\dim X_p/\mathfrak{m}X_p \le \mu$.

Let J' again denote the abelian variety dual to J and let Y'_q again be the analogue of Y_q for J'. Because J has multiplicative reduction at q, one has an exact sequence of $\mathbf{T}[\mathrm{Gal}(\overline{\mathbf{Q}}_q/\mathbf{Q}_q)]$-modules

$$0 \to \mathrm{Hom}(Y_q/\ell Y_q, \mu_\ell) \to J[\ell] \to Y'_q/\ell Y'_q \to 0. \tag{7.9}$$

Indeed, Grothendieck has constructed a "canonical filtration" [18,11.6.5] on the $\mathrm{Gal}(\overline{\mathbf{Q}}_q/\mathbf{Q}_q)$-module $A[\ell]$, whenever A is an abelian variety with semistable reduction over $\mathbf{Q}_q$. This filtration has the form

$$0 \subset A[\ell]^{\mathrm{t}} \subset A[\ell]^{\mathrm{f}} \subset A[\ell],$$

where $A[\ell]^{\mathrm{t}}$, the "toric part" of $A[\ell]$, may be written $\mathrm{Hom}(Y/\ell Y, \mu_\ell)$; here, Y is the character group of the toric part of the special fiber of the Néron model of A. Analogously (and dually), the quotient $A[\ell]/A[\ell]^{\mathrm{f}}$ of Galois modules may be identified with $Y'/\ell Y'$, where Y' is the analogue of Y for the abelian variety dual to A ([18,11.6.6], [18,11.6.7]). In case A has purely multiplicative reduction, Y and Y' are free of rank $\dim A$, and the middle terms $A[\ell]^{\mathrm{t}}$ and $A[\ell]^{\mathrm{f}}$ of the filtration coincide. One therefore obtains an exact sequence like (7.9) whenever the reduction of A is multiplicative. (For a discussion of abelian varieties with *split* multiplicative reduction, the reader may consult the expository account in Chapter III of [36]. The exact sequence (7.9) appears as Lemma 3.3.1 in [36].)

Localization of (7.9) at $\mathfrak{m}$ yields an exact sequence of $\mathbf{T}[\mathrm{Gal}(\overline{\mathbf{Q}}_q/\mathbf{Q}_q)]$-modules

$$0 \to \mathrm{Hom}(Y_q/\ell Y_q, \mu_\ell)_{\mathfrak{m}} \to J[\ell]_{\mathfrak{m}} \to (Y'_q/\ell Y'_q)_{\mathfrak{m}} \to 0. \tag{7.10}$$

This sequence *splits* as a sequence of $\mathbf{T}$-modules. Indeed, consider again a Frobenius element Frob_q of $\mathrm{Gal}(\overline{\mathbf{Q}}_q/\mathbf{Q}_q)$. This operator commutes with the action of $\mathbf{T}$ and operates on each of $\mathrm{Hom}(Y_q/\ell Y_q, \mu_\ell)_{\mathfrak{m}}$ and $(Y'_q/\ell Y'_q)_{\mathfrak{m}}$ as an element of $\mathbf{T}$. Indeed, Frob_q operates on Y_q as the involution $U = T_q$; therefore it operates as qU on $\mathrm{Hom}(Y_q/\ell Y_q, \mu_\ell)_{\mathfrak{m}}$ and as U on $(Y'_q/\ell Y'_q)_{\mathfrak{m}}$. These elements of $\mathbf{T}$ are incongruent mod $\mathfrak{m}$, since q is -1 (mod ℓ). This forces the desired splitting.

Since (7.10) splits, it remains exact after we take "kernels of $\mathfrak{m}$." In other words, we have a sequence of $(\mathbf{T}/\mathfrak{m})[\mathrm{Gal}(\overline{\mathbf{Q}}_q/\mathbf{Q}_q)]$-modules

$$0 \to \mathrm{Hom}(Y_q/\ell Y_q, \mu_\ell)[\mathfrak{m}] \to J[\mathfrak{m}] \to (Y'_q/\ell Y'_q)[\mathfrak{m}] \to 0.$$

Since Frob_q operates on $\mathrm{Hom}(Y_q/\ell Y_q, \mu_\ell)[\mathfrak{m}]$ and $(Y'_q/\ell Y'_q)[\mathfrak{m}]$ as homotheties, the dimensions of each of these modules must be μ. To see this, we observe that the eigenvalues of Frob_q on the two-dimensional representation V are $+1$ and -1. Therefore, the characteristic polynomial arising from the action of Frob_q on $J[\mathfrak{m}]$ is $(T-1)^\mu(T+1)^\mu$. Suppose that $\mathrm{Hom}(Y_q/\ell Y_q, \mu_\ell)[\mathfrak{m}]$ has dimension d_1 and that Frob_q operates on $\mathrm{Hom}(Y_q/\ell Y_q, \mu_\ell)[\mathfrak{m}]$ as the scalar ξ_1. Define d_2 and ξ_2 similarly, using $(Y'_q/\ell Y'_q)[\mathfrak{m}]$ in place of $\mathrm{Hom}(Y_q/\ell Y_q, \mu_\ell)[\mathfrak{m}]$. Then we clearly have

$$(T-1)^\mu(T+1)^\mu = (T-\xi_1)^{d_1}(T-\xi_2)^{d_2}.$$

Since -1 and $+1$ are distinct, we have $d_1 = d_2 = \mu$.

Let Ψ_q again be the group of components associated with the mod q reduction of J. As was recalled in §6, Ψ_q may be constructed from the monodromy pairing on Y_q. Indeed, if we view this pairing as an inclusion $Y_q' \hookrightarrow Y_q^*$, where $Y_q^* = \mathrm{Hom}(Y_q, \mathbf{Z})$, then Ψ_q is naturally the cokernel of this inclusion. By considering the "multiplication by ℓ" maps in the exact sequence $0 \to Y_q' \to Y_q^* \to \Psi_q \to 0$, and using the Snake Lemma, we obtain an inclusion $\Psi_q[\ell] \subseteq Y_q'/\ell Y_q'$. In particular, we have $\Psi_q[\mathfrak{m}] \subseteq (Y_q'/\ell Y_q')[\mathfrak{m}]$. Thus, μ is bounded from below by the dimension of $\Psi_q[\mathfrak{m}]$. By (6.9), this dimension coincides with that of $X_p/\mathfrak{m}X_p$, since η lies in $\mathfrak{m}$. Thus we have the desired inequality $\dim X_p/\mathfrak{m}X_p \le \mu$. ■

Using (7.4), Proposition 7.6, and Proposition 7.7, we obtain

$$2\mu \le \lambda,$$

which implies $\mu \le \lambda$, since μ is a natural number. From (7.5) and Propositions 7.6, 7.7, and 7.8, we get

$$2\lambda \le 2\mu.$$

Clearly, these inequalities imply $\lambda = \mu = 0$. As indicated above, this contradiction completes the proof of Theorem 1.5.

8. Mazur's Principle

In this §, ℓ is an odd prime, and $\mathbf{F}$ is again a fixed algebraic closure of $\mathbf{F}_\ell$. Consider an odd irreducible representation $\rho\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{GL}(2, \mathbf{F})$, and a prime number q different from ℓ. (The case $q = \ell$ may be included if one introduces the concept of "finiteness" of a representation; see [49, p. 189] and [41, Lemma 6.2].)

The following result is a variant of (4.7):

(8.1) Theorem. *Suppose that ρ arises from the space of weight-two cusp forms on $\Gamma_1(N) \cap \Gamma_o(q)$, where N is prime to q. Assume that ρ is unramified at q and that $q \not\equiv 1$ mod ℓ. Then ρ arises from the space of weight-two cusp forms on $\Gamma_1(N)$.*

The space $\mathcal{S}$ of weight-two cusp forms on $\Gamma_1(N) \cap \Gamma_o(q)$ contains (as its "q-old" subspace) a direct sum $\mathcal{S}_o$ of two copies of $S_2(\Gamma_1(N))$. This subspace is stable under the action of the Hecke operators T_n (for $n \ge 1$) on $\mathcal{S}$. Let $\mathbf{T}$ be the subring of $\mathrm{End}\,\mathcal{S}$ generated by the T_n, and let $\mathbf{T}_o$ be the q-old quotient of $\mathbf{T}$, i.e., the image of $\mathbf{T}$ in $\mathrm{End}\,\mathcal{S}_o$. These algebras contain the diamond bracket operators $\langle d \rangle$ for $d \in (\mathbf{Z}/N\mathbf{Z})^*$. The assumption that ρ arises from $\mathcal{S}$ implies that there is a homomorphism

$$\varphi\colon \mathbf{T} \to \mathbf{F}$$

satisfying

$$\varphi(T_r) = \mathrm{trace}(\rho(\mathrm{Frob}_r)), \qquad \varphi(\langle r \rangle r) = \det(\rho(\mathrm{Frob}_r))$$

for all prime numbers r prime to $\ell q N$. We view φ as an embedding $k \hookrightarrow \mathbf{F}$, where $k := \mathbf{T}/\mathfrak{m}$, and observe that there exists a model for ρ over the field k. This model is a two-dimensional k-vector space V furnished with an action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$. The $\mathbf{F}$-vector space $V \otimes_k \mathbf{F}$ affords the representation ρ. As in §5, to prove that ρ arises from $\mathcal{S}_o$, it suffices to show that $\mathfrak{m}$ arises by pullback from the quotient $\mathbf{T}_o$ of $\mathbf{T}$. We will sketch a proof that this is so; a more detailed proof, but with $\Gamma_1(N)$ replaced by $\Gamma_o(N)$, is given in [41, §6].

In the discussion below, we view V as a $\mathbf{T}$-module in the obvious way: $\mathbf{T}$ acts through its quotient k.

Consider the modular curve C attached to the group $\Gamma_1(N)\cap\Gamma_o(q)$. The operators T_n and $\langle d\rangle$ act on C as correspondences. The correspondences T_n and $\langle d\rangle$ of C induce a faithful action of $\mathbf{T}$ on the Jacobian $J := \operatorname{Pic}^{\mathrm{o}} C$. The endomorphisms of J induced by the T_n and $\langle d\rangle$ will be denoted, as usual, by the same symbols T_n and $\langle d\rangle$. The kernel $J[\mathfrak{m}]$ of $\mathfrak{m}$ on $J(\overline{\mathbf{Q}})$, which is non-zero, is a direct sum of copies of V (cf. [4]). In particular, there exist $\mathbf{T}[\operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})]$-equivariant maps $V \overset{\tilde\iota}{\hookrightarrow} J(\overline{\mathbf{Q}})$. Because V is unramified at q, and in view of the result of Serre-Tate used above, each choice $\tilde\iota$ determines a map of $\mathbf{T}[\operatorname{Gal}(\overline{\mathbf{Q}}_q/\mathbf{Q}_q)]$-modules

$$V \overset{\iota}{\hookrightarrow} \underline{J}(\overline{\mathbf{F}}_q).$$

Here, $\underline{J}$ represents the fiber over $\mathbf{F}_q$ of the Néron model of J. Fix one $\tilde\iota$, and use the resulting map ι to view V as a subgroup of $\underline{J}(\overline{\mathbf{F}}_q)$.

By results of Deligne-Rapoport [11] and Raynaud, $\underline{J}$ is semiabelian. In other words, $\underline{J}$ is an extension of its component group Φ by a connected group $\underline{J}^o$, which in turn is an extension of an abelian variety A by a torus T. More precisely, A is a product of two copies of $J_1(N)_{/\mathbf{F}_q}$, while the character group X of T may be interpreted as the group of degree-zero divisors on $X_1(N)_{/\mathbf{F}_q}$ which are concentrated at the set of supersingular points of $X_1(N)_{/\overline{\mathbf{F}}_q}$. The ring $\mathbf{T}$ acts naturally on Φ, T and A by functoriality. To show that $\mathfrak{m}$ arises by pullback from $\mathbf{T}_o$, it suffices to verify the following three points:

— The ring $\mathbf{T}$ acts on A through its quotient $\mathbf{T}_o$.

— The $\mathbf{T}$-module Φ is prime to $\mathfrak{m}$ in the sense that $\mathfrak{m}$ does not lie in the support of Φ.

— The submodule V of $\underline{J}(\overline{\mathbf{Q}}_q)$ cannot be contained in the torus T.

For the first point, the standard degeneracy maps $\alpha, \beta\colon C \rightrightarrows X_1(N)$ induce (in characteristic zero) a map on Picard varieties $\delta\colon J_1(N)\times J_1(N) \to J$ which is known to be injective (cf. [13] or [38]). It is clear that $\mathbf{T}$ acts on the abelian subvariety $J_1(N)\times J_1(N)$ of J through its old quotient $\mathbf{T}_o$. Indeed, if we dualize δ and then pass to cotangent spaces, we obtain the canonical inclusion $\mathcal{S}_o \hookrightarrow \mathcal{S}$. The map δ induces in characteristic q a map $\delta_{/\mathbf{F}_q}\colon J_1(N)_{/\mathbf{F}_q}\times J_1(N)_{/\mathbf{F}_q} \to \underline{J}^o$, and the point to be checked is that $\pi\circ\delta_{/\mathbf{F}_q}$ is an isogeny, where π is the structural map $\underline{J}^o \to A$. If we regard A as $J_1(N)_{/\mathbf{F}_q}\times J_1(N)_{/\mathbf{F}_q}$, then $\pi\circ\delta_{/\mathbf{F}_q}$ becomes a two-by-two matrix of endomorphisms of $J_1(N)_{/\mathbf{F}_q}$. This matrix, as computed by Deligne-Rapoport [11, p. 287], has diagonal components equal to the identity endomorphism, and off-diagonal components equal to the Verschiebung endomorphism (dual of the Frobenius). This matrix is then visibly an isogeny, since it acts on the cotangent space of $J_1(N)_{/\mathbf{F}_q}\times J_1(N)_{/\mathbf{F}_q}$ as the identity map.

The following circumstance is perhaps worth stressing. In the functorial action of $\mathbf{T}$ on A, the Hecke operators $\langle d\rangle$ and T_n with n prime to q act "as expected." Namely, they are induced by their namesakes on $J_1(N)$, which operate diagonally on $J_1(N)\times J_1(N)$ and then by reduction (mod q) on A. However, the functorial action of T_q on A belongs naturally to characteristic q.

For the second point, we show that Φ is *Eisenstein* in the strong sense that we have on Φ the equations $T_p = 1 + p$ and $\langle p\rangle = 1$ for all p prime to qN. These equations imply immediately that $\mathfrak{m}$ is prime to the support of Φ because ρ is irreducible; cf. [41, Theorem 5.2c]. That Φ is Eisenstein was proved in [39] for the abelian variety $J_o(qN)$. The result we seek then follows for $N \leq 3$ because the natural map $J_o(qN) \to J$ is an isomorphism.

Assume that $N \geq 4$. Recall that Φ is the cokernel of the map $X \to \mathrm{Hom}(X, \mathbf{Z})$ coming from the monodromy pairing on X. View X as the kernel of the degree map $\tilde{X} \to \mathbf{Z}$, where $\tilde{X}$ is the free abelian group on the set Σ of supersingular points of $X_1(N)_{/\overline{\mathbf{F}}_q}$. The monodromy pairing on X is the restriction to X of a diagonal pairing on $\tilde{X}$. This pairing takes the value $e(\sigma)$ on (σ, σ), where $e(\sigma)$ is essentially the number of automorphisms of any pair (E, P) which represents σ; we understand that E is a supersingular elliptic curve over $\overline{\mathbf{F}}_q$ and $P \in E(\overline{\mathbf{F}}_q)$ is a point of order N. The precise formula for (σ, σ) may be derived from [11, p. 286, Théorème 6.9]: we have

$$e(\sigma) = \begin{cases} \frac{1}{2} \# \mathrm{Aut}(E, P) & \text{if } -1 \in \mathrm{Aut}\, \sigma, \\ \# \mathrm{Aut}(E, P) & \text{if } -1 \notin \mathrm{Aut}\, \sigma. \end{cases}$$

The assumption $N \geq 4$ implies that $e(\sigma) = 1$ by [27, Corollary 2.7.4]. Hence the monodromy pairing on X is just the restriction to $X \times X$ of the standard diagonal pairing on $\tilde{X}$. It follows that the natural map

$$\xi \colon \tilde{X}/X \longrightarrow \Phi = \mathrm{Hom}(X, \mathbf{Z})/X, \qquad \sum n_\sigma \sigma \longmapsto \left\langle \sum n_\sigma \sigma, \cdot \right\rangle$$

is surjective. Since $\tilde{X}/X$ is mapped isomorphically to $\mathbf{Z}$ by the degree map, Φ is a finite cyclic group.

Fix p prime to Nq, and let T denote the usual pth Hecke operator on $\tilde{X}$: this is the $\mathbf{Z}$-linear map of $\tilde{X}$ which sends (the class of) (E, P) to $\sum(E/\mathcal{C}, P \bmod \mathcal{C})$. Here, the sum ranges over the $p{+}1$ subgroups $\mathcal{C}$ of $E[p]$ whose order is p. Let T' denote the composite $T \circ \langle p \rangle^{-1}$, where $\langle p \rangle \colon (E, p) \mapsto (E, pP)$. One checks the formula $\langle Tx, y \rangle = \langle x, T'y \rangle$, where $\langle\, , \rangle$ is the Euclidean pairing on $\tilde{X}$. Also, it is clear that both T and T' preserve the subgroup X of $\tilde{X}$ and induce the map "multiplication by $p{+}1$" on the quotient $\tilde{X}/X$.

The operator T_p on Φ is the map on $\mathrm{Hom}(X, \mathbf{Z})/X$ induced by the operators T on X and $\mathrm{Hom}(T', \mathbf{Z})$ on $\mathrm{Hom}(X, \mathbf{Z})$. The formula relating T, T', and $\langle\, , \rangle$ shows that ξ is T_p-equivariant if T_p acts naturally on $\tilde{X}/X$ by multiplication by $p{+}1$. Hence $T_p = p + 1$ on Φ. In a similar manner, we find that the diamond-bracket operators are trivial on Φ.

For the last point, we must determine the action of Frob_q on $T(\overline{\mathbf{F}}_q)$. Let w be the automorphism of C which corresponds to the map

$$(E, P, \mathcal{C}) \mapsto (E/\mathcal{C}, P \bmod \mathcal{C}, E[q]/\mathcal{C})$$

on triples consisting of an elliptic curve E, a point P on E of order M, and a cyclic subgroup $\mathcal{C}$ of E which has order q. A short computation shows that we have $w^2 = \langle q \rangle$. One checks as well that the correspondence $T_q + w_q$ of C may be written $\beta' \circ \alpha$ where α and β are again the degeneracy maps $C \rightrightarrows X_1(N)$. It follows from this that $T_q + w_q$ induces an endomorphism of $\underline{J}$ which kills the torus T, since this endomorphism factors through a map $\underline{J} \to J_1(N)$. On the other hand, w coincides with the Frobenius map on Σ but permutes the two components of $C_{/\mathbf{F}_q}$. Therefore $\mathrm{Frob}_q = -w = T_q$ on X, since X is canonically the first integral homology group of the dual graph associated with $C_{/\mathbf{F}_q}$. Thus $\mathrm{Frob}_q = qT_q$ on $T(\overline{\mathbf{F}}_q)$. Therefore, if $V \hookrightarrow T(\overline{\mathbf{F}}_q)$, then Frob_q acts on V as a scalar, i.e., as an element of k. Since the square of this scalar is $q^2\langle q\rangle$, the determinant the action of Frob_q on V is $q^2\langle q\rangle$.

On the other hand, it is clear from the Čebotarev Density Theory that the determinant of V is the character $\chi\epsilon\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to k^*$, where ϵ is the Dirichlet character $\langle\ \rangle$, i.e., where $\epsilon(\mathrm{Frob}_r) = \langle r\rangle$ for every prime number r prime to M. In particular, the determinant of the action of Frob_q on V is $q\langle q\rangle$. Therefore, we cannot have $V \hookrightarrow T(\overline{\mathbf{F}}_q)$ unless $q \equiv 1 \bmod \ell$.

References

1. A. Atkin and W. Li, *Twists of newforms and pseudo-eigenvalues of W-operators*, Invent. Math. **48** (1978), 221–243.
2. A. Ash and G. Stevens, *Modular forms in characteristic ℓ and special values of their L-functions*, Duke Math. J. **53** (1986), 849–868.
3. L. Bégueri, *Dualité sur un corps local à corps résiduel algébriquement clos*, Mémoires de la Société Mathématique de France, nouvelle série **4** (1980).
5. N. Boston, *Families of Galois representations—increasing the ramification*, Duke Math. J. **66** (1992) (to appear).
4. N. Boston, H. W. Lenstra, Jr., and K. A. Ribet, *Quotients of group rings arising from two-dimensional representations*, C. R. Acad. Sci. Paris Sér. I Math. **312** (1991), 323–328.
6. H. Carayol, *Sur les représentations ℓ-adiques associées aux formes modulaires de Hilbert*, Ann. scient. Éc. Norm. Sup., 4$^{\mathrm{e}}$ série **19** (1986), 409–468.
7. ———, *Sur les représentations galoisiennes modulo ℓ attachées aux formes modulaires*, Duke Math. J. **59** (1989), 785–801.
8. I. V. Cerednik, *Uniformization of algebraic curves by discrete arithmetic subgroups of* $\mathrm{PGL}_2(k_w)$ *with compact quotients*, Mat. Sb. **100** (1976), 59–88; English transl. in Math USSR Sb. **29** (1976).
9. R. Coleman and J. Voloch, *Companion forms and Kodaira-Spencer theory* (to appear).
10. P. Deligne, *Formes modulaires et représentations ℓ-adiques*, Sém. Bourbaki n$^{\mathrm{o}}$ 355 (1968/69), Lecture Notes in Math., vol. 179, Springer-Verlag, Berlin and New York, 1971, pp. 139–172.
11. P. Deligne and M. Rapoport, *Les schémas de modules de courbes elliptiques*, Lecture Notes in Math., vol. 349, Springer-Verlag, Berlin and New York, 1973, pp. 143–316.
12. P. Deligne and J-P. Serre, *Formes modulaires de poids* 1, Ann. scient. Éc. Norm. Sup., 4$^{\mathrm{e}}$ série **7** (1974), 507–530.
13. F. Diamond, *Congruence primes for cusp forms of weight $k \geq 2$*, Astérisque **196–197** (1991), 205–213.
14. V. G. Drinfeld, *Coverings of p-adic symmetric regions*, Functional. Anal. i Prilozen. **10** (1976), 29–40; English transl. in Functional Anal. Appl. **10** (1976).
15. B. Edixhoven, *The weight in Serre's conjectures on modular forms*, Invent. Math. **109** (1992), 563–594.
16. G. Faltings and B. Jordan, *Crystalline cohomology and* $\mathbf{GL}(2,\mathbf{Q})$ (to appear).
17. B. H. Gross, *A tameness criterion for Galois representations associated to modular forms* mod p, Duke Math. J. **61** (1990), 445–517.
18. A. Grothendieck, *SGA7 I, Exposé IX*, Lecture Notes in Math., vol. 288, Springer-Verlag, Berlin and New York, 1972, pp. 313–523.
19. H. Hida, *Congruences of cusp forms and special values of their zeta function*, Invent. Math. **63** (1981), 225–261.
20. H. Hida, *Iwasawa modules attached to congruences of cusp forms*, Ann. scient. Éc. Norm. Sup., 4$^{\mathrm{e}}$ série **19** (1986), 231–273.
21. H. Hida, *Galois representations into* $\mathbf{GL}_2(\mathbf{Z}_p[[X]])$ *attached to ordinary cusp forms*, Invent. Math. **85** (1986), 545–613.
22. K. Iwasawa, *Lectures on p-adic L-functions*, Annals of Math. Studies **74**, Princeton University Press, Princeton, 1972.
23. B. Jordan and R. Livné, *On the Néron model of Jacobians of Shimura curves*, Compositio Math. **60**, 227–236.
24. ———, *Conjecture "epsilon" for weight $k > 2$*, Bull. AMS **21** (1989), 51–56.

25. N. M. Katz, *Higher congruences between modular forms*, Ann. of Math **101** (1975), 332–367.
26. ______, *A result on modular forms in characteristic p*, Lecture Notes in Math., vol. 601, Springer-Verlag, Berlin and New York, 1977, pp. 53–61.
27. N. M. Katz and B. Mazur, *Arithmetic moduli of elliptic curves*, Annals of Math. Studies **108**, Princeton University Press, Princeton, 1985.
28. R. P. Langlands, *Modular forms and ℓ-adic representations*, Lecture Notes in Math., vol. 349, Springer-Verlag, Berlin and New York, 1973, pp. 361–500.
29. W. Li, *Newforms and functional equations*, Math. Ann. **212** (1975), 285–315.
30. R. Livné, *On the conductors of mod ℓ Galois representations coming from modular forms*, Journal of Number Theory **31** (1989), 133–141.
31. B. Mazur, *Letter to J-F. Mestre (16 August 1985)*, unpublished.
32. B. Mazur and K. A. Ribet, *Two-dimensional representations in the arithmetic of modular curves*, Astérisque **196–197** (1991), 215–255.
33. W. G. McCallum, *Duality theorems for Néron models*, Duke Math. J. **53** (1986), 1093–1124.
34. T. Miyake, *Modular Forms*, Springer-Verlag, Berlin and New York, 1989.
35. C. Queen, *The existence of p-adic abelian L-functions*, Number Theory and Algebra, Edited by H. Zassenhaus, Academic Press, New York, 1977.
36. K. A. Ribet, *Galois action on division points on abelian varieties with many real multiplications*, Am. J. Math. **98** (1976), 751–804.
37. ______, *The ℓ-adic representations attached to an eigenform with Nebentypus: a survey*, Lecture Notes in Math., vol. 601, Springer-Verlag, Berlin and New York, 1977, pp. 17–52.
38. ______, *Congruence relations between modular forms*, Proc. Int. Cong. of Mathematicians 1983, pp. 503–514.
39. ______, *On the component groups and the Shimura subgroup of $J_o(N)$*, exposé 6, Sém. Th. Nombres, Université Bordeaux (1987–88).
40. ______, *Bimodules and Abelian surfaces*, Advanced Studies in Pure Mathematics **17** (1989), 359–407.
41. ______, *On modular representations of* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ *arising from modular forms*, Invent. Math. **100** (1990), 431–476.
42. ______, *Raising the levels of modular representations*, Progress in Math. **81** (1990), 259–271.
43. ______, *Lowering the levels of modular representations without multiplicity one*, International Mathematics Research Notices (1991), 15–19.
44. J-P. Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Invent. Math. **15** (1972), 259–331.
45. ______, *Congruences et formes modulaires [d'après H. P. F. Swinnerton-Dyer]*, Sém. Bourbaki n^o 416 (1971/72), Lecture Notes in Math., vol. 317, Springer-Verlag, Berlin and New York, 1973, pp. 319–338.
46. ______, *Formes modulaires et fonctions zêta p-adiques*, Lecture Notes in Math., vol. 350, Springer-Verlag, Berlin and New York, 1973, pp. 191–268.
47. ______, *Divisibilité de certaines fonctions arithmétiques*, Ens. Math. **22** (1976), 227–260.
48. ______, *Lettre à J-F. Mestre (13 août 1985)*, Contemporary Mathematics **67** (1987), 263–268.
49. ______, *Sur les représentations modulaires de degré 2 de* $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, Duke Math. J. **54** (1987), 179–230.
50. ______, *Letter to K. Ribet (15 April 1987)*, unpublished.
51. J-P. Serre and J. Tate, *Good reduction of abelian varieties*, Ann. of Math. **88** (1968), 492–517.
52. G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, Princeton, 1971.
53. H. P. F. Swinnerton-Dyer, *On ℓ-adic representations and congruences for modular forms*, Lecture Notes in Math., vol. 350, Springer-Verlag, Berlin and New York, 1973, pp. 1–55.
54. John T. Tate, *Number Theoretic Background*, Proceedings of Symposia in Pure Mathematics **33** (1979), (2) 3–26.

UC Mathematics Department, Berkeley, CA 94720 USA
E-mail address: ribet@math.berkeley.edu