Optimal Levels for Modular Mod 2 Representations over Totally Real Fields

To John Coates, with thanks and best wishes

Frazer Jarvis

Received: August 30, 2005
Revised: January 26, 2006

Abstract. In this paper, we study the level lowering problem for mod 2 representations of the absolute Galois group of a totally real field F. In the case $\mathrm{F} = \mathbb{Q}$, this was done by Buzzard; here, we generalise some of Buzzard's results to higher weight and arbitrary totally real fields, using Rajaei's generalisation of Ribet's theorem and previous work of Fujiwara and the author.

2000 Mathematics Subject Classification: 11F33, 11F41, 11G18, 14G35
Keywords and Phrases: Hilbert modular forms, totally real fields, Galois representations

The main theorem of this paper is the following result, which reduces level lowering for the prime $\ell = 2$ for totally real fields to a multiplicity one hypothesis, thus showing that multiplicity one is the only obstruction to level lowering in characteristic 2.

Theorem 0.1 *Let* F *be a totally real number field. Let*

$$\overline{\rho} : \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_2)$$

be a continuous irreducible representation such that $\overline{\rho}$ is not induced from a character of $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}(i))$*. Let $\mathfrak{n}(\overline{\rho})$ denote the Artin conductor away from 2 of $\overline{\rho}$. Suppose that there is some Hilbert cuspidal eigenform of arithmetic weight k and level $U_1(\mathfrak{n})$ that gives rise to $\overline{\rho}$, where $(2, \mathfrak{n}) = 1$. Suppose also that $\overline{\rho}$ satisfies a certain multiplicity one hypothesis (see Definition 6.1). Then there is a Hilbert cuspidal eigenform of weight k and level $U_1(\mathfrak{n}(\overline{\rho}))$ that gives rise to $\overline{\rho}$.*

We note that it is part of the hypothesis that $\overline{\rho}$ occurs at some level prime to 2. This will not be true in general, but it makes for a comparatively clean statement, and the reader will easily be able to extend the statement if needed. The question of possible level structures at 2 is more naturally considered in the context of optimising the weight, we shall not address this problem here; this is the subject of work in progress with Kevin Buzzard and Fred Diamond. We remark (as Fujiwara [11] also explains) that the methods in this paper show that if ℓ is odd, the same result holds for characteristic ℓ representations without the multiplicity one hypothesis. We have:

Theorem 0.2 *Let ℓ be an odd prime. Let*

$$\overline{\rho} : \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$$

be a continuous irreducible representation. If $[\mathrm{F}(\mu_\ell) : \mathrm{F}] = 2$, suppose that $\overline{\rho}$ is not induced from a character of $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}(\mu_\ell))$. Let $\mathfrak{n}(\overline{\rho})$ denote the Artin conductor away from ℓ of $\overline{\rho}$. Suppose that there is some Hilbert cuspidal eigenform of arithmetic weight k and level $U_1(\mathfrak{n}\mathfrak{p}^r)$ that gives rise to $\overline{\rho}$. If $r > v_\mathfrak{p}(\mathfrak{n}(\overline{\rho}))$, then there is a Hilbert cuspidal eigenform of weight k and level $U_1(\mathfrak{n}\mathfrak{p}^{r-1})$ that gives rise to $\overline{\rho}$.

This result has no multiplicity one hypothesis, and also allows us to lower the level if ℓ and $\mathfrak{n}$ are not coprime, so as to lower the level by all primes not dividing the characteristic. We recall that Fujiwara's work remains unpublished, but alternative references are available for all but his version of Mazur's Principle when $[\mathrm{F} : \mathbb{Q}]$ is even. In particular, when $[\mathrm{F} : \mathbb{Q}]$ is odd, this theorem does not depend on Fujiwara's unpublished work.
We will concentrate on the case $\ell = 2$, as Theorem 0.2 is an easy corollary of previous results of Fujiwara ([11]) and Rajaei ([18]). However, the case $\ell = 2$ requires additional work, and combines Rajaei's results with ideas of Buzzard ([3]), which in turn are based on work of Ribet, for the case $\mathrm{F} = \mathbb{Q}$.

1 Notation

Our notation completely follows [15], and we summarise it next. Throughout this paper, F will denote a totally real number field of degree d over $\mathbb{Q}$. Let $I = \{\tau_1, \ldots, \tau_d\}$ denote the set of embeddings $\mathrm{F} \hookrightarrow \mathbb{R}$. If $\mathfrak{p}$ is a prime of F, then we will denote the local ring at $\mathfrak{p}$ by $\mathcal{O}_\mathfrak{p}$ and its residue field by $\kappa_\mathfrak{p}$. We will be considering continuous semisimple representations

$$\overline{\rho} : \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_2),$$

and we study such representations which are associated to Hilbert modular forms.
The *weight* of a Hilbert modular form will be a d-tuple of integers, $k \in \mathbb{Z}^I$, so that there is one component for each infinite place of F.

Definition 1.1 We say that k is *arithmetic* if $k_\tau \geq 2$ for all $\tau \in I$, and further, if all k_τ have the same parity.

Throughout this paper, weights will always be supposed arithmetic.
Given a weight k, we define a d-tuple $v \in \mathbb{Z}^I_{\geq 0}$ so that $k+2v$ is *parallel* (i.e., $k_\tau + 2v_\tau$ is independent of τ), and some $v_\tau = 0$. We also write $t = (1, \ldots, 1)$, so that $k+2v$ is a multiple of t. The transformation law for Hilbert modular forms is normalised by choosing a vector $w = k + v - \alpha.t$ for some integer α, and we choose $\alpha = 1$ (as in [12]). If $x \in \mathbb{Z}^I$ is parallel, we write $x = [x]t$ so $[x] \in \mathbb{Z}$.
If G denotes the algebraic group $\mathrm{Res}_{\mathrm{F}/\mathbb{Q}}(\mathrm{GL}_2)$ with centre Z, then the *level* of a Hilbert modular form is an open compact subgroup U of $G(\mathbb{A}^\infty)$, where $\mathbb{A}^\infty$ denotes the finite adeles of $\mathbb{Q}$. In this paper, we will only ever consider subgroups U of the form $U = \prod_{\mathfrak{p}} U_{\mathfrak{p}}$, where $U_{\mathfrak{p}}$ is an open compact subgroup of $\mathrm{GL}_2(\mathrm{F}_{\mathfrak{p}})$. The (finite-dimensional complex) vector space of Hilbert cusp forms of weight k and level U will be denoted $S_k(U)$ (see [12], (2.3), for the precise definition of this space, where it is denoted $S_{k,w,I}(U;\mathbb{C})$; in [15], it is denoted $S_{k,w}(U)$).
Suppose B is a quaternion algebra over F and let $S(B)$ denote the set of finite places ramifying in B. If S is any finite set of finite places of F containing $S(B)$, we define the Hecke algebra $\mathbb{T}^{S,B}$ as the $\mathbb{Z}$-algebra generated by all the Hecke operators $T_{\mathfrak{q}}$ with $\mathfrak{q} \notin S(B)$, and the operators $S_{\mathfrak{q}}$ for $\mathfrak{q} \notin S$. If U is as above, and S contains all finite places at which $U_{\mathfrak{q}}$ is not maximal compact, then $\mathbb{T}^{S,B}$ acts on $S_k(U)$ through a quotient which we denote $\mathbb{T}^{S,B}_k(U)$. If $B = \mathrm{GL}_2$, we will omit it from the notation. If S consists precisely of $S(B)$ together with the places $\mathfrak{q}$ such that $U_{\mathfrak{q}}$ is not maximal compact in $(\mathcal{O}_B \otimes \mathcal{O}_{\mathfrak{q}})^\times \cong \mathrm{GL}_2(\mathcal{O}_{\mathfrak{q}})$, then we will omit it from the notation.

2 Preliminaries

Carayol ([5]) and Taylor ([19]) have proven that to any Hilbert cuspidal eigenform, one may attach a compatible system of global Galois representations compatible with the local Langlands correspondence. For a statement, see [19] or [15].
This result leads us to examine analogues of the Serre conjectures for Galois representations over totally real fields.

Definition 2.1 Given an irreducible modulo ℓ representation,

$$\overline{\rho} : \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_\ell),$$

we say that $\overline{\rho}$ is *modular of level U and weight k* if there exists a Hilbert cuspidal eigenform $f \in S_k(U)$ and a prime $\lambda | \ell$ of $\mathcal{O}_f$ (the ring of integers of the number field generated by the Hecke eigenvalues) such that $\overline{\rho}$ is isomorphic to the reduction of $\rho_{f,\lambda}$ mod λ. As we will primarily be interested in the case $U = U_1(\mathfrak{n})$, we will simply say that $\overline{\rho}$ is *modular of level* $\mathfrak{n}$ if it is modular of level $U_1(\mathfrak{n})$.

(Note that we only consider here Hilbert modular forms coming from characteristic zero, and do not think about mod ℓ forms in the sense of Katz.)
In general, a modular $\overline{\rho}$ may have many different weights and levels, even when we insist that f is a newform (i.e., does not occur at a lower level). We are interested in this paper in the smallest possible levels that may arise.
We define an "optimal" level as in [15]. Given $\overline{\rho}$, define

$$\mathfrak{n}(\overline{\rho}) = \text{the Artin conductor (away from } \ell\text{) of } \overline{\rho}.$$

Note that $\mathfrak{n}(\overline{\rho})$ is prime to ℓ. It is not true that if $\overline{\rho}$ is modular, then it is modular of level $\mathfrak{n}(\overline{\rho})$ (one can see that one cannot always remove primes dividing the characteristic simply by looking at det $\overline{\rho}$). To get a clean statement, however, we will be assuming that our representation is modular at a level prime to ℓ, and try to remove primes not dividing the characteristic. The reader will be able to adapt the statement to more general situations if required.
Let $\mathfrak{p} \nmid \ell$ be prime. For any ℓ-adic character χ of $D_{\mathfrak{p}}$, the decomposition group at $\mathfrak{p}$, let $\overline{\chi}$ denote the reduction modulo ℓ, and let $a(\chi)$ denote the $\mathfrak{p}$-adic valuation of the conductor. In [15], one finds the following generalisation of a result of Carayol ([6]) and Livné ([17]).

Theorem 2.2 *Suppose π is an automorphic representation of $\mathrm{GL}_{2/\mathrm{F}}$ giving rise to $\overline{\rho}$. If π has conductor $\mathfrak{n}$, write $n_{\mathfrak{p}} = v_{\mathfrak{p}}(\mathfrak{n})$. Write $\overline{n}_{\mathfrak{p}} = v_{\mathfrak{p}}(\mathfrak{n}(\overline{\rho}))$. Then one always has $\overline{n}_{\mathfrak{p}} \leq n_{\mathfrak{p}}$ (so $\mathfrak{n}(\overline{\rho})|\mathfrak{n}$), and one has equality except possibly in the following cases:*

1. *$\pi_{\mathfrak{p}}$ is special, associated to a character χ of $\mathrm{F}_{\mathfrak{p}}^{\times}$ which is unramified.*
2. *$\pi_{\mathfrak{p}}$ is special, associated to a character χ of $\mathrm{F}_{\mathfrak{p}}^{\times}$ which degenerates, in that $a(\chi) = 1$ and $a(\overline{\chi}) = 0$.*
3. *$\pi_{\mathfrak{p}}$ is principal series, associated to two characters χ and ψ of $\mathrm{F}_{\mathfrak{p}}^{\times}$, with at least one of the characters degenerating.*
4. *$\pi_{\mathfrak{p}}$ is a supercuspidal Weil representation, associated to a character of $\Omega^{\times}$ which degenerates, where Ω is the unramified quadratic extension of $\mathrm{F}_{\mathfrak{p}}$.*

For a character of $\mathrm{F}_{\mathfrak{p}}^{\times}$ to degenerate, we require that $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{p}) \equiv 1 \pmod{\ell}$, and for a character of $\Omega^{\times}$ to degenerate, we require that $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{p}) \equiv -1 \pmod{\ell}$.

Conjecture 2.3 Suppose that $\overline{\rho}$ is modular of weight k and level $U_1(\mathfrak{n})$ with $(\mathfrak{n}, \ell) = 1$. If $v_{\mathfrak{p}}(\mathfrak{n}) > v_{\mathfrak{p}}(\mathfrak{n}(\overline{\rho}))$ and $\mathfrak{p} \nmid \ell$, then it is modular of weight k and level $U_1(\mathfrak{n}/\mathfrak{p})$.

As usual, the method is to remove one prime at a time from the level. The theorem above classifies the primes which may occur.

Remark 2.4 We first remark that cases (2)–(4) of the classification above (Theorem 2.2) will be treated with the existing methods, once we have a good notion of auxiliary prime. Indeed, [15] treats cases (2) and (4), and [11] also treats (3) (as well as doing (2) and (4) independently). One readily verifies that all proofs continue to hold in the case of mod 2 representations, as long as auxiliary primes are available. We will explain later that Buzzard's construction of auxiliary primes generalises to the totally real case.

Remark 2.5 We also point out for later use that the proof of "Mazur's Principle" (the case $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{p}) \not\equiv 1 \pmod{\ell}$ of case (1)) given in [14] for $[\mathrm{F}:\mathbb{Q}]$ odd, and [11] for $[\mathrm{F}:\mathbb{Q}]$ even, is valid more generally when $\overline{\rho}(\mathsf{Frob}_{\mathfrak{p}})$ is not a scalar (see [14], Corollary 18.8, and [11], end of §5). As the ratio of the diagonal entries of $\overline{\rho}(\mathsf{Frob}_{\mathfrak{p}})$ is equal to $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{p})$, this is implied by the congruence condition; however, we will later need the stronger version—after all, the congruence condition will never be satisfied when $\ell = 2$. The condition that ℓ be odd was imposed in [14] and [11] only in order that auxiliary primes should exist; as in the previous remark, our generalisation of Buzzard's construction of auxiliary primes then implies some cases of Mazur's Principle even when $\ell = 2$.

3 Characters and Carayol's Lemma

One of the crucial technical tricks used in the theory of level lowering is a result known as Carayol's Lemma (after [6]). The version we want to use was proven in [15].

Roughly, Carayol's Lemma allows us to show that if a mod ℓ representation $\overline{\rho}$ is modular, associated to some modular form with character ϕ, then given any character $\psi \equiv \phi \pmod{\ell}$, there is some modular form with character ψ which gives rise to the given mod ℓ representation. In other words, whether or not there is a modular form of given character giving rise to $\overline{\rho}$ depends only on the character modulo ℓ.

In [15], we proved the following (for additional notation, as well as a more general statement, see [15]):

Theorem 3.1 (Carayol's Lemma) *Let ℓ be an odd prime. Suppose that S is a finite set of places of* F *containing all infinite places of* F*. Let k be arithmetic, and let U and U' be S-subgroups with U' normal in U. Suppose r (resp. χ) is an irreducible representation (resp. a character with ℓ-power order) of $U/U'(U \cap Z(\mathbb{Q}))$. Let $\overline{\theta} : \mathbb{T}_k^S(U,r) \longrightarrow \overline{\mathbb{F}}_\ell$ be a homomorphism for which $\overline{\rho} = \overline{\rho}_{\overline{\theta}}$ is irreducible. If $[\mathrm{F}(\mu_\ell):\mathrm{F}] = 2$, suppose that $\overline{\rho}$ is not induced from a character of the kernel of the mod ℓ cyclotomic character. Then there exists a homomorphism $\overline{\theta}' : \mathbb{T}_k^S(U, r\otimes\chi) \longrightarrow \overline{\mathbb{F}}_\ell$ such that the two maps $\mathbb{T}^S \longrightarrow \overline{\mathbb{F}}_\ell$ induced by $\overline{\theta}$ and $\overline{\theta}'$ coincide.*

The additional hypothesis when $[\mathrm{F}(\mu_\ell):\mathrm{F}] = 2$ was not explicitly stated in [15], as it was a running hypothesis throughout the paper. (The author apologises if this has caused any confusion.) This hypothesis on ℓ, as well as the stipulation that ℓ be odd, was only invoked to allow us to introduce an auxiliary prime (see next section) so that U—and therefore U'—may be assumed *sufficiently small* in the sense of Carayol ([4], 1.4.1.1, 1.4.1.2). However, as a corollary of the proof, we can omit this hypothesis, and still deduce the same result (even for $\ell = 2$), so long as U is sufficiently small.

Corollary 3.2 *Suppose the notation and hypotheses are as above, except that we replace the hypotheses on ℓ by the hypothesis that U is sufficiently small. Then Carayol's Lemma is again true.*

4 Auxiliary primes

A crucial trick that we will use is to alter the level by making it sufficiently small, and then return to the original level. This trick originated in [8] and [9] and easily generalises to totally real fields; the case of mod 2 representations was treated in [3].

Let G be a finite group. Suppose that

$$\overline{\rho} : G \longrightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_2)$$

is an irreducible continuous representation and let

$$\chi : G \longrightarrow \{\pm 1\}$$

be a surjective character.

In our applications, $\overline{\rho}$ will be the given Galois representation. We will let χ be the mod 4 cyclotomic character giving the action of $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F})$ on the fourth roots of unity. As F is totally real, its absolute Galois group contains complex conjugation elements, so this mod 4 cyclotomic character is non-trivial, and hence maps surjectively onto $\{\pm 1\}$. As $\overline{\rho}$ is continuous, it factors through a finite group; we let G be a finite group through which $\overline{\rho} \oplus \chi$ factors.

We say that $g \in G$ is *special* if $\mathrm{tr}\, \overline{\rho}(g) = 0$.

We will need the following lemma.

Lemma 4.1 *Suppose that $\overline{\rho}$ is not induced from a character of $\ker \chi$. Then there exists $g \in G$ which is not in $\ker \chi$ and which is not special.*

Proof. See [3]. □

When χ is the mod 4 cyclotomic character, its kernel is precisely $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}(i))$.

We now apply this lemma to construct "auxiliary" primes. From the lemma, we see that if

$$\overline{\rho} : G \longrightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_2)$$

is an irreducible mod 2 representation, then there exists an element $g \in G$ such that $\chi(g) = -1$ and $\mathrm{tr}\, \overline{\rho}(g) \neq 0$. If χ is the mod 4 cyclotomic character, and $g \in G$ is the image of $\mathsf{Frob}_{\mathfrak{q}}$, then $\chi(g) = -1$ is equivalent to $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{q}) \equiv 3 \pmod 4$.

Lemma 4.2 *Let $\mathfrak{q}$ be a prime ideal of $\mathcal{O}_{\mathrm{F}}$ with $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{q}) \equiv 3 \pmod 4$. Suppose that E is a finite extension of $\mathbb{Q}_2$ and that*

$$\psi : (\mathcal{O}_{\mathrm{F}}/\mathfrak{q})^{\times} \longrightarrow \mathcal{O}_E^{\times}$$

is a character with trivial reduction. If $\psi(-1) = 1$, then ψ is trivial.

Proof. As ψ has trivial reduction, it is valued in $\ker(\mathcal{O}_E^{\times} \longrightarrow k_E^{\times})$, where k_E denotes the residue field of $\mathcal{O}_E$. It is easy to see (using Hensel's lemma, for example) that the only torsion in this kernel is killed by a power of 2. But

$$|(\mathcal{O}_{\mathrm{F}}/\mathfrak{q})^{\times}| = N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{q}) - 1 \equiv 2 \pmod 4.$$

It follows that ψ is valued in $\{\pm 1\}$. If also $\psi(-1) = 1$, then the 2-torsion of $(\mathcal{O}_{\mathrm{F}}/\mathfrak{q})^{\times}$ is killed by ψ, so ψ must be trivial. □

The following result generalises Corollary 2.6 of [3] to totally real fields.

Theorem 4.3 *Suppose that $\overline{\rho}$ is an irreducible continuous mod 2 representation not induced from a character of $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}(i))$. Then there exists a prime $\mathfrak{r}$ with the property that if $\overline{\rho}$ is associated to a cuspidal eigenform f of weight k and level $U_1(\mathfrak{n}) \cap U_1^1(\mathfrak{r})$ for some $\mathfrak{n}$ prime to $\mathfrak{r}$, then there is a cuspidal eigenform g of weight k and level $U_1(\mathfrak{n})$ which also gives rise to $\overline{\rho}$.*

Proof. For this, we apply Carayol's Lemma and the lemmas above. Let G be a finite group through which $\overline{\rho} \oplus \chi$ factors. By the hypotheses above on $\overline{\rho}$, we can find an element $g \in G$ which is neither in $\ker \chi$, nor is special. Let $\mathfrak{r}$ be any prime such that $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{r}) > 4^d$ which is unramified for $\overline{\rho}$ such that $\mathsf{Frob}_{\mathfrak{r}}$ maps to g. Then

- $\chi(\mathsf{Frob}_{\mathfrak{r}}) = -1$, i.e., $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{r}) \equiv 3 \pmod 4$, and
- $\mathrm{tr}\,\overline{\rho}(\mathsf{Frob}_{\mathfrak{r}}) \neq 0$.

As $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{r}) > 4^d$, the group $U_1^1(\mathfrak{r})$ is sufficiently small ([14], §12) so that we may apply Corollary 3.2. We may regard f as an eigenform on $U_0(\mathfrak{n}\mathfrak{r})$ with character $\chi = \chi_{\mathfrak{n}}\chi_{\mathfrak{r}}$ of the abelian group $U_0(\mathfrak{n})/U_1(\mathfrak{n}) \times U_0(\mathfrak{r})/U_1^1(\mathfrak{r})$. We apply Corollary 3.2 with $U = U_0(\mathfrak{n}) \cap U_1^1(\mathfrak{r})$ and $U' = U_1(\mathfrak{n}) \cap U_1^1(\mathfrak{r})$ to see that there is a cuspidal eigenform f' on $U_1(\mathfrak{n}) \cap U_1^1(\mathfrak{r})$ with character $\chi' = \chi'_{\mathfrak{n}}\chi'_{\mathfrak{r}}$ which is congruent to f and such that $\chi'_{\mathfrak{n}}(-1) = (-1)^{[k+2v]}$.
However, one knows that $\chi'(-1) = (-1)^{[k+2v]}$, so that $\chi'_{\mathfrak{r}}(-1) = 1$. As $\overline{\rho}$ is unramified at $\mathfrak{r}$, the reduction of $\chi'_{\mathfrak{r}}$ is trivial. By Lemma 4.2, $\chi'_{\mathfrak{r}}$ itself is trivial. It follows that f' is actually a cuspidal eigenform on $U_1(\mathfrak{n}) \cap U_0(\mathfrak{r})$. This implies that the component at $\mathfrak{r}$ of the automorphic representation corresponding to f' is either unramified principal series or is special unramified. The latter is ruled out as then we would have $\mathrm{tr}\,\overline{\rho}(\mathsf{Frob}_{\mathfrak{r}}) = 0$. Thus f' is old at $\mathfrak{r}$, and we may choose an eigenform g for level $U_1(\mathfrak{n})$ with the same Hecke eigenvalues as those for f' except possibly at $\mathfrak{r}$. The result follows. □

5 Shimura curves

In this section, we summarise results (in §5.2) of [14] on integral models and their reductions in characteristic p of Shimura curves whose level structure involves primes dividing p, but where such primes do not ramify in the quaternion algebra, and the relevant results (in §5.3) of Boutot-Zink ([2]) and Varshavsky ([20], [21]) on integral models and their reductions in characteristic p of Shimura curves whose level structure does not involve primes dividing p, but where such primes do ramify in the quaternion algebra. These results also appear in [18]; this section is as much to fix notation as it is to remind the reader of previous results.

5.1 Formalism of vanishing cycles

Here we summarise the theory of vanishing cycles from SGA 7, XIII, XV. For a beautiful introduction to the theory, see also [18].

Let V be a mixed characteristic henselian discrete valuation ring with fraction field K and residue field k of characteristic p. If $\mathcal{C}$ is a proper generically smooth curve over $S = \operatorname{spec} V$ with semistable reduction, and $\mathcal{F}$ is a constructible sheaf on $\mathcal{C}$ with torsion prime to p, then we have the following exact sequence:

$$\begin{aligned} 0 \longrightarrow H^1(\mathcal{C}\otimes\overline{k},\mathcal{F}) \longrightarrow H^1(\mathcal{C}\otimes\overline{K},\mathcal{F}) \stackrel{\beta}{\longrightarrow} H^1(\mathcal{C}\otimes\overline{k},R\Phi\mathcal{F}) \\ \longrightarrow H^2(\mathcal{C}\otimes\overline{k},\mathcal{F}) \longrightarrow H^2(\mathcal{C}\otimes\overline{K},\mathcal{F}) \longrightarrow 0 \end{aligned}$$

where $R\Phi\mathcal{F}$ is a complex of sheaves supported on the singular points Σ of the special fibre, and $R^i\Phi\mathcal{F} \neq 0$ only when $i = 1$. In particular,

$$H^1(\mathcal{C}\otimes\overline{k},R\Phi\mathcal{F}) = \bigoplus_{x\in\Sigma}(R^1\Phi\mathcal{F})_x.$$

Furthermore, we have a complete understanding of the sheaf $R^1\Phi\mathcal{F}$ coming from the Picard-Lefschetz formula. We define

$$X(\mathcal{C},\mathcal{F}) = \operatorname{im}(\beta)(1)$$

where the (1) denotes the Tate twist.

The cohomology $H^1(\mathcal{C}\otimes\overline{K},\mathcal{F})$ may be computed by means of another complex $R\Psi\mathcal{F}$ of sheaves on $\mathcal{C}\otimes\overline{k}$.

There is a trace pairing in this situation; Illusie ([13]) has explained that this induces a second exact sequence, dual to the first:

$$\begin{aligned} 0 \longrightarrow H^0(\widehat{\mathcal{C}\otimes\overline{k}},R\Psi\mathcal{F}) \longrightarrow H^0(\widehat{\mathcal{C}\otimes\overline{k}},\mathcal{F}) \longrightarrow \bigoplus_{x\in\Sigma} H^1_x(\mathcal{C}\otimes\overline{k},R\Psi\mathcal{F}) \\ \stackrel{\beta'}{\longrightarrow} H^1(\mathcal{C}\otimes\overline{K},\mathcal{F}) \longrightarrow H^1(\mathcal{C}\otimes\overline{k},\mathcal{F}) \longrightarrow 0, \end{aligned}$$

where $\widehat{\mathcal{C}\otimes\overline{k}}$ denotes the normalisation of the special fibre of $\mathcal{C}$. We set

$$\widehat{X}(\mathcal{C},\mathcal{F}) = \mathrm{im}(\beta').$$

Rajaei ([18], prop.1) points out that $\widehat{X}(\mathcal{C},\mathcal{F})$ actually lies inside $H^1(\mathcal{C}\otimes\overline{k},\mathcal{F})$, regarded a subspace of $H^1(\mathcal{C}\otimes\overline{K},\mathcal{F})$ by the first exact sequence. The first exact sequence is called the *specialisation exact sequence* and the second is called the *cospecialisation exact sequence*.
Deligne defines a *variation map*

$$\mathrm{Var}(\sigma)_x : (R^1\Phi\mathcal{F})_x \longrightarrow H^1_x(\mathcal{C}\otimes\overline{k},\mathcal{F})$$

for $\sigma\in I_K$, the inertia group of $\mathrm{Gal}(\overline{K}/K)$. The action of $\sigma\in I_K$ on both exact sequences may be expressed using this map; for example, the action of σ on $H^1(\mathcal{C}\otimes\overline{K},\mathcal{F})$ is given by $\mathrm{id}+(\beta'\circ\bigoplus_{x\in\Sigma}\mathrm{Var}(\sigma)_x\circ\beta)$. From the form of the variation map, one may define a canonical *monodromy logarithm*

$$N_x : (R^1\Phi\mathcal{F})_x(1) \longrightarrow H^1_x(\mathcal{C}\otimes\overline{k},\mathcal{F}).$$

Rajaei explains that this monodromy map induces an injective map $\lambda : X(\mathcal{C},\mathcal{F}) \longrightarrow \widehat{X}(\mathcal{C},\mathcal{F})$; we write $\Phi(\mathcal{C},\mathcal{F}) = \mathrm{coker}\,\lambda$, and call it the *component group* (by analogy with Jacobians).
We end this survey with an alternative, more concrete, description of $\widehat{X}(\mathcal{C},\mathcal{F})$. Let

$$r : \widehat{\mathcal{C}\otimes\overline{k}} \longrightarrow \mathcal{C}\otimes\overline{k}$$

denote the normalisation map. Then we define the sheaf $\mathcal{G}$ by

$$0\longrightarrow\mathcal{F}\longrightarrow r_*r^*\mathcal{F}\longrightarrow\mathcal{G}\longrightarrow 0;$$

as in [14], §17, or [18], §1.3, we may also write

$$\widehat{X}(\mathcal{C},\mathcal{F}) = \mathrm{im}(H^0(\mathcal{C}\otimes\overline{k},\mathcal{G}) \longrightarrow H^1(\mathcal{C}\otimes\overline{k},\mathcal{F})).$$

5.2 Integral models of Shimura curves: the split case

Next, we fix a quaternion algebra B over F, and suppose that $\mathfrak{p}$ is a finite prime of F at which B is split. We will suppose that B is split at the infinite place τ_1 and ramified at the other infinite places of F. At each split place v, we fix an isomorphism $B(\mathrm{F}_v)\cong M_2(\mathrm{F}_v)$, and if v is a finite place, we even fix $B(\mathcal{O}_v)\cong M_2(\mathcal{O}_v)$. We regard F as a subfield of $\mathbb{C}$ via τ_1. In the usual way, we let $G=\mathrm{Res}_{\mathrm{F}/\mathbb{Q}}(B^\times)$, and if U is an open compact subgroup of $G(\mathbb{A}^\infty)$, we may form a Shimura curve M_U, defined over F, whose complex points are

$$G(\mathbb{Q})\backslash G(\mathbb{A}^\infty)\times X/U,$$

where $X=\mathfrak{h}^\pm=\mathbb{C}-\mathbb{R}$ is two copies of the upper-half complex plane.

Now we suppose that $U = U_0(\mathfrak{p}) \times H$, as in [14]. Thus U, the level, has a component $U_0(\mathfrak{p}) \subset \mathrm{GL}_2(\mathcal{O}_\mathfrak{p})$ at $\mathfrak{p}$, and a level $H \subset \prod_{v\neq\mathfrak{p}}(B \times \mathrm{F}_v)^\times$ away from $\mathfrak{p}$. Write $\Gamma = \prod_{v\neq\mathfrak{p}}(B \times \mathrm{F}_v)^\times$. Here,

$$U_0(\mathfrak{p}) = \left\{ \left(\begin{array}{cc} a & b \\ c & d \end{array} \right) \in \mathrm{GL}_2(\mathcal{O}_\mathfrak{p}) \middle| c \in \mathfrak{p} \right\},$$

using the above identification. We suppose that H is sufficiently small in the sense of Carayol ([4], 1.4.1.1 and 1.4.1.2). When the level structure at $\mathfrak{p}$ is maximal compact, then Carayol ([4]) proved that the Shimura curve has an integral model $\mathbf{M}_{0,H}$ over $\operatorname{spec}\mathcal{O}_{(\mathfrak{p})}$ with good reduction (i.e., is proper and smooth).

Then in [14], we proved the following:

Theorem 5.1 *1. If H is sufficiently small (as above), then there exists a regular model $\mathbf{M}_{U_0(\mathfrak{p}),H}$ over $\operatorname{spec}\mathcal{O}_{(\mathfrak{p})}$ of M_U.*

2. The special fibre $\mathbf{M}_{U_0(\mathfrak{p}),H} \times \overline{\kappa}_\mathfrak{p}$ looks like two copies of $\mathbf{M}_{0,H} \times \overline{\kappa}_\mathfrak{p}$ intersecting transversely above a finite set of points Σ_H.

The set of points Σ_H are the *supersingular points* of $\mathbf{M}_{0,H} \times \overline{\kappa}_\mathfrak{p}$, and we use the same notation Σ_H for the points which lie above them in $\mathbf{M}_{U_0(\mathfrak{p}),H} \times \overline{\kappa}_\mathfrak{p}$, the singular points of the special fibre. Carayol ([4], §11) describes Σ_H as follows: Let $\overline{B}(\mathfrak{p})$ denote the quaternion algebra got from B by changing the invariants at $\mathfrak{p}$ and at τ_1 (so it is now ramified at both these places, and is totally definite). We write $\overline{B} = \overline{B}(\mathfrak{p})$. Let $\overline{G}$ denote the algebraic group $\mathrm{Res}_{\mathrm{F}/\mathbb{Q}}\overline{B}^\times$, and fix, for all places $v \neq \mathfrak{p}, \tau_1$, an isomorphism between $B \otimes \mathrm{F}_v$ and $\overline{B} \otimes \mathrm{F}_v$. Then $\overline{G}(\mathbb{A}^\infty)$ may be identified with $\Gamma \times \overline{B}_\mathfrak{p}^\times$. By [4], 11.2(3), there is a bijection

$$\begin{aligned} \Sigma_H &\cong \overline{G}(\mathbb{Q})\backslash\overline{G}(\mathbb{A}^\infty)/H \times \mathcal{O}_{\overline{B}_\mathfrak{p}}^\times \\ &\cong \overline{G}(\mathbb{Q})\backslash\Gamma \times \mathrm{F}_\mathfrak{p}^\times/H \times \mathcal{O}_\mathfrak{p}^\times \end{aligned}$$

where the second isomorphism is induced by the reduced norm $\overline{B}_\mathfrak{p}^\times \longrightarrow \mathrm{F}_\mathfrak{p}^\times$.

5.3 Integral models of Shimura curves: the ramified case

Again, we consider a quaternion algebra B over F, and again suppose that B is split at the infinite place τ_1 and ramified at the other infinite places of F. This time, however, we suppose that $\mathfrak{p}$ is a finite prime of F at which B is ramified. Fix isomorphisms at split places in the same way as in the previous subsection. Again, if U is an open compact subgroup of $G(\mathbb{A}^\infty)$, we may form a Shimura curve M_U, defined over F.

Now we suppose that the level structure may be written $U = K_\mathfrak{p} \times H$, where $K_\mathfrak{p} = \mathcal{O}_{B,\mathfrak{p}}^\times$. In this case (and more generally), the integral models were studied by Boutot-Zink ([2]) and by Varshavsky ([20], [21]). Their methods generalise

the case $\mathrm{F} = \mathbb{Q}$ due to Drinfeld and Čerednik respectively. As in the case $\mathrm{F} = \mathbb{Q}$, the results of Boutot and Zink apply for more general level structures. Again, the main result depends on defining another quaternion algebra $\overline{B}(\mathfrak{p})$, whose invariants are the same as B, except with the invariants at $\mathfrak{p}$ and τ_1 changed. So $\overline{B} = \overline{B}(\mathfrak{p})$ is now split at $\mathfrak{p}$, and is totally definite. We define the algebraic group $\overline{G}$ in the usual way, and fix isomorphisms between B and $\overline{B}$ everywhere except at $\mathfrak{p}$ and at τ_1.
Let $\overline{H}$ denote the subgroup of $(\overline{B} \otimes \mathbb{A}_{\mathrm{F}}^{\infty,\mathfrak{p}})^\times$ corresponding to H under the isomorphism

$$(\overline{B} \otimes \mathbb{A}_{\mathrm{F}}^{\infty,\mathfrak{p}})^\times \xrightarrow{\sim} (B \otimes \mathbb{A}_{\mathrm{F}}^{\infty,\mathfrak{p}})^\times.$$

Theorem 5.2 *1. In the above situtíon, the Shimura curve M_U has an integral model $\mathbf{M}_U$ defined over* $\operatorname{spec}\mathcal{O}_{(\mathfrak{p})}$*, and the completion of this model along its special fibre is isomorphic as a formal $\mathcal{O}_{\mathfrak{p}}$-scheme (and the isomorphism is $G(\mathbb{A}^{\infty,\mathfrak{p}})$-equivariant) to*

$$\mathrm{GL}_2(\mathrm{F}_{\mathfrak{p}})\backslash(\mathfrak{h}_{\mathfrak{p}} \times_{\mathrm{Spf}\,\mathcal{O}_{\mathfrak{p}}} \mathrm{Spf}\,\mathcal{O}_{\mathfrak{p}}^{\mathrm{unr}}) \times X_H,$$

where X_H denotes the finite set $\overline{H}\backslash\overline{G}(\mathbb{A}^\infty)/\overline{G}(\mathbb{Q})$ and $\mathfrak{h}_{\mathfrak{p}}$ is Mumford's $\mathfrak{p}$-adic upper half-plane.

2. In particular, the dual graph associated to the special fibre of $\mathbf{M}_U$ is $\mathrm{GL}_2(\mathrm{F}_{\mathfrak{p}})^+\backslash(\Delta \times X_H)$, where Δ denotes the Bruhat-Tits building of $\mathrm{SL}_2(\mathrm{F}_{\mathfrak{p}})$; here, $\mathrm{GL}_2(\mathrm{F}_{\mathfrak{p}})^+$ denotes the set of elements of $\mathrm{GL}_2(\mathrm{F}_{\mathfrak{p}})$ with even $\mathfrak{p}$-adic valuation.

6 Ribet's theorem

As already remarked, the existence of auxiliary primes and Carayol's Lemma proves that one may lower the level for characteristic 2 representations in Cases (2)–(4) of Theorem 2.2 in the same way as [15] or [11] does for odd characteristic representations. To finish the proof of Theorem 0.1, it remains to verify Case (1). For odd characteristic, this is done in [18], except for certain cases where $[\mathrm{F} : \mathbb{Q}]$ is even. In this section, we deal with the case of characteristic 2 representations, also indicating how to deal with all cases when $[\mathrm{F} : \mathbb{Q}]$ is even. This analysis is also valid for odd characteristic, and thus completes the proof of level lowering for primes not dividing the characteristic in this case also.
The proof synthesizes the techniques of Buzzard ([3]) with the work of Rajaei ([18]) to find a version of Ribet's theorem for $\ell = 2$ applicable for totally real fields. Most of the hard work has been done in these two sources, and we refer to them for certain details.
Our target is to prove Theorem 0.1. For simplicity, we shall first describe the case where $[\mathrm{F} : \mathbb{Q}]$ is odd, and will later indicate how to adapt the argument to the even degree case.
We therefore fix a modular mod 2 Galois representation

$$\overline{\rho} : \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_2)$$

which is continuous, irreducible and not induced from a character of $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}(i))$. Part of the hypotheses of the statement is that $\overline{\rho}$ is associated to some modular form of level prime to 2. If D denotes the quaternion algebra over F ramified at exactly all infinite places of F except τ_1, then the Jacquet-Langlands correspondence provides examples of automorphic representations on D of conductor prime to 2 and of some weight whose mod 2 Galois representation are isomorphic to $\overline{\rho}$.

Fix isomorphisms between D and GL_2 at split places in the usual way. If π is such an automorphic representation, with a fixed vector under $U \subset \mathrm{GL}_2(\mathbb{A}_\mathrm{F}^\infty)$ and of weight k, then π corresponds to a maximal ideal m in the Hecke algebra $\mathbb{T} = \mathbb{T}_k^D(U)$ defined in §1. The level structure U gives a Shimura curve M_U as in §5.2, and Carayol ([5], §2) defines a sheaf $\mathcal{F}_k^D$ on M_U corresponding to the weight k.

Definition 6.1 We say that $\overline{\rho}$ satisfies *multiplicity one at weight k* if for all such maximal ideals m coming from automorphic representations of conductor prime to 2 and weight k, we have

$$\dim_{\mathbb{T}/\mathsf{m}}(H^1(M_U \otimes_\mathrm{F} \overline{\mathrm{F}}, \mathcal{F}_k^D) \otimes \mathbb{T}/\mathsf{m}) = 1.$$

Although we expect multiplicity one to hold often (after all, Fujiwara [10] has shown that at least in the ordinary case the minimal Hecke algebra is a complete intersection), Kilford ([16]) has shown that it sometimes fails for $\mathrm{F} = \mathbb{Q}$ when $\ell = 2$.

Having defined the notion of multiplicity one, we now turn to the proof of Theorem 0.1. By Remark 2.4, it suffices to consider Case (1) of Theorem 2.2, i.e., the special unramified case. Thus we suppose that $\overline{\rho}$ is modular of some weight k and some level $U_1(\mathfrak{n}) \cap U_0(\mathfrak{p})$. Here, $\mathfrak{n}$ is coprime to 2, and $\mathfrak{p} \nmid 2\mathfrak{n}$.
We must prove the following result.

Theorem 6.2 *Let $f \in S_k(U_1(\mathfrak{n}) \cap U_0(\mathfrak{p}))$ be a Hilbert cuspidal eigenform, where $(\mathfrak{n}, 2) = 1$ and $\mathfrak{p} \nmid 2\mathfrak{n}$ is a prime ideal. Suppose that the mod 2 representation associated to f,*

$$\overline{\rho} : \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_2),$$

1. *is absolutely irreducible and unramified at $\mathfrak{p}$,*
2. *is not induced from a character of $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}(i))$,*
3. *satisfies multiplicity one at weight k.*

Then there is a Hilbert cuspidal eigenform $g \in S_k(U_1(\mathfrak{n}))$ that gives rise to $\overline{\rho}$.

In order to apply geometric arguments, we first add some auxiliary level structure with the aid of Theorem 4.3. This theorem guarantees the existence of infinitely many primes $\mathfrak{r}_0 \nmid 2\mathfrak{n}\mathfrak{p}$ such that $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{r}_0) > 4^d$ and such that $\mathfrak{r}_0$ is an example of a prime such that Theorem 4.3 holds. We may then add auxiliary level $U_1^1(\mathfrak{r}_0)$-structure, and we showed in [14], §12, that this is sufficiently small so that integral models exist for Shimura curves with this level structure. We may thus use geometric arguments when this auxiliary level is present; however, all modular forms occuring are automatically old at $\mathfrak{r}_0$.

Write U for the level structure $U_1(\mathfrak{n}) \cap U_1^1(\mathfrak{r}_0)$. Thus $f \in S_k(U \cap U_0(\mathfrak{p}))$.

Write $\mathbb{T}_U$ (resp. $\mathbb{T}_{\mathfrak{p},U}$) for the Hecke algebra $\mathbb{T}_k^D(U)$ (resp. $\mathbb{T}_k^D(U \cap U_0(\mathfrak{p}))$). So $\mathbb{T}_U$ is generated by operators $T_\mathfrak{r}$ and $S_\mathfrak{r}$ for primes $\mathfrak{r} \nmid \mathfrak{n}\mathfrak{r}_0$ and operators $U_\mathfrak{r}$ for primes $\mathfrak{r}|\mathfrak{n}\mathfrak{r}_0$, and $\mathbb{T}_{\mathfrak{p},U}$ has the same generators, except that there is an operator $U_\mathfrak{p}$ replacing the pair $T_\mathfrak{p}$ and $S_\mathfrak{p}$. On the $\mathfrak{p}$-old subspace of $S_k(U \cap U_0(\mathfrak{p}))$, the operators $S_\mathfrak{p}$, $T_\mathfrak{p}$ and $U_\mathfrak{p}$ are related by the Eichler-Shimura relation

$$U_\mathfrak{p}^2 - U_\mathfrak{p} T_\mathfrak{p} + N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{p}) S_\mathfrak{p} = 0.$$

As $\overline{\rho}$ is modular of level $U \cap U_0(\mathfrak{p})$ by hypothesis, there is a non-trivial maximal ideal $\widehat{\mathsf{m}}$ of $\mathbb{T}_{\mathfrak{p},U}$ containing 2 and a Hilbert cuspidal eigenform f whose mod 2 eigenvalues are given by the map

$$\mathbb{T}_{\mathfrak{p},U} \longrightarrow \mathbb{T}_{\mathfrak{p},U}/\widehat{\mathsf{m}} \hookrightarrow \overline{\mathbb{F}}_2$$

such that $\overline{\rho}$ is the mod 2 Galois representation associated to f.

Next, we add another auxiliary prime to the level. Its function is rather different to the first.

Let G denote the image of $\overline{\rho}$. As in [3], G must have even order; if it were to have odd order, then it could not have any degree 2 absolutely irreducible representations: its representation theory would be the same as in characteristic 0, and then the degree of any absolutely irreducible representation would divide the order of the group. We may therefore find an involution $\sigma \in G$. By the Čebotarev density theorem, there are infinitely many primes $\mathfrak{q}$ such that $\overline{\rho}(\mathrm{Frob}_\mathfrak{q}) = \sigma$. All involutions in $\mathrm{GL}_2(\overline{\mathbb{F}}_2)$ have trace 0, so we conclude that these Frobenius elements $\mathrm{Frob}_\mathfrak{q}$ are special. We fix such a prime $\mathfrak{q} \nmid \mathfrak{n}\mathfrak{p}$. We will also be considering the Hecke algebra $\mathbb{T}_{\mathfrak{p}\mathfrak{q},U}$ associated to $U \cap U_0(\mathfrak{p}\mathfrak{q})$. We say that an ideal m of $\mathbb{T}_{\mathfrak{p}\mathfrak{q},U}$ is *compatible with* $\widehat{\mathsf{m}}$ if the restrictions of the two maps

$$\begin{array}{rcl}\mathbb{T}_{\mathfrak{p},U} \longrightarrow \mathbb{T}_{\mathfrak{p},U}/\widehat{\mathsf{m}} & \hookrightarrow & \overline{\mathbb{F}}_2 \\ \mathbb{T}_{\mathfrak{p}\mathfrak{q},U} \longrightarrow \mathbb{T}_{\mathfrak{p}\mathfrak{q},U}/\mathsf{m} & \hookrightarrow & \overline{\mathbb{F}}_2\end{array}$$

agree on the intersection of the two Hecke algebras.

We have the following level raising result:

Theorem 6.3 *If* m *is a maximal ideal of* $\mathbb{T}_{\mathfrak{p}\mathfrak{q},U}$ *which is* $\mathfrak{p}$*-new, and compatible with* $\widehat{\mathsf{m}}$*, then* m *is also* $\mathfrak{q}$*-new.*

Proof. This is exactly as in [18], Theorem 5, noting that $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{q})$ is odd and $\mathrm{tr}\,\overline{\rho}(\mathrm{Frob}_\mathfrak{q}) = 0$, so that $T_\mathfrak{q} \in \widehat{\mathsf{m}}$. □

So there is a Hilbert cusp form in $S_k(U \cap U_0(\mathfrak{p}\mathfrak{q}))$, which is $\mathfrak{p}$-new and $\mathfrak{q}$-new and which gives rise to the representation $\overline{\rho}$. If there were a form in

$S_k(U \cap U_0(\mathfrak{q}))$, then we could apply Mazur's Principle of [14] or [11] as in Remark 2.5, as $\overline{\rho}(\mathsf{Frob}_\mathfrak{q})$ is of order 2 in $\mathrm{GL}_2(\overline{\mathbb{F}}_2)$, and is therefore not a scalar. Mazur's Principle now implies that there must be a form in $S_k(U)$ giving $\overline{\rho}$, and so the theorem above will hold.

Thus we assume for a contradiction that there is no Hilbert cusp form in $S_k(U \cap U_0(\mathfrak{q}))$ giving $\overline{\rho}$. Write $\mathbb{T}$ for the Hecke algebra $\mathbb{T}_{\mathfrak{pq},U}$.

We let B denote the quaternion algebra over F ramified at all infinite places of F except τ_1, and also at both $\mathfrak{p}$ and $\mathfrak{q}$, so differs from D only in that the invariants at $\mathfrak{p}$ and $\mathfrak{q}$ have been switched. We fix an isomorphism $B \otimes \mathrm{F}_v \cong D \otimes \mathrm{F}_v$ at all other places v, and also integral versions at all finite $v \neq \mathfrak{p}, \mathfrak{q}$. As we have already fixed isomorphisms between D and GL_2 at finite places, we obtain isomorphisms between B and GL_2 at all finite split places.

As in [18], we let $\mathcal{C}$ denote the Shimura curve associated to the quaternion algebra B with level structure U, and we write $M_{\mathfrak{pq},U}$ (resp. $M_{\mathfrak{p},U}$, $M_{\mathfrak{q},U}$) for the Shimura curve associated to D with level structure $U_0(\mathfrak{pq}) \cap U$ (resp. $U_0(\mathfrak{p}) \cap U$, $U_0(\mathfrak{q}) \cap U$). Rajaei points ([18], §3.1) out that the Hecke algebra $\mathbb{T}$ acts on the cohomology of all of these objects (which is not *a priori* clear for $\mathcal{C}$).

As remarked above, Carayol ([5], §2, §4) defines a sheaf $\mathcal{F}_k^D$ on $M_{\mathfrak{pq},U}$ corresponding to the weight k, and explains how to extend the definition to the integral model. The same construction (see [18], §3.1) gives a sheaf $\mathcal{F}_k^B$ on $\mathcal{C}$.

We make the following abbreviations for objects defined in §5.1:

$$\begin{aligned} X_\mathfrak{p}(\mathfrak{p}) &= X(M_{\mathfrak{p},U} \otimes \overline{\mathrm{F}}_\mathfrak{p}, \mathcal{F}_k^D), \\ X_\mathfrak{p}(\mathfrak{pq}) &= X(M_{\mathfrak{pq},U} \otimes \overline{\mathrm{F}}_\mathfrak{p}, \mathcal{F}_k^D), \\ X_\mathfrak{q}(\mathfrak{p}) &= X(M_{\mathfrak{p},U} \otimes \overline{\mathrm{F}}_\mathfrak{q}, \mathcal{F}_k^D), \\ X_\mathfrak{q}(\mathfrak{pq}) &= X(M_{\mathfrak{pq},U} \otimes \overline{\mathrm{F}}_\mathfrak{q}, \mathcal{F}_k^D), \end{aligned}$$

and similarly for $\widehat{X}$ and the "component group" Φ. We will also use these definitions with $\mathfrak{p}$ and $\mathfrak{q}$ interchanged. As for $\mathcal{C}$, we define:

$$\begin{aligned} Y_\mathfrak{p}(\mathfrak{q}) &= X(\mathcal{C} \otimes \overline{\mathrm{F}}_\mathfrak{p}, \mathcal{F}_k^B), \\ \widehat{Y}_\mathfrak{p}(\mathfrak{q}) &= \widehat{X}(\mathcal{C} \otimes \overline{\mathrm{F}}_\mathfrak{p}, \mathcal{F}_k^B), \\ \Psi_\mathfrak{p}(\mathfrak{q}) &= \Phi(\mathcal{C} \otimes \overline{\mathrm{F}}_\mathfrak{p}, \mathcal{F}_k^B); \end{aligned}$$

again we will use these definitions with $\mathfrak{p}$ and $\mathfrak{q}$ interchanged.

The detailed studies of the dual graphs of the special fibres of $M_{\mathfrak{pq},U}$ mod $\mathfrak{q}$ and $\mathcal{C}$ mod $\mathfrak{p}$ shows that the combinatorics of the two reductions have much in common. In particular, the vertices (resp. edges) of the dual graph of the special fibre of $\mathcal{C}$ mod $\mathfrak{p}$ are in bijection with the singular points of $M_{\mathfrak{q},U}$ mod $\mathfrak{q}$ (resp. of $M_{\mathfrak{pq},U}$ mod $\mathfrak{q}$). Using this, Rajaei proves ([18], (3.15)) that there is an analogue of Ribet's exact sequence in this general weight, general totally real field, case:

$$0 \longrightarrow \widehat{X}_{\mathfrak{q}}(\mathfrak{q})^2_{\mathsf{m}} \longrightarrow \widehat{X}_{\mathfrak{q}}(\mathfrak{pq})_{\mathsf{m}} \longrightarrow \widehat{Y}_{\mathfrak{p}}(\mathfrak{q})_{\mathsf{m}} \longrightarrow 0$$

and similarly with $\mathfrak{p}$ and $\mathfrak{q}$ interchanged.

By the theorem of Boston, Lenstra and Ribet ([1]), the $\mathbb{T}$-module $H^1(M_{\mathfrak{pq},U} \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \mathcal{F}_k^D) \otimes \mathbb{T}/\mathsf{m}$ is a semisimple $\mathbb{T}/\mathsf{m}[\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F})]$-module, isomorphic to $\check{\overline{\rho}}^{\lambda}$ for some $\lambda \geq 1$.

As we are assuming that $\overline{\rho}$ satisfies multiplicity one, we have $\lambda = 1$.

In the same way, $H^1(\mathcal{C} \otimes_{\mathrm{F}} \overline{\mathrm{F}}, \mathcal{F}_k^B) \otimes \mathbb{T}/\mathsf{m}$ is a semisimple $\mathbb{T}/\mathsf{m}[\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F})]$-module, isomorphic to $\check{\overline{\rho}}^{\mu}$ for some integer μ. As m corresponds to a cuspidal eigenform on $U_0(\mathfrak{pq}) \cap U$ which is new at $\mathfrak{p}$ and at $\mathfrak{q}$, the corresponding automorphic representation is special at $\mathfrak{p}$ and $\mathfrak{q}$. Then the Jacquet-Langlands correspondence furnishes a cuspidal automorphic representation on B of level $\mathcal{O}_{B,\mathfrak{p}}^{\times} \times \mathcal{O}_{B,\mathfrak{q}}^{\times} \times U$ whose associated Galois representation is $\overline{\rho}$. It follows that $\mu > 0$.

Our assumptions on $\overline{\rho}$ imply that $X_{\mathfrak{q}}(\mathfrak{q})_{\mathsf{m}} = 0$ and $\widehat{X}_{\mathfrak{q}}(\mathfrak{q})_{\mathsf{m}} = 0$.

Proposition 6.4 *We have:*

1. $\dim_{\mathbb{T}/\mathsf{m}}(\widehat{Y}_{\mathfrak{p}}(\mathfrak{q}) \otimes \mathbb{T}/\mathsf{m}) = 2\mu$,
2. $\dim_{\mathbb{T}/\mathsf{m}}(\widehat{X}_{\mathfrak{q}}(\mathfrak{pq}) \otimes \mathbb{T}/\mathsf{m}) \leq 1$.

Proof.

1. This is proven in the same way as the first claim of [18], Proposition 10. We have the following isomorphisms:

$$\begin{aligned} \widehat{Y}_{\mathfrak{p}}(\mathfrak{q}) \otimes \mathbb{T}/\mathsf{m} &\cong (H^1(\mathcal{C} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \mathcal{F}_k^B) \otimes \mathbb{T}/\mathsf{m})^{I_{\mathfrak{p}}} \\ &\cong H^1(\mathcal{C} \otimes \overline{\mathrm{F}}_{\mathfrak{p}}, \mathcal{F}_k^B) \otimes \mathbb{T}/\mathsf{m} \\ &\cong H^1(\mathcal{C} \otimes \overline{\mathrm{F}}, \mathcal{F}_k^B) \otimes \mathbb{T}/\mathsf{m} \\ &\cong \check{\overline{\rho}}^{\mu} \end{aligned}$$

where the first isomorphism comes from the theory of vanishing cycles (see [18], Lemma 1), and the second occurs as $\overline{\rho}$ is unramified at $\mathfrak{p}$. Now $\overline{\rho}$ is 2-dimensional as a $\mathbb{T}/\mathsf{m}$-vector space, and so the result follows.

2. From the specialisation exact sequence for $M_{\mathfrak{pq},U}$ mod $\mathfrak{q}$, we get the following exact sequence (as in [18]):

$$0 \longrightarrow H^1(\mathbf{M}_{\mathfrak{pq},U} \otimes \overline{\kappa}_{\mathfrak{q}}, \mathcal{F}_k^D)_{\mathsf{m}} \longrightarrow H^1(\mathbf{M}_{\mathfrak{pq},U} \otimes \overline{\mathrm{F}}_{\mathfrak{q}}, \mathcal{F}_k^D)_{\mathsf{m}} \longrightarrow X_{\mathfrak{q}}(\mathfrak{pq})_{\mathsf{m}}(-1) \longrightarrow 0.$$

We see that $X_{\mathfrak{q}}(\mathfrak{pq})(-1) \otimes \mathbb{T}/\mathsf{m}$ is a quotient of $H^1(\mathbf{M}_{\mathfrak{pq},U} \otimes \overline{\mathrm{F}}_{\mathfrak{q}}, \mathcal{F}_k^D) \otimes \mathbb{T}/\mathsf{m}$. However, this latter space is precisely $\overline{\rho}$ (at least restricted to a decomposition group at $\mathfrak{q}$), using the multiplicity one hypothesis. We know from Carayol's Theorem ([5], Théorème (A)) that $\mathsf{Frob}_{\mathfrak{q}}$ acts on $X_{\mathfrak{q}}(\mathfrak{pq}) \otimes \mathbb{T}/\mathsf{m}$ by a scalar. However, $\mathfrak{q}$ was chosen so that $\overline{\rho}(\mathsf{Frob}_{\mathfrak{q}})$ is an involution on a 2-dimensional space. It follows that any quotient space on which $\mathsf{Frob}_{\mathfrak{q}}$ acts as a scalar must be at most 1-dimensional.

Finally, as the integral model $\mathbf{M}_{\mathfrak{pq},U}$ is regular (as U is sufficiently small), the theory of vanishing cycles implies that $X_{\mathfrak{q}}(\mathfrak{pq})$ is isomorphic (as $\mathbb{T}$-modules, though not as Galois modules) to $\widehat{X}_{\mathfrak{q}}(\mathfrak{pq})$. They therefore have the same dimension, and the result follows.

□

Finally, however, we consider Ribet's exact sequence:

$$0 \longrightarrow \widehat{X}_{\mathfrak{q}}(\mathfrak{q})^2_{\mathsf{m}} \longrightarrow \widehat{X}_{\mathfrak{q}}(\mathfrak{pq})_{\mathsf{m}} \longrightarrow \widehat{Y}_{\mathfrak{p}}(\mathfrak{q})_{\mathsf{m}} \longrightarrow 0.$$

Our assumptions on $\overline{\rho}$ imply that $\widehat{X}_{\mathfrak{q}}(\mathfrak{q})_{\mathsf{m}} = 0$. It follows that the remaining two terms are isomorphic, but we have just shown that the first has dimension ≤ 1, while the second has dimension 2μ. As μ is a strictly positive integer, this is a contradiction. This contradiction establishes the desired result.

In the even degree case, we will construct yet another auxiliary prime $\mathfrak{r}_1$ such that $\overline{\rho}(\mathsf{Frob}_{\mathfrak{r}_1})$ is an involution as above. This implies that $T_{\mathfrak{r}_1} \in \mathsf{m}$, and we may then use Taylor's level raising result ([19], Theorem 1) to add $\mathfrak{r}_1$ to the level. We then use exactly the same argument as above, except where all of the quaternion algebras involved are also ramified at $\mathfrak{r}_1$, and where all Hecke algebras contain the operator $U_{\mathfrak{r}_1}$ rather than $S_{\mathfrak{r}_1}$ and $T_{\mathfrak{r}_1}$. At the end of the argument we remove the prime $\mathfrak{r}_1$ from the level using Fujiwara's version of Mazur's Principle ([11], §5) for the even degree case.

In fact, this approach also works when ℓ is an odd prime. One adds an auxiliary prime to the level using Taylor's result, lowers the level using Rajaei's result (in which there is no multiplicity one hypothesis), and removes the auxiliary prime using Fujiwara's result. This therefore completes level lowering away from the characteristic for all odd primes, and completes the proof of Theorem 0.1 and Theorem 0.2. We stress that in this case, all necessary results are already due to Fujiwara and Rajaei, and the only new results in this paper concern the case $\ell = 2$.

An alternative to this method might be to compare the Shimura curve of level $U_0(\mathfrak{p}) \times H$ in characteristic $\mathfrak{q}$ for the quaternion algebra ramified at all but one infinite places and at $\mathfrak{p}$ with the Shimura curve of level $U_0(\mathfrak{q}) \times H$ in characteristic $\mathfrak{p}$ for the quaternion algebra ramified at all but one infinite places and at $\mathfrak{q}$. One might hope to derive a version of Ribet's theorem without introducing auxiliary primes, which would be rather cleaner. However, the theory of level raising already exists in the even degree case, and so we make use of it freely.

References

[1] N.Boston, H.Lenstra, K.Ribet, Quotients of group rings arising from two-dimensional representations, C. R. Acad. Sci. Paris 312 (1991) 323–328

[2] J.-F.Boutot, T.Zink, The p-adic uniformisation of Shimura curves, preprint

[3] K.Buzzard, On level-lowering for mod 2 representations, Math. Res. Lett. 7 (2000) 95–110

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

[5] H.Carayol, Sur les représentations ℓ-adiques associées aux formes modulaires de Hilbert, Ann. Sci. Ec. Norm. Sup. 19 (1986) 409–468

[6] H.Carayol, Sur les représentations galoisiennes modulo ℓ attachées aux formes modulaires, Duke Math. J. 59 (1989) 785–801

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

[8] F.Diamond, R.Taylor, Non-optimal levels of mod ℓ modular representations, Invent. Math. 115 (1994) 435–462

[9] F.Diamond, R.Taylor, Lifting modular mod ℓ representations, Duke Math. J. 74 (1994) 253–269

[10] K.Fujiwara, Deformation rings and Hecke algebras in the totally real case, preprint (1996)

[11] K.Fujiwara, Level optimisation in the totally real case, preprint (1999, revised 2004)

[12] H.Hida, On p-adic Hecke algebras for GL_2 over totally real fields, Ann. Math. 128 (1988) 295–384

[13] L.Illusie, Réalisation ℓ-adique de l'accouplement de monodromie, d'après A.Grothendieck, Astérisque 196–97 (1991) 27–44

[14] F.Jarvis, Mazur's Principle for totally real fields of odd degree, Comp. Math. 116 (1999) 39–79

[15] F.Jarvis, Level lowering for modular mod ℓ representations over totally real fields, Math. Ann. 313 (1999) 141–160

[16] L.Kilford, Some examples of non-Gorenstein Hecke algebras associated to modular forms, J. Number Theory 97 (2002) 157–164

[17] R.Livné, On the conductors of modulo ℓ representations coming from modular forms, Journal of Number Theory 31 (1989) 133–141

[18] A.Rajaei, On the levels of mod ℓ-Hilbert modular forms, J. reine angew. Math. 537 (2001) 33–65

[19] R.Taylor, On Galois representations associated to Hilbert modular forms, Invent. Math. 98 (1989) 265–280

[20] Y.Varshavsky, p-adic uniformisation of unitary Shimura varieties I, Publ. Math. IHES 87 (1998) 57–119

[21] Y.Varshavsky, p-adic uniformisation of unitary Shimura varieties II, J. Diff. Geom. 49 (1998) 75–113

Frazer Jarvis
Department of Pure Mathematics
Hicks Building
University of Sheffield
Sheffield S3 7RH
Great Britain
a.f.jarvis@sheffield.ac.uk