# Level lowering for modular mod $\ell$ representations over totally real fields*

Frazer Jarvis
School of Mathematics and Statistics
University of Sheffield
Hicks Building
Sheffield S3 7RH
U.K.
jarvisa@maths.ox.ac.uk

**Abstract**

In this paper, we continue the study of part of the analogue of Serre's conjecture for mod $\ell$ Galois representations for totally real fields. More precisely, one knows, through results of Carayol and Taylor, that to any Hilbert cuspidal eigenform over a totally real field F, one can attach a compatible system of $\lambda$-adic representations of the corresponding absolute Galois group. One may ask if a given $\lambda$-adic or modulo $\ell$ representation is attached by this process to a Hilbert modular form, and, if so, what weights and levels this form can have. We prove some analogues of results known in the case $\mathrm{F} = \mathbb{Q}$.

## Introduction

Fix a totally real field F and an odd rational prime $\ell$. We will be considering continuous semisimple representations

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

As all of the results of this paper are already known for $\mathrm{F} = \mathbb{Q}$, we will assume $\mathrm{F} \neq \mathbb{Q}$. To certain Hilbert modular forms which are eigenforms for

*AMS subject classification: 11F33, 11F41, 11F80, 11G18 the action of a Hecke algebra, it is known by work of Carayol and Taylor that one may associate a compatible family of $\lambda$-adic representations of $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F})$, as $\lambda$ runs over primes of the number field generated by the Hecke eigenvalues. See Theorem 1.1 below for a more precise statement. Reducing these modulo those $\lambda$ of residue characteristic $\ell$, and taking the semisimplifications, give examples of representations $\overline{\rho}$ of the above form. Such $\overline{\rho}$ are said to be *modular*. If a particular Hilbert modular form gives rise to $\overline{\rho}$, we may say (abusively) in this situation that $\overline{\rho}$ has weight and level equal to that of this modular form, even though, in general, there will be many different Hilbert modular forms giving rise to a given modular representation. Thus a modular representation will have many different weights and levels. This should cause no confusion in the remainder of the paper.

The aim of this paper is to prove an analogue of some results of Carayol [4] on the possible levels of Hilbert modular forms associated to a given modular modulo $\ell$ representation of the absolute Galois group of a totally real field. In conjunction with our paper [12], this enables us to prove the following unconditional result, in which our notation for Hilbert modular forms (explained more precisely below) follows Hida [11]:

**Theorem 0.1** *Let $\overline{\rho} : \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ be an irreducible modular Galois representation associated to a Hilbert modular form $f$ of arithmetic weight $k$ and level $U_1(\mathfrak{n}\mathfrak{p}^r)$, where $r > v_\mathfrak{p}(a(\overline{\rho}))$, the $\mathfrak{p}$-adic valuation of the Artin conductor of $\overline{\rho}$. (Recall that $\ell$ is odd.) Write $\pi$ for the automorphic representation of $\mathrm{GL}_{2/\mathrm{F}}$ associated to $f$. Suppose that the following supplementary hypotheses hold:*

- $\mathfrak{p} \nmid \ell$
- *if $\pi_\mathfrak{p}$ is principal series, then $r = 1$*
- *if $\pi_\mathfrak{p}$ is special, then $r > 1$ or $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{p}) \not\equiv 1 \pmod{\ell}$*
- *if $[\mathrm{F}(\mu_\ell) : \mathrm{F}] = 2$ then $\overline{\rho}$ is not induced from a character of the kernel of the reduction of the cyclotomic character*
- *if $[\mathrm{F} : \mathbb{Q}]$ is even, suppose that there exists a prime $\mathfrak{q} \neq \mathfrak{p}$ such that $\pi_\mathfrak{q}$ is special or supercuspidal*

*Then $\overline{\rho}$ is modular of weight $k$ on $U_1(\mathfrak{n}\mathfrak{p}^{r-1})$.*

The condition that $k$ be arithmetic, defined in §1 of this paper, implies that the Hecke eigenvalues of an eigenform generate a number field. Throughout the paper, all Hilbert modular forms will be assumed to have arithmetic weight. Throughout the paper, the hypotheses above on F and $\ell$ will be assumed to hold.

**Remark 0.2** One might expect stronger versions of these results to have implications for the solubility of certain Diophantine equations over totally real fields. While this is probably true, we recall that Ribet's proof that the Shimura-Taniyama-Weil conjecture implies Fermat's Last Theorem relies on the fact that $S_2(\Gamma_0(2))$ is trivial. The analogous statement for totally real fields is rarely true; even for real quadratic fields of the form $\mathbb{Q}(\sqrt{p})$ with $p$ prime, the space of Hilbert cusp forms of weight $(2,2)$ on the full Hilbert modular group is trivial only for $p = 2$, 3, 5, 7, 13 and 17 (from calculations in Freitag [9]). The Fermat cubic $X^3 + Y^3 = Z^3$ has non-trivial solutions over many totally real quadratic fields; it is an elliptic curve and one rewrites it as $u^2 = f(v)$ where $f$ is a cubic with rational coefficients. Then there are non-trivial solutions to the Fermat cubic over the totally real field $\mathbb{Q}(\sqrt{f(v)})$ for values of $v$ with $f(v) > 0$.

In outline this paper is similar to that of [4]. We begin by giving a result classifying all cases where $r$ can be strictly greater than $v_{\mathfrak{p}}(a(\overline{\rho}))$. We then prove an analogue of Carayol's Lemma, and the main result of this paper is similar to one in [7]. Indeed, it will be clear that our methods are direct generalisations of those of Carayol ([4]) and Diamond and Taylor ([6], [7]). The final part of the paper is devoted to explaining how to deduce the above result from the main theorem, together with the main result of [12].

We thank Kazuhiro Fujiwara for helpful remarks on a previous version of this paper; he has also independently obtained very similar results, and indeed stronger results in the case of weight 2. Many thanks also to the anonymous referee for making many useful comments.

# 1 Preliminaries

Let F be a totally real number field of degree $d$ over $\mathbb{Q}$. Let $I = \{\tau_1, \dots, \tau_d\}$ be the set of embeddings $\mathrm{F} \hookrightarrow \mathbb{R}$.

We first fix our notation for Hilbert modular forms, which is based closely on that of Hida ([11]). Precise definitions of these objects are to be found in [11], and this is merely intended to be a brief summary of notation.

The weight of a Hilbert modular form is a vector $k \in \mathbb{Z}^I$, so that there is one component for every infinite place of F. We say that $k$ is *arithmetic* if every $k_\tau \geq 2$ and all are of the same parity. Throughout this paper, all Hilbert modular forms will be taken to have arithmetic weight. Write $t = (1, \dots, 1) \in \mathbb{Z}^I$. Then define $v \in \mathbb{Z}^I_{\geq 0}$ so that $k + 2v$ is a multiple of $t$ and some $v_\tau = 0$. The transformation law for Hilbert modular forms is normalised by the choice of a vector $w$ of the form $k + v - \alpha.t$ for some integer $\alpha$. Following Hida, we fix the choice $\alpha = 1$ throughout. If $G$ denotes the algebraic group $\mathrm{Res}_{\mathrm{F}/\mathbb{Q}}\mathrm{GL}_2$, then the level of a Hilbert modular form is an open compact subgroup $U$ of $G(\mathbb{A}^\infty)$. The (finite-dimensional complex) vector space of Hilbert cusp forms of weight $k$ and level $U$ is denoted $S_{k,w}(U)$ (see [11], (2.3), for the precise definition of this space, where it is denoted $S_{k,w,I}(U;\mathbb{C})$). Later we will also consider Hilbert modular forms on more general quaternion algebras.

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 in $\mathrm{GL}_2(\mathrm{F}_{\mathfrak{p}})$. Of most interest will be the groups $U_1(\mathfrak{n})$ introduced in [11], §2. If $S$ is any finite set of finite places of F, we define the Hecke algebra $\mathbb{T}^S$ as the $\mathbb{Z}$-algebra generated by the Hecke operators $T_{\mathfrak{q}}$ and $S_{\mathfrak{q}}$ for $\mathfrak{q} \not\in S$. The Hecke algebra $\mathbb{T}^S$ acts on the space $S_{k,w}(U)$ through its quotient $\mathbb{T}_{k,w}(U)$ (this is our definition of this latter algebra), where $S$ is chosen so as to consist of precisely those finite places where $U_{\mathfrak{p}}$ is not maximal compact in $\mathrm{GL}_2(\mathrm{F}_{\mathfrak{p}})$. The condition that $k$ be arithmetic implies that the field generated by the eigenvalues of the Hecke operators on a given eigenform is a number field.

Let $A_{k,w}(U)$ denote the set of all $f \in S_{k,w}(U)$ (up to scalars) which are eigenfunctions for all of the operators in the Hecke algebra $\mathbb{T}_{k,w}(U)$. Thus, elements $f \in A_{k,w}(U)$ correspond to automorphic representations $\pi(f)$ on $\mathrm{GL}_{2/\mathrm{F}}$ of weight $k$ with a fixed vector under $U$.

Similarly, $A^0_{k,w}(U_1(\mathfrak{n}))$ will denote those elements of $A_{k,w}(U_1(\mathfrak{n}))$ which are *new*, in the usual sense of not arising at any lower level $\mathfrak{n}'$. These correspond to automorphic representations of conductor $\mathfrak{n}$. If $\mu$ is a character of $(\mathcal{O}_{\mathrm{F}}/\mathfrak{n})^\times$, we can make similar definitions for the spaces $A_{k,w}(U_0(\mathfrak{n}),\mu)$ and $A^0_{k,w}(U_0(\mathfrak{n}),\mu)$.

We recall the following theorem:

**Theorem 1.1 (Carayol–Taylor)** *Let $f \in A_{k,w}(U_1(\mathfrak{n}))$, with $k$ arithmetic. Let $\mathcal{O}_f$ be the ring of integers of a number field such that there exists a morphism*

$$\theta_f : \mathbb{T}_{k,w}(U_1(\mathfrak{n})) \longrightarrow \mathcal{O}_f$$

*with $f|T = \theta_f(T)f$. Then if $\lambda$ is a prime of $\mathcal{O}_f$, there is a continuous representation*

$$\rho_{f,\lambda} : \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\mathcal{O}_{f,\lambda})$$

*such that, for all primes $\mathfrak{p}$ of F with different residue characteristic to that of $\lambda$, one has $\rho_{f,\lambda}|_{D_{\mathfrak{p}}} \sim \sigma^\lambda(\pi(f)_{\mathfrak{p}})$, the $\lambda$-adic representation associated (by the local Langlands correspondence) to the local component at $\mathfrak{p}$ of $\pi(f)$.*

**Definition 1.2** Given an irreducible modulo $\ell$ 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* $\mathfrak{n}$ *and weight* $(k, w)$ if there exists $f \in A_{k,w}(U_1(\mathfrak{n}))$ and a prime $\lambda | \ell$ of $\mathcal{O}_f$ such that $\overline{\rho}$ coincides with $\rho_{f,\lambda}$ mod $\lambda$.

**Remark 1.3** Note that this is a slightly non-standard notion of level—we do not insist that $f$ be new of level $\mathfrak{n}$. If this holds, i.e., if $f \in A^0_{k,w}(U_1(\mathfrak{n}))$, we will say $f$ or $\overline{\rho}$ has *new level* $\mathfrak{n}$.

We will always denote the reduction of a representation by adding a bar—this will also apply to characters.

In general, a modular $\overline{\rho}$ may have many possible weights and new levels (see [1], [6], [7], [8], and [16] for a full discussion of the situation when $\mathrm{F} = \mathbb{Q}$).

However, following Serre, we will define an *optimal* level (Serre [17] also predicts an optimal weight and character when $\mathrm{F} = \mathbb{Q}$). Define

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

Then, in the same way as [17], we may ask:

**Question 1** *If $\overline{\rho}$ is modular, then is it modular of level $\mathfrak{n}(\overline{\rho})$?*

The terminology "optimal" is justified by the following result.

**Theorem 1.4 (Carayol-Livné)** *Suppose $\overline{\rho}$ is modular of level $\mathfrak{n}$. Then $\mathfrak{n}(\overline{\rho}) | \mathfrak{n}$.*

This was proven by Carayol [4] and Livné [14]; Carayol only considers the case $\mathrm{F} = \mathbb{Q}$, but the general case is proven in exactly the same way.

It follows that, if the question above is true, then $\overline{\rho}$ is modular of *new* level $\mathfrak{n}(\overline{\rho})$.

To answer the question (at least when $\mathrm{F} = \mathbb{Q}$), one "lowers the level" by one prime at a time. Thus, if $\overline{\rho}$ is modular of level $\mathfrak{n}$, one chooses a prime $\mathfrak{p}$ such that $\mathfrak{p} | (\mathfrak{n}/\mathfrak{n}(\overline{\rho}))$, and tries to show that $\overline{\rho}$ is also modular of level $\mathfrak{n}/\mathfrak{p}$. This is the procedure adopted in [4], [5] and [16]. More precisely,

**Conjecture 1** *Let $\overline{\rho} : \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}) \longrightarrow \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ be a modular Galois representation of weight $k$ and level $U_1(\mathfrak{n}\mathfrak{p}^r)$, where $r > v_{\mathfrak{p}}(a(\overline{\rho}))$. Then $\overline{\rho}$ is modular of weight $k$ on $U_1(\mathfrak{n}\mathfrak{p}^{r-1})$.*

Both Carayol and Livné also classify the possible situations in which the inequality $v_{\mathfrak{p}}(\mathfrak{n}(\overline{\rho})) \leq v_{\mathfrak{p}}(\mathfrak{n})$ can be strict, at least when $\mathrm{F} = \mathbb{Q}$. One readily checks that this analysis is also valid for any totally real field, using [3].

Suppose $\pi$ is an automorphic representation of $\mathrm{GL}_{2/\mathrm{F}}$ of conductor $\mathfrak{n}$ giving rise to the representation $\overline{\rho}$. In other words, $\pi$ corresponds to a Hilbert modular form $f \in A^0_{k,w}(U_1(\mathfrak{n}))$ such that $\overline{\rho} = \overline{\rho}_{f,\lambda}$.

Let $\overline{n}_{\mathfrak{p}} = v_{\mathfrak{p}}(\mathfrak{n}(\overline{\rho}))$, and let $n_{\mathfrak{p}} = v_{\mathfrak{p}}(\mathfrak{n})$.

Then

**Theorem 1.5** *Suppose $\mathfrak{p} \nmid \ell$. Then $\overline{n}_{\mathfrak{p}} = n_{\mathfrak{p}}$ except in the following cases:*

1. *$\pi_{\mathfrak{p}}$ is a special representation, associated to a character $\chi$ which is unramified. Then it may happen that*

$$n_{\mathfrak{p}} = 1, \text{ and } \overline{n}_{\mathfrak{p}} = 0.$$

2. *$\pi_{\mathfrak{p}}$ is a special representation associated to a character $\chi$ which degenerates, i.e., $a(\chi) = 1$ but $a(\overline{\chi}) = 0$. (Here, for any character $\chi$, $a(\chi)$ denotes the $\mathfrak{p}$-adic valuation of its conductor.) For such a character to degenerate, we require that $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{p}) \equiv 1 \pmod{\ell}$. Here $n_{\mathfrak{p}} = 2$.*

3. *$\pi_{\mathfrak{p}}$ is a principal series representation, associated to two characters $\chi$ and $\psi$, with at least one of the characters, $\chi$ say, degenerating. Then $n_{\mathfrak{p}} = 1 + a(\psi)$. Again, for this to occur, it is necessary that $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{p}) \equiv 1 \pmod{\ell}$.*

4. *$\pi_{\mathfrak{p}}$ is a cuspidal Weil representation, $W(\Omega, \xi)$, where $\Omega$ is the unramified quadratic extension of $\mathrm{F}_{\underline{\mathfrak{p}}}$, and $\xi$ is a character of $\Omega^\times$ which degenerates (so that $a(\xi) = 1$ and $a(\overline{\xi}) = 0$). Then $n_{\mathfrak{p}} = 2$. For such a character to degenerate, we require $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{p}) \equiv -1 \pmod{\ell}$.*

We make the following conjecture, which is merely a reformulation of Conjecture 1 in the unramified special case (Case (1) of Theorem 1.5).

**Conjecture 2** *Suppose $\ell$ is as above. Suppose $f \in A^0_{k,w}(U_0(\mathfrak{n}\mathfrak{p}), \mu)$, where $\mathfrak{p} \nmid \mathfrak{n}\ell$, $\mu$ factors through $(\mathcal{O}_{\mathrm{F}}/\mathfrak{n})^\times$, and that $v_{\mathfrak{p}}(\mathfrak{n}(\overline{\rho})) = 0$. Then there exists $f' \in A_{k,w}(U_0(\mathfrak{n}), \mu')$ for some character $\mu'$ of $(\mathcal{O}_{\mathrm{F}}/\mathfrak{n})^\times$ giving the same mod $\ell$ representation.*

In the case $\mathrm{F} = \mathbb{Q}$, the main theorem of [4] is that Conjecture 2 (together with a level raising conjecture) implies Conjecture 1.

**Remark 1.6** In this paper we develop methods analogous to those in the case $\mathrm{F} = \mathbb{Q}$ to lower the level in some of the cases of Theorem 1.5. More precisely, we can completely lower the level in cases (2) and (4), and in some examples of case (3), at least under the hypotheses on F and $\ell$ of Theorem 0.1. The main theorem of [12] proves Conjecture 2, which is no more than a reformulation of Case (1), when $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{p}) \not\equiv 1 \pmod{\ell}$. Thus to complete the answer to Question 1, and lower the level to the optimal level, it remains still to deal with the following cases:

- Case (1), where $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{p}) \equiv 1 \pmod{\ell}$, and
- Case (3), where both defining characters are ramified.

The dictionary between automorphic representations and Galois representations (as in the Carayol-Taylor theorem above) allow us to determine explicitly the form of the restrictions to $D_{\mathfrak{p}}$ of $\overline{\rho}$ and $\rho_{f,\lambda}$ in cases of degeneracy. Once again, this analysis is exactly the same as in the case $\mathrm{F} = \mathbb{Q}$ ([4], Proposition 2), and we will omit the details.

Many of the proofs in this paper are closely modelled on those of [6] and [7].

# 2 Modular forms over quaternion algebras

In this section, we introduce the objects which we will study.

Let $B_{/\mathrm{F}}$ be a quaternion algebra, and let $S(B)$ denote the set of all ramified places of $B$. Denote by $\nu$ the reduced norm on $B$. Fix a maximal order $\mathcal{O}_B$ of $B$, and let $G$ be the algebraic group $\mathrm{Res}_{\mathrm{F}/\mathbb{Q}} B^\times$, and let $Z$ denote its centre. Fix an isomorphism $\mathcal{O}_B \otimes_{\mathcal{O}_\mathrm{F}} \mathcal{O}_{\mathrm{F}_v} \cong \mathrm{M}_2(\mathcal{O}_{\mathrm{F}_v})$ at all finite places of F at which $B$ is split. Write $\mathbb{A}^\infty$ for the *finite* adeles of $\mathbb{Q}$, and write $G(\mathbb{A}^\infty)$ (resp. $G(\widehat{\mathbb{Z}})$) for the $\mathbb{A}^\infty$-valued points of $G$ (resp. $\prod_{\text{finite } v} (\mathcal{O}_B \otimes_{\mathcal{O}_\mathrm{F}} \mathcal{O}_{\mathrm{F}_v})^\times$). Fix a finite Galois extension $\mathrm{E}/\mathbb{Q}$ contained in $\mathbb{C}$ and containing F which splits $B$, and choose an isomorphism

$$B \otimes_{\mathbb{Q}} \mathrm{E} \cong M_2(\mathrm{E})^I$$

satisfying [11], (1.1). Fix also a prime $\lambda$ of E above $\ell$.

Fix $k \in \mathbb{Z}^I_{\geq 2}$, where all of the $k_\tau$ have the same parity, and let $v$ be as in §1. Write $k_0 = \max\{k_\tau\}$. As in [11], we will investigate modular forms of weight $(k, w)$ on $B$.

Write

$$S^B_{k,w} = \bigoplus_{\pi} \pi^\infty,$$

where $\pi$ runs over the set of cuspidal automorphic representations of weight $(k, w)$ (we include those of dimension 1 for the moment). In other words, we require that

$$\pi_\tau = \begin{cases} D_{k_\tau, k_0-2}, & \text{if } \tau \in I - S(B), \\ \overline{D}_{k_\tau, k_0-2}, & \text{if } \tau \in I \cap S(B), \end{cases}$$

where the representations $D_{k_\tau,k_0-2}$ and $\overline{D}_{k_\tau,k_0-2}$ are defined in [3], §0.

The space $S^B_{k,w}$ is an admissible $G(\mathbb{A}^\infty)$-module, where the action is via right translation. If $U$ is an open compact subgroup of $G(\mathbb{A}^\infty)$, write $S_{k,w,B}(U)$ for $(S^B_{k,w})^U$. Then $g \in G(\mathbb{A}^\infty)$ maps $S_{k,w,B}(U)$ to $S_{k,w,B}(gUg^{-1})$. Finally, we need the notion of modular form with representation. Suppose that $G(\mathbb{A}^\infty) \supset U \supset U'$, with $U'$ normal in $U$, and let

$$r : U/U'(U \cap Z(\mathbb{Q})) \longrightarrow \mathrm{Aut}(V)$$

be an irreducible complex representation. Then we write

$$S_{k,w,B}(U, r) = \mathrm{Hom}_U(V, S_{k,w,B}(U')).$$

We now define some $\lambda$-adic objects.

One has an isomorphism

$$G(\mathbb{R}) \cong \mathrm{GL}_2(\mathbb{R})^{I-S(B)} \times (\mathbb{H}^\times)^{I \cap S(B)},$$

where $\mathbb{H}$ denotes the Hamiltonian quaternions. Let $X = (\mathfrak{h}^\pm)^{I-S(B)}$, which has a natural action of $G(\mathbb{R})$. Let $C_\infty$ denote the elements $g \in G(\mathbb{R})$ such that $g$ fixes $(\sqrt{-1}, \ldots, \sqrt{-1})$. Then one can identify $X$ with the quotient $G(\mathbb{R})/C_\infty$. For any open compact subgroup $U$ of $G(\mathbb{A}^\infty)$, one forms the set of points

$$M_U(\mathbb{C}) = G(\mathbb{Q})\backslash G(\mathbb{A}^\infty) \times X/U.$$

The cases which will interest us are those in which either 0 or 1 infinite places of F are split in $B$. Then $M_U(\mathbb{C})$ has dimension 0 and 1 respectively.

When there is 1 infinite place split in $B$, $M_U(\mathbb{C})$ are the complex points of a smooth proper Shimura curve $M_U$, defined over a subfield of $\mathbb{C}$ which we may identify with F. More generally, if there are $r$ split infinite places of $B$, and $r > 0$, then $M_U(\mathbb{C})$ are the complex points of a Shimura variety $M_U$ which is smooth and proper over (a subfield of) the Galois closure of F if $B$ is not split.

We follow Carayol [3], §2.1, in defining a sheaf on $M_U(\mathbb{C})$.

For each $i = 1, \ldots, d$, fix an isomorphism $B \otimes_{\mathrm{F},\tau_i} \mathrm{E} \cong M_2(\mathrm{E})$ such that $\mathcal{O}_B \otimes_{\mathcal{O}_\mathrm{F},\tau_i} \mathcal{O}_\mathrm{E} \cong M_2(\mathcal{O}_\mathrm{E})$, and this defines an equivalence class, written $\xi_i$, of representations of $B^\times = G(\mathbb{Q})$ on $W_i = \mathrm{E}^2$. Consider

$$\xi = \bigotimes_{i=1}^{d} [(\tau_i \circ \nu)^{v_i} \mathrm{Sym}^{k_i-2}(\xi_i)]$$

of $G(\mathbb{Q})$ acting (on the left) on the space $W = \bigotimes_{i=1}^{d} \mathrm{Sym}^{k_i-2} W_i$. We also denote this space by $L_{k,w,B}(\mathrm{E})$. This action naturally extends to a unique action of the algebraic group $G$ on the $\mathbb{Q}$-space underlying $W$. In fact, this even extends to an integral model; if we choose the lattice $L_i = \mathcal{O}_{\mathrm{E}}^2 \subset W_i$ which is stable under $\mathcal{O}_B$, the same procedure gives an action $\xi$ of $\mathcal{O}_B$ on $L = L_{k,w,B}(\mathcal{O}_{\mathrm{E}}) = \bigotimes_{i=1}^{d} \mathrm{Sym}^{k_i-2} L_i$.

Write $W_\lambda$ for $W \otimes_{\mathrm{E}} \mathrm{E}_\lambda$.

We first define a (total space of a) locally constant sheaf on the complex variety $M_U(\mathbb{C})$ by

$$\mathcal{F}^U_{k,w,B}(\mathbb{Q}_\ell)_{\mathbb{C}} = G(\mathbb{Q})\backslash G(\mathbb{A}^\infty) \times X \times W_\lambda / U,$$

with $G(\mathbb{Q})$ acting on the left on $W_\lambda$ via $G(\mathbb{Q}_\ell)$, and $U$ acting trivially. In particular, a section $f$ of this sheaf is a locally constant map

$$f : G(\mathbb{A}) \longrightarrow W_\lambda$$

such that

$$\begin{array}{lrcl} \text{for } \gamma \in G(\mathbb{Q}), & f(\gamma g) & = & \xi(\gamma_\ell) f(g), \\ \text{for } u \in UC_\infty, & f(gu) & = & f(g). \end{array}$$

Note that the function

$$F_f(g) = \xi(g_\ell)^{-1} f(g)$$

satisfies

$$\begin{array}{lrcl} \text{for } \gamma \in G(\mathbb{Q}), & F_f(\gamma g) & = & F_f(g), \\ \text{for } u \in UC_\infty, & F_f(gu) & = & \xi(u_\ell)^{-1} F_f(g). \end{array}$$

In the case where the complex analytic variety $M_U(\mathbb{C})$ has positive dimension, we may interpret this sheaf as a locally constant étale sheaf on $M_U$ as follows, providing $U$ is sufficiently small ([2], 1.4.1.1, or [12], §12). On $M(\mathbb{C}) = \varprojlim M_K(\mathbb{C})$, the projective limit as $K$ runs through open compact subgroups of $G(\mathbb{A}^\infty)$, this sheaf is constant, with fibre $W_\lambda$. Thus we see that

$$\mathcal{F}^U_{k,w,B}(\mathbb{Q}_\ell)_{\mathbb{C}} = (M(\mathbb{C}) \times W_\lambda)/(U/(U \cap \widehat{Z(\mathbb{Q})})),$$

as $U/(U \cap \widehat{Z(\mathbb{Q})})$ is the group of the covering $M(\mathbb{C}) \longrightarrow M_U(\mathbb{C})$ when $U$ is sufficiently small, i.e., the quotient $M(\mathbb{C})/(U/(U \cap \widehat{Z(\mathbb{Q})}))$ may be identified with $M_U(\mathbb{C})$. Here $\widehat{Z(\mathbb{Q})}$ is the closure of $Z(\mathbb{Q})$ in $Z(\mathbb{A}^\infty)$, and $u \in U$ acts on the right on $W_\lambda$ by $\xi(u_\ell)^{-1}$.

Choose a lattice $L_\lambda$ inside $W_\lambda$ stabilised by $U$, and pick a normal subgroup $U' \subset U$ such that $U'$ acts trivially on $L_\lambda/\ell^n L_\lambda$.

Then we may define a sheaf $\mathcal{F}_{k,w,B}(\mathbb{Z}_\ell)_{\mathbb{C}}$ by

$$\mathcal{F}^U_{k,w,B}(\mathbb{Z}_\ell)_{\mathbb{C}} = (M(\mathbb{C}) \times L_\lambda) \Big/ \Big(U/(U \cap \widehat{Z(\mathbb{Q})})\Big),$$

and

$$\mathcal{F}^U_{k,w,B}(\mathbb{Z}/\ell^n\mathbb{Z})_{\mathbb{C}} = \Big(M_{U'}(\mathbb{C}) \times (L_\lambda/\ell^n L_\lambda)\Big) \Big/ (U/U').$$

Carayol then defines an étale sheaf of $(\mathbb{Z}/\ell^n\mathbb{Z})$-modules $\mathcal{F}^U_{k,w,B}(\mathbb{Z}/\ell^n\mathbb{Z})$ on $M_U$ by

$$\mathcal{F}^U_{k,w,B}(\mathbb{Z}/\ell^n\mathbb{Z}) = \Big(M_{U'} \times (L_\lambda/\ell^n L_\lambda)\Big) /(U/U').$$

Write $\mathcal{F}^U_{k,w,B}$ for the inverse limit of these sheaves (as $n$ varies).

Write $\mathcal{O} = \mathcal{O}_{\mathrm{E},\lambda}$, and fix throughout an embedding of $\mathcal{O}$ into $\mathbb{C}$. If $N$ is an $\mathcal{O}$-module, define

$$\mathcal{L}^i_{k,w,B}(U; N) = H^i_{et}(M_U \otimes \overline{\mathrm{F}}, \mathcal{F}^U_{k,w,B} \otimes_{\mathcal{O}} N).$$

Again this makes sense whenever the dimension of $M_U(\mathbb{C})$ is positive.

We next define a sheaf "with representation" $r : U/U'(U \cap Z(\mathbb{Q})) \longrightarrow \mathrm{Aut}(V)$. (The reason for insisting $r$ be trivial on $U \cap Z(\mathbb{Q})$ is that the group $U/U'(U \cap Z(\mathbb{Q}))$ is exactly the group of the étale covering $M_{U'} \longrightarrow M_U$ when $U$ is sufficiently small.) Let $A$ be a finitely generated free $\mathcal{O}$-module such that there is a representation, which we also denote by $r$,

$$\begin{array}{rcl} r : U/U'(U \cap Z(\mathbb{Q})) & \longrightarrow & \mathrm{Aut}(A) \\ u & \longmapsto & r_u \end{array}$$

such that $A \otimes_{\mathcal{O}} \mathbb{C} \cong V$. We amend the above construction of the sheaf $\mathcal{F}^U_{k,w,B}$, by replacing $W_\lambda$ by $\mathrm{Hom}_{\mathrm{E}_\lambda}(A \otimes_{\mathcal{O}} \mathrm{E}_\lambda, W_\lambda)$, and choosing the lattice $\mathrm{Hom}_{\mathcal{O}}(A, L_\lambda)$ inside this space instead. Here $U$ acts on the right on $f \in \mathrm{Hom}_{\mathcal{O}}(A, L_\lambda)$ by $(fu)(\alpha) = \xi(u_\ell)^{-1} f(r_u(\alpha))$. In this case, one obtains a sheaf $\mathcal{F}^U_{k,w,B}(r)$. Write

$$\mathcal{L}^i_{k,w,B}(U, r; N) = H^i_{et}(M_U \otimes \overline{\mathrm{F}}, \mathcal{F}^U_{k,w,B}(r) \otimes_{\mathcal{O}} N)$$

for an $\mathcal{O}$-module $N$. We reiterate that these étale sheaves only make sense when $U$ is sufficiently small.

Write $\mathcal{L}_{k,w,B}(U, r; N)$ for $\mathcal{L}^i_{k,w,B}(U, r; N)$ if $B$ has $i$ split infinite places. When $\dim M_U(\mathbb{C}) = 0$, similar constructions give spaces

$$\mathcal{L}_{k,w,B}(U, r; N) = H^0(M_U(\mathbb{C}), \mathcal{F}^U_{k,w,B}(r)_{\mathbb{C}} \otimes_{\mathcal{O}} N),$$

the global sections of a locally free sheaf on a discrete set of points. Later, however, we will re-interpret this in terms of modules.

# 3 Eisenstein ideals

Suppose $S$ is a finite set of places of F containing all infinite places. We will introduce modules for a Hecke algebra $\mathbb{T}^S$, defined as the polynomial ring over $\mathbb{Z}$ generated by $T_{\mathsf{q}}$ and $S_{\mathsf{q}}$ for primes $\mathfrak{q} \notin S$.

Suppose $S$ also contains all primes over which $B$ is ramified. Then $\mathbb{T}^S$ acts naturally, through a quotient which we denote $\mathbb{T}^S_{k,w,B}(U)$, on the space $S_{k,w,B}(U)$ for any open compact subgroup $U = \prod_{\mathfrak{q}} U_{\mathfrak{q}}$ of $G(\widehat{\mathbb{Z}})$ such that $U_{\mathfrak{p}} = \mathrm{GL}_2(\mathcal{O}_{\mathrm{F},\mathfrak{p}})$ for all $\mathfrak{p} \notin S$. We say that $U$ is an *$S$-subgroup* when these conditions are verified. If again $G(\mathbb{A}^\infty) \supset U \supset U'$, with $U'$ normal in $U$, $U'$ an $S$-subgroup, then for any irreducible complex representation

$$r : U/U'(U \cap Z(\mathbb{Q})) \longrightarrow \mathrm{Aut}(V)$$

as before, we can define $\mathbb{T}^S_{k,w,B}(U, r)$.

In particular, suppose $S$ contains all prime ideals dividing some ideal $\mathfrak{n}$. Suppose $\mathsf{m}$ is a maximal ideal of $\mathbb{T}^S$ of residue characteristic $\ell$ in the support of $S_{k,w}(U_1(\mathfrak{n}))$ for $k$ arithmetic. Associated to $\mathsf{m}$ is the homomorphism

$$\overline{\theta}_{\mathsf{m}} : \mathbb{T}^S \longrightarrow\!\!\!\!\!\rightarrow \mathbb{T}^S/\mathsf{m} \hookrightarrow \overline{\mathbb{F}}_\ell$$

and an eigenform $f_{\mathsf{m}} \in S_{k,w}(U_1(\mathfrak{n}))$ such that, for $T \in \mathbb{T}^S$, one has $f_{\mathsf{m}}|T = \theta_{\mathsf{m}}(T) f_{\mathsf{m}}$ for some character $\theta_{\mathsf{m}}$ of $\mathbb{T}^S$ whose reduction modulo $\ell$ coincides with $\overline{\theta}_{\mathsf{m}}$. Then the semisimplification of the modulo $\ell$ reduction of the Galois representation associated (by Theorem 1.1) to $f_{\mathsf{m}}$ gives a representation

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

Its properties may be deduced from Theorem 1.1. In particular, we note that for primes $\mathfrak{q} \nmid \mathfrak{n}\ell$, one has

$$\begin{aligned} \mathrm{tr}\,\overline{\rho}_{\mathsf{m}}(\mathsf{Frob}_{\mathfrak{q}}) &= \overline{\theta}_{\mathsf{m}}(T_{\mathfrak{q}}) \\ \det \overline{\rho}_{\mathsf{m}}(\mathsf{Frob}_{\mathfrak{q}}) &= N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{q})\overline{\theta}_{\mathsf{m}}(S_{\mathfrak{q}}). \end{aligned}$$

For topological reasons, the image of $\overline{\rho}_{\mathsf{m}}$ is finite, so that $\overline{\rho}_{\mathsf{m}}$ factors through a finite group.

We now generalise from [6] the condition that a maximal ideal $\mathsf{m}$ as above should be Eisenstein.

Recall first from class field theory that there is an isomorphism

$$\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F})^{\mathrm{ab}} \cong \mathrm{F}^\times \backslash \mathbb{A}_{\mathrm{F}}^\times / (\mathrm{F}_\infty^\times)_+.$$

Any finite quotient of $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F})^{\mathrm{ab}}$ is then a quotient of some

$$Cl(\mathfrak{f}) = \mathrm{F}^\times \backslash \mathbb{A}_{\mathrm{F}}^\times / (\mathrm{F}_\infty^\times)_+ A_{\mathfrak{f}},$$

for some integral ideal $\mathfrak{f}$ of $\mathcal{O}_{\mathrm{F}}$, where

$$A_{\mathfrak{f}} = \{\pi \in \mathbb{A}_{\mathrm{F}}^{\infty\times} | \pi - 1 \in \mathfrak{f}\}.$$

$Cl(\mathfrak{f})$ is isomorphic to the narrow ray class group of F modulo $\mathfrak{f}$. Corresponding to $Cl(\mathfrak{f})$ is the ray class field $\mathrm{F}_{\mathfrak{f}}$, satisfying $\mathrm{Gal}(\mathrm{F}_{\mathfrak{f}}/\mathrm{F}) \cong Cl(\mathfrak{f})$.

We say that a maximal ideal $\mathsf{m}$ of $\mathbb{T}^S$ of residue characteristic $\ell$ as above is *Eisenstein* if there exists some integral ideal $\mathfrak{f}$ such that for all but finitely many prime ideals $\mathfrak{q}$ which are trivial in the narrow ray class group $Cl(\mathfrak{f})$, one has $T_{\mathfrak{q}} - 2 \in \mathsf{m}$ and $S_{\mathfrak{q}} - 1 \in \mathsf{m}$.

In a similar way to [6], Proposition 2, $\mathsf{m}$ is Eisenstein if and only if $\overline{\rho}_{\mathsf{m}}$ is reducible. (The proof is exactly the same, except that for the forward implication, one uses $H = \mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F}_{\mathfrak{f}}(\mu_\ell))$, and for the reverse implication, choose $\mathfrak{f}$ so that $\mathrm{F}_{\mathfrak{f}}$ contains $\mathrm{F}_{\mathfrak{n}\mathfrak{c}}(\mu_\ell)$, where $\mathfrak{c}$ is the conductor of the character $\chi$ appearing in the proof.)

Finally, we say that a $\mathbb{T}^S$-module is *Eisenstein* if all maximal ideals in its support are Eisenstein.

# 4 Carayol's Lemma

For the proof of Theorem 0.1, we will essentially follow the original method of Carayol ([4]) to lower the level at supercuspidal places. Namely, we use the Jacquet-Langlands correspondence to switch to a quaternion algebra which is ramified at $\mathfrak{p}$, and use an analogue of Lemme 1 of [4].

As Carayol uses geometric techniques, he switches to an indefinite quaternion algebra of discriminant $pq$, taking advantage of a level raising result to add an auxiliary prime $q$ to the level. He then applies his lemma to the corresponding Shimura curve.

In the setting of Hilbert modular forms, the level raising result is not yet known in full generality, although there are some unpublished results of Fujiwara along these lines (when $[\mathrm{F} : \mathbb{Q}]$ is even, the result is known—see Taylor [18] or the final section of this paper). We therefore follow the alternative approach outlined by Diamond ([5], §5), and prove an analogue of Carayol's Lemma by switching to a definite quaternion algebra.

The form of the statement of the theorem is taken from [7], Theorem 9.

**Theorem 4.1** *Suppose $B$ is a quaternion algebra over* F *(possibly split), and suppose $S$ is a finite set of places of* F *containing all infinite places of* F *and all ramified places of $B$. Let $k$ be arithmetic, and let $U$ and $U'$ be $S$-subgroups with $U'$ normal in $U$. Suppose $r$ (resp. $\chi$) is an irreducible representation (resp. a character with $\ell$-power order) of $U/U'(U \cap Z(\mathbb{Q}))$. Let*

$\overline{\theta} : \mathbb{T}^S_{k,w,B}(U,r) \longrightarrow \overline{\mathbb{F}}_\ell$ *be a homomorphism for which* $\overline{\rho}_{\overline{\theta}}$ *is irreducible. Then there exists a homomorphism* $\overline{\theta}' : \mathbb{T}^S_{k,w,B}(U, r \otimes \chi) \longrightarrow \overline{\mathbb{F}}_\ell$ *such that the two maps* $\mathbb{T}^S \longrightarrow \overline{\mathbb{F}}_\ell$ *induced by* $\overline{\theta}$ *and* $\overline{\theta}'$ *coincide.*

We first observe that under the hypothesis on $\ell$ made in Theorem 0.1, it suffices to prove the result for $S$-subgroups $U$ which are "sufficiently small" in the sense of §2. For this, one follows [12], §12, to find a prime ideal $\mathfrak{q}_0 \notin S$ and $\mathfrak{q}_0 \nmid \ell$ such that

- $U_1^1(\mathfrak{q}_0)$ is "sufficiently small"
- there are no congruences between forms of level $U$ and $\mathfrak{q}_0$-new forms of level dividing $U \cap U_1^1(\mathfrak{q}_0)$.

Here,

$$U_1^1(\mathfrak{q}_0) = \left\{ \alpha \in G(\widehat{\mathbb{Z}}) \,\middle|\, \alpha \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod{\mathfrak{q}_0} \right\}$$

Then applying Theorem 4.1 with $S$ replaced by $S \cup \{\mathfrak{q}_0\}$, $U$ by $U \cap U_1^1(\mathfrak{q}_0)$ and $U'$ by $U' \cap U_1^1(\mathfrak{q}_0)$, and using the second property of $U_1^1(\mathfrak{q}_0)$ above, one sees that the conclusion of Theorem 4.1 is also valid for $U$, even if $U$ is not "sufficiently small".

In particular, we will be able to consider the case $U = U_1(\mathfrak{n})$.

By the Jacquet-Langlands correspondence, it suffices to prove this result if $B$ has at most one split infinite place. Thus we consider the two cases in which $B$ is totally definite, and where $B$ has exactly one split infinite place. (As $\mathrm{F} \neq \mathbb{Q}$, $B$ is then never split.)

Exactly as in [7], it suffices to prove

**Proposition 4.2** *Suppose that* $\overline{\theta} : \mathbb{T}^S \longrightarrow \overline{\mathbb{F}}_\ell$ *is a homomorphism with non-Eisenstein kernel* $\mathsf{m}$*,* $U$ *a sufficiently small* $S$*-subgroup. Let* $\kappa$ *be the residue field of* $\mathcal{O}$*. Then* $\overline{\theta}$ *factors through* $\mathbb{T}^S_{k,w,B}(U,r)$ *if and only if* $\mathsf{m} \in \operatorname{supp} \mathcal{L}_{k,w,B}(U, r; \kappa)$*.*

We will prove this proposition in the following two sections.

# 5 The totally definite case

In this case, $B$ is totally definite, and $M_U(\mathbb{C})$ consists of a finite number of points. Note that $G(\mathbb{R}) = C_\infty$. We can interpret the space of cusp forms in the following manner.

Recall that in §2 we defined the space $L_{k,w,B}(\mathcal{O}_{\mathrm{E}})$ as a representation space (denoted $\xi$) of $M_2(\mathcal{O}_{\mathrm{E}})^I \cong \mathcal{O}_B \otimes_{\mathbb{Z}} \mathcal{O}_{\mathrm{E}}$. If $R$ is an $\mathcal{O}_{\mathrm{E}}$-algebra, the same procedure gives a representation $\xi$ of $(\mathcal{O}_B \otimes_{\mathbb{Z}} \mathcal{O}_{\mathrm{E}}) \otimes_{\mathcal{O}_{\mathrm{E}}} R \cong M_2(R)^I$ on a space $L_{k,w,B}(R)$. If $N$ is an $R$-module, we write $L_{k,w,B}(N) = L_{k,w,B}(R) \otimes_R N$. We will usually apply this with $R = \mathcal{O}$.

If $f : G(\mathbb{A}) \longrightarrow L_{k,w,B}(\mathbb{C})$, and $u = u^\infty u_\infty \in G(\mathbb{A}^\infty) \times G(\mathbb{R})$, define the transform

$$(f|_{k,w}u)(g) = \xi(u_\infty)^{-1} f(gu^{-1}).$$

Then for an open compact subgroup $U \subset G(\widehat{\mathbb{Z}})$, a model for $S_{k,w,B}(U)$ is

$$S_{k,w,B}(U) = \{f : G(\mathbb{Q})\backslash G(\mathbb{A}) \longrightarrow L_{k,w,B}(\mathbb{C}) |\ f|_{k,w}u = f \text{ for } u \in UG(\mathbb{R})\}.$$

Note that this condition is equivalent to

$$f(gu) = \xi(u_\infty)^{-1} f(g).$$

Define also

$$S_{k,w,B}(U)^{\mathrm{triv}} = \begin{cases} \{f \in S_{2t,t,B}(U) | f \text{ factors through } \nu\}, & \text{if } k = 2t, \\ 0, & \text{if } k \neq 2t. \end{cases}$$

Write

$$\widetilde{S}_{k,w,B}(U) = S_{k,w,B}(U)/S_{k,w,B}(U)^{\mathrm{triv}}.$$

It is this space which really corresponds to the space of cusp forms. The inductive limit of these spaces (as $U$ gets smaller and smaller) becomes an admissible $G(\mathbb{A}^\infty)$-module via

$$(gf)(h) = f(hg).$$

We now turn to the $\ell$-adic model denoted $\mathcal{L}_{k,w,B}(U; N)$ by analogy with the notation of §2. Here $N$ will be an $\mathcal{O}$-module. Then the action of $\mathcal{O}_B^\times$ on $L_{k,w,B}(N)$ extends to an action of $(\mathcal{O}_B \otimes_{\mathbb{Z}} \mathbb{Z}_\ell)^\times$ by continuity, and we also denote this action by $\xi$. If $U$ is as above, and also $U_\ell = \prod_{\lambda|\ell} U_\lambda$ is contained in $(\mathcal{O}_B \otimes_{\mathbb{Z}} \mathbb{Z}_\ell)^\times$ (as will be the case for all $U$ of interest in this paper), then $\mathcal{L}_{k,w,B}(U; N)$, for an $\mathcal{O}$-module $N$, is the set

$$\{f : G(\mathbb{Q})\backslash G(\mathbb{A}^\infty) \longrightarrow L_{k,w,B}(N) | \xi(u_\ell)^{-1} f(g) = f(gu) \text{ for } u \in U\}.$$

Define $\mathcal{L}_{k,w,B}(U; N)^{\mathrm{triv}}$ and $\widetilde{\mathcal{L}}_{k,w,B}(U; N)$ in the above manner. Again, the inductive limit of these spaces is an admissible $G(\mathbb{A}^\infty)$-module, via

$$(gf)(h) = \xi(g_\ell) f(hg).$$

We give models for the corresponding objects with representation $r : U/U'(U \cap Z(\mathbb{Q})) \longrightarrow \mathrm{Aut}(V)$ as at the end of §2. Again,

$$S_{k,w,B}(U, r) = \mathrm{Hom}_U(V, S_{k,w,B}(U')),$$

and define its trivial subspace to consist of those homomorphisms which are valued in $S_{k,w,B}(U')^{\mathrm{triv}}$. We find that $\mathcal{L}_{k,w,B}(U, r; N)$ is the set of functions

$$f : G(\mathbb{Q})\backslash G(\mathbb{A}^\infty) \longrightarrow \mathrm{Hom}_{\mathcal{O}}(A, L_{k,w,B}(N))$$

satisfying

$$f_{gu}(a) = \xi(u_\ell)^{-1} f_g(r_u(a)) \qquad \text{for } u \in U,\ a \in A,$$

where we write $f_g$ for the map $f(g)$. The spaces $\mathcal{L}_{k,w,B}(U, r; N)^{\mathrm{triv}}$ and $\widetilde{\mathcal{L}}_{k,w,B}(U, r; N)$ are defined in the now familiar way.

Again, the inductive limits of these constructions are admissible $G(\mathbb{A}^\infty)$-modules. The actions of $G(\mathbb{A}^\infty)$ give rise to an action of Hecke operators at finite level (see [13] or [6] for the definitions). Suppose $U$ and $U'$ are $S$-subgroups. Then $\mathbb{T}^S$ acts on all of the above spaces. Recall that we fixed an embedding $\mathcal{O} \hookrightarrow \mathbb{C}$; this induces a $\mathbb{T}^S$-equivariant isomorphism

$$S_{k,w,B}(U, r) \cong \mathcal{L}_{k,w,B}(U, r; \mathbb{C})$$

under which the trivial subspaces correspond. For prime ideals $\mathfrak{q}$ trivial in the narrow ray class group $Cl(\mathfrak{n})$ (where $\mathfrak{n}$ is the level of $U$), one finds from the formulae (see [11], (3.9)), that $S_{\mathfrak{q}}$ acts on the trivial subspaces by 1, and that $T_{\mathfrak{q}}$ acts by $(1 + N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{q}))$ (as functions in the trivial subspaces factor through the norm). It follows that the space $\mathcal{L}_{k,w,B}(U, r; \mathbb{C})^{\mathrm{triv}}$ is Eisenstein.

Given $U$, fix a decomposition

$$G(\mathbb{A}^\infty) = \bigoplus_{i=1}^{r} G(\mathbb{Q}) g_i U.$$

We have an isomorphism

$$\mathcal{L}_{k,w,B}(U, r; N) \cong \bigoplus_{i=1}^{r} \mathrm{Hom}_i,$$

where $\mathrm{Hom}_i$ denotes the set of homomorphisms $\theta \in \mathrm{Hom}_{\mathcal{O}}(A, L_{k,w,B}(N))$ such that

$$\theta(a) = \xi(u_\ell)^{-1} \theta(r_u(a)) \text{ for } a \in A, u \in U \cap g_i^{-1} G(\mathbb{Q}) g_i.$$

The isomorphism is given by sending $f$ to $(f_{g_i})_{i=1}^r$.

The condition that $U$ is sufficiently small implies that

$$U \cap g_i^{-1} G(\mathbb{Q}) g_i \subset U \cap Z(\mathbb{Q})_+ \subset (\mathcal{O}_{\mathrm{F}}^\times)_+,$$

the totally positive units in $\mathcal{O}_{\mathrm{F}}$ (see [2], 1.4.1.1 and its proof). In this case, if $a \in A$ and $u \in U \cap g_i^{-1}G(\mathbb{Q})g_i$, $r_u(a) = a$ by definition of the representation $r$. Further, if $\alpha \in (\mathcal{O}_{\mathrm{F}}^\times)_+$, it is easy to show that $\xi(\alpha) = 1$. Thus the condition above on homomorphisms $f \in \mathrm{Hom}_{\mathcal{O}}(A, L_{k,w,B}(N))$ is automatic, and we see that, for any $\mathcal{O}$-module $N$,

$$\begin{aligned}
\mathcal{L}_{k,w,B}(U,r;N) &\cong \bigoplus_{i=1}^{r} \mathrm{Hom}_{\mathcal{O}}(A, L_{k,w,B}(N)) \\
&= \bigoplus_{i=1}^{r} \mathrm{Hom}_{\mathcal{O}}(A, L_{k,w,B}(\mathcal{O}) \otimes_{\mathcal{O}} N) \\
&\cong (\bigoplus_{i=1}^{r} \mathrm{Hom}_{\mathcal{O}}(A, L_{k,w,B}(\mathcal{O}))) \otimes_{\mathcal{O}} N \\
&\qquad \text{(as } A \text{ and } L_{k,w,B}(\mathcal{O}) \text{ are finite free } \mathcal{O}\text{-modules)} \\
&\cong \mathcal{L}_{k,w,B}(U,r;\mathcal{O}) \otimes_{\mathcal{O}} N
\end{aligned}$$

We immediately deduce the following lemma (as in [7], Lemma 5):

**Lemma 5.1** *Suppose $\theta : \mathbb{T}^S \longrightarrow \overline{\mathbb{F}}_\ell$ is a homomorphism with kernel $\mathsf{m}$ which is not Eisenstein. Then $\theta$ factors through $\mathbb{T}^S_{k,w,B}(U,r)$ if and only if $\mathsf{m} \in$* $\mathrm{supp}\,\mathcal{L}_{k,w,B}(U,r;\kappa)$.

# 6 One split infinite place

A method similar to that of Carayol ([3], 2.2.4) shows that

$$\mathcal{L}^1_{k,w,B}(U,r;\mathbb{C}) \cong S_{k,w,B}(U,r)^2,$$

and this isomorphism is equivariant for the action of $\mathbb{T}^S$. (This is also a special case of the Matsushima-Shimura theorem: see, for example, [11], 6.2.) Write $\mathcal{L}_{k,w,B}(U,r;N)$ for $\mathcal{L}^1_{k,w,B}(U,r;N)$.

One has the following lemma, which acts as a substitute for the final two isomorphisms of §5:

**Lemma 6.1 ([7], Lemma 6)** *Let $\mathsf{m}$ be a maximal ideal of $\mathbb{T}^S$ with residue characteristic $\ell$. Suppose either that $\mathsf{m} \in \mathrm{supp}\,\mathcal{L}_{k,w,B}(U,r;\mathcal{O})^{\mathrm{tors}}$ or that $\mathsf{m} \in \mathrm{supp}\,\mathcal{L}_{k,w,B}(U,r;\kappa)/\mathcal{L}_{k,w,B}(U,r;\mathcal{O}) \otimes \kappa$, where $\kappa$ is as in the statement of Proposition 4.2. Then $\mathsf{m}$ is Eisenstein.*

**Proof.** In the same way as [7], we see that $\mathcal{L}_{k,w,B}(U,r;\mathcal{O})^{\mathrm{tors}}$ is a quotient of $\mathcal{L}^0_{k,w,B}(U,r;\ell^{-n}\mathcal{O}/\mathcal{O})$ for some $n$, and $\mathcal{L}_{k,w,B}(U,r;\kappa)/\mathcal{L}_{k,w,B}(U,r;\mathcal{O})\otimes\kappa$ injects into $\mathcal{L}^2_{k,w,B}(U,r;\ell^{-n}\mathcal{O}/\mathcal{O})$.

But the action of $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F})$ on $\mathcal{L}^0_{k,w,B}(U,r;N)$ factors through $\mathrm{Gal}(\mathrm{F}^{\mathrm{ab}}/\mathrm{F})$ as the components of $M_U\times_{\mathrm{F}}\overline{\mathrm{F}}$ are defined over $\mathrm{F}^{\mathrm{ab}}$ ([2]); the same holds for $\mathcal{L}^2_{k,w,B}(U,r;N)$ by Poincaré duality. Exactly as in [6], Lemma 3, it follows that $\mathsf{m}$ is Eisenstein.

Now one deduces in the same way as §5 (exactly as in [7]), that

**Lemma 6.2** *Suppose* $\theta:\mathbb{T}^S\longrightarrow\overline{\mathbb{F}}_\ell$ *is a homomorphism with kernel* $\mathsf{m}$ *which is not Eisenstein. Then* $\theta$ *factors through* $\mathbb{T}^S_{k,w,B}(U,r)$ *if and only if* $\mathsf{m}\in\operatorname{supp}\mathcal{L}_{k,w,B}(U,r;\kappa)$.

Lemmas 5.1 and 6.2 together complete the proof of Theorem 4.1.

# 7 Lowering the level results–$[\mathrm{F}:\mathbb{Q}]$ odd

In this section, we prove analogues of Carayol's results in the case where $d=[\mathrm{F}:\mathbb{Q}]$ is odd. We remind the reader that we still insist that $\ell$ satisfies the hypothesis of Theorem 0.1.

Fix a mod $\ell$ representation $\overline{\rho}$ of $\mathrm{Gal}(\overline{\mathrm{F}}/\mathrm{F})$ which is modular.

As mentioned above, we will assume the conjectural analogue of Ribet's theorem ([16] and [5]), namely, Conjecture 2, although we remind the reader (§7.1 below) that it is known in many cases. We will now consider the cases of Theorem 1.5 in turn, and try to remove a prime from the level.

## 7.1 Case (1)

Conjecture 2 is exactly a reformulation of case (1). Although we will assume it true in general, the main result of [12] proves this conjecture in many cases, as we now explain.

Here, $\pi_{\mathfrak{p}}$ is special unramified at $\mathfrak{p}$, but $v_{\mathfrak{p}}(\mathfrak{n}(\overline{\rho}))=0$. It thus corresponds to a Hilbert modular form in $A^0_{k,w}(U_0(\mathfrak{n}\mathfrak{p}),\mu)$ where $\mu$ factors through $(\mathcal{O}_{\mathrm{F}}/\mathfrak{n})^\times$. We use the Jacquet-Langlands correspondence to switch from $\mathrm{GL}_{2/\mathrm{F}}$ to the quaternion algebra $B_{/\mathrm{F}}$ ramified at $\{\tau_2,\ldots,\tau_d\}$. Applying the main result of [12] if the appropriate hypotheses hold, and switching back to $\mathrm{GL}_{2/\mathrm{F}}$ gives the required result. Thus there is a form in $A_{k,w}(U_0(\mathfrak{n}),\mu')$ (for some $\mu'$) with the same modulo $\ell$ Galois representation (i.e., proving Conjecture 2) under the supplementary hypotheses listed in the statement of Theorem 0.1:

**Theorem 7.1** *Conjecture 2 holds if* $N_{\mathrm{F}/\mathbb{Q}}(\mathfrak{p}) \not\equiv 1$ *(mod* $\ell$*).*

## 7.2 Case (2)

When $\mathrm{F} = \mathbb{Q}$, Carayol uses a local twisting argument to solve this case. Such an argument is not always available for more general totally real number fields. Instead we adapt slightly the argument of [4], §5, originally intended to solve the analogue of Case (4) (see Theorem 7.3 below).

Then $\pi_{\mathfrak{p}}$ is a special representation, corresponding to a degenerating character of $\mathrm{F}_{\mathfrak{p}}^{\times}$. If $\mathfrak{n}$ is the conductor of $\pi$, then $v_{\mathfrak{p}}(\mathfrak{n}) = 2$. Write $\mathfrak{n} = \mathfrak{n}'\mathfrak{p}^2$.

**Theorem 7.2** *Assume Conjecture 2. If* $\pi_{\mathfrak{p}}$ *is as in Case (2) of Theorem 1.5, then there exists* $\pi'$*, of conductor dividing* $\mathfrak{n}'\mathfrak{p}$*, which also gives rise to* $\overline{\rho}$*.*

**Proof.** We switch, using the Jacquet-Langlands correspondence to the quaternion algebra $B$, ramified at $\{\mathfrak{p}, \tau_1, \ldots, \tau_d\}$. Choose a maximal order $\mathcal{O}_B$ of $B$. Suppose that, under this correspondence, $\pi$ is sent to $\widetilde{\pi}$. We can describe the local component $\widetilde{\pi}_{\mathfrak{p}}$ following [7]. We find that $\widetilde{\pi}_{\mathfrak{p}}$ is $\xi \circ \nu$, where $\nu$ denotes the reduced norm on $B_{\mathfrak{p}}$, and $\xi$ is a tamely ramified character with unramified reduction.

We apply Theorem 4.1 above, with $U = U_1(\mathfrak{n}')\mathcal{O}_{B_{\mathfrak{p}}}^{\times}$ and $U' = U_1(\mathfrak{n}')U(\mathfrak{p})$, where $U(\mathfrak{p})$ denotes the units of $\mathcal{O}_{B_{\mathfrak{p}}}^{\times}$ congruent to 1 modulo $\mathfrak{m}_{B_{\mathfrak{p}}}$, the maximal ideal of $\mathcal{O}_{B_{\mathfrak{p}}}$, and we think of $U_1(\mathfrak{n}')$ as contained in $\prod_{\mathfrak{q}\neq\mathfrak{p}} \mathrm{GL}_2(\mathcal{O}_{\mathrm{F}_{\mathfrak{q}}})$. Let $r = \widetilde{\pi}_{\mathfrak{p}}|_{\mathcal{O}_{B_{\mathfrak{p}}}^{\times}}$ (regarded in the obvious way as a representation of $U/U'(U \cap Z(\mathbb{Q}))$) which is itself a character of $\ell$-power order, and let $\chi = r^{-1}$. Let $S$ be a finite set of places including all infinite places and all primes dividing $\mathfrak{n}$. Let

$$\overline{\theta} : \mathbb{T}_{k,w,B}^{S}(U, r) \longrightarrow \overline{\mathbb{F}}_{\ell}$$

correspond to $\widetilde{\pi}$. Then apply Theorem 4.1 to find an automorphic representation $\widetilde{\pi}'$ corresponding to

$$\overline{\theta}' : \mathbb{T}_{k,w,B}^{S}(U, 1) \longrightarrow \overline{\mathbb{F}}_{\ell}.$$

Then from the dictionary of the Jacquet-Langlands correspondence, $\widetilde{\pi}'$ corresponds to an automorphic representation $\pi'$ on $\mathrm{GL}_{2/\mathrm{F}}$ which is special unramified at $\mathfrak{p}$. The result follows.

## 7.3 Case (3)

We say a few words here about the principal series case, in which one of the two defining characters is tamely ramified with unramified reduction. Carayol uses a version of Theorem 4.1 when the other character is unramified, and a local twisting argument when the other character is ramified. As remarked above, this local twisting argument is not valid for general totally real fields, and we do not consider this case here. However, we note that our version of Theorem 4.1 suffices to prove Case (3) when one of the characters is tamely ramified but with unramified reduction, and the other character is unramified. In this case, if the conductor of $\pi$ is $\mathfrak{n}$, then $v_{\mathfrak{p}}(\mathfrak{n}) = 1$. Write $\mathfrak{n} = \mathfrak{n}'\mathfrak{p}$. Exactly as in [4], we apply Theorem 4.1 with $r$ being the central character of $\pi$, regarded as a character of $U_1(\mathfrak{n}')/U_1(\mathfrak{n})$. As in Case (2), $r$ is 1-dimensional, and we let $\chi = r^{-1}$. Theorem 4.1 provides an automorphic representation on $U_1(\mathfrak{n}')$ which is unramified principal series at $\mathfrak{p}$, and so has conductor $\mathfrak{n}'$.

## 7.4 Case (4)

For this case, we follow [4], §5, as in §7.2 above.

Here $\pi_{\mathfrak{p}}$ is a supercuspidal representation, associated to a degenerating character $\xi$ of $\Omega^\times$, where $\Omega$ is the unramified quadratic extension of $\mathrm{F}_{\mathfrak{p}}$, which does not factor through $N_{\Omega/\mathrm{F}_{\mathfrak{p}}}$. If $\mathfrak{n}$ is the conductor $a(\pi)$ of $\pi$, then this implies that $v_{\mathfrak{p}}(\mathfrak{n}) = 2$. Write $\mathfrak{n} = \mathfrak{n}'\mathfrak{p}^2$.

**Theorem 7.3** *Assume Conjecture 2. If $\pi_{\mathfrak{p}}$ is as in case (4) of Theorem 1.5, then there exists $\pi'$, of conductor dividing $\mathfrak{n}'\mathfrak{p}$, which also gives rise to $\overline{\rho}$.*

**Proof.** We switch, using the Jacquet-Langlands correspondence once more, to the quaternion algebra $B$ of 7.2. Suppose that $\pi$ is sent to $\widetilde{\pi}$. Using [10], we may describe the image $\widetilde{\pi}_{\mathfrak{p}}$, a representation of $B_{\mathfrak{p}}^\times = (B \otimes \mathrm{F}_{\mathfrak{p}})^\times$. This is exactly as in [4].

Let $\kappa_\Omega$ denote the residue field of $\Omega$, and choose an embedding $\Omega \hookrightarrow B_{\mathfrak{p}}$. Let

$$U(\mathfrak{p}) = \{\alpha \in \mathcal{O}_{B_{\mathfrak{p}}}^\times | \alpha - 1 \in \mathfrak{m}_{B_{\mathfrak{p}}}\}.$$

Then

$$\widetilde{\pi}_{\mathfrak{p}} = \mathrm{Ind}_{\Omega^\times U(\mathfrak{p})}^{B_{\mathfrak{p}}^\times} \widetilde{\xi},$$

where $\widetilde{\xi}$ is a character trivial on $U(\mathfrak{p})$ and agreeing with $\xi$ on $\Omega^\times$. (Note that $\Omega^\times U(\mathfrak{p})$ is of index 2 in $B_{\mathfrak{p}}^\times$.)

We apply Theorem 4.1 with $U = U_1(\mathfrak{n}')\mathcal{O}_{B_{\mathfrak{p}}}^\times$, $U' = U_1(\mathfrak{n}')U(\mathfrak{p})$, where we think of $U_1(\mathfrak{n}')$ as contained in $\prod_{\mathfrak{q}\neq\mathfrak{p}} \mathrm{GL}_2(\mathcal{O}_{\mathrm{F}_{\mathfrak{q}}})$. Then

$$U/U' \cong \mathcal{O}_{B_{\mathfrak{p}}}^\times / U(\mathfrak{p}) \cong \kappa_\Omega^\times.$$

Then $\widetilde{\pi}_{\mathfrak{p}}|_{\mathcal{O}_{B_{\mathfrak{p}}}^{\times}}$ is the sum $\widetilde{\xi} \oplus \widetilde{\xi}^{\sigma}$. We let $r = \widetilde{\xi}|_{\mathcal{O}_{B_{\mathfrak{p}}}^{\times}}$, which we may think of as a representation of $U/U' \cong \kappa_{\Omega}^{\times}$. Let $S$ be a finite set of places including all infinite places and finite primes dividing $\mathfrak{n}$. Let

$$\overline{\theta} : \mathbb{T}_{k,w,B}^{S}(U, r) \longrightarrow \overline{\mathbb{F}}_{\ell}$$

correspond to $\widetilde{\pi}$. But $\xi$ is a character of $\Omega^{\times}$ with $a(\xi) = 1$, so $\xi$ is trivial on $1 + \mathfrak{m}_{\Omega}$. Thus $\xi$ restricts to a character of $\mathcal{O}_{\Omega}^{\times}/(1 + \mathfrak{m}_{\Omega}) \cong \kappa_{\Omega}^{\times}$. In this way, we can think of $\xi$ (and $\widetilde{\xi}$) as characters of $U/U'$. Note also that as $a(\overline{\xi}) = 0$, $\xi$ has trivial reduction.

Put $\chi = \widetilde{\xi}^{-1}$. Then Theorem 4.1 furnishes an automorphic representation $\widetilde{\pi}'$ corresponding to

$$\overline{\theta}' : \mathbb{T}_{k,w,B}^{S}(U, 1) \longrightarrow \overline{\mathbb{F}}_{\ell}$$

Switching back to $\mathrm{GL}_{2/\mathrm{F}}$, and recalling the Jacquet-Langlands dictionary from [7], we obtain an automorphic representation $\pi'$ whose component at $\mathfrak{p}$ is special unramified (as the trivial character does factor through the norm), and thus $v_{\mathfrak{p}}(a(\pi')) = 1$.

# 8 Lowering the level results–[F : $\mathbb{Q}$] even

Suppose that $\overline{\rho}$ arises from the cuspidal automorphic representation $\pi$ on $\mathrm{GL}_{2/\mathrm{F}}$. If there is a finite prime $\mathfrak{q}_1$ of F at which $\pi_{\mathfrak{q}_1}$ is square-integrable (special or supercuspidal), then, by the Jacquet-Langlands correspondence, $\overline{\rho}$ also arises from an automorphic representation on the quaternion algebra $B$ ramified at $\{\mathfrak{q}_1, \tau_2, \ldots, \tau_d\}$. In this case, the results and proofs of §7 go through, using this quaternion algebra to solve Case (1), and using the algebra got by changing the invariants at $\mathfrak{p}$ and at $\tau_1$ to solve Case (4). In this case, the proof of [12] of Conjecture 2 under the supplementary hypotheses goes through.

If, on the other hand, no such prime $\mathfrak{q}_1$ exists, we use the following result of Taylor ([18]):

**Theorem 8.1** *Suppose* [F : $\mathbb{Q}$] *is even. Let* $f \in A_{k,w}(U_0(\mathfrak{n}), \mu)$. *Then there exists an auxiliary prime* $\mathfrak{q} \nmid \mathfrak{n}\ell$, *and* $f' \in A_{k,w}(U_0(\mathfrak{n}\mathfrak{q}), \mu')$, *with* $\mu'$ *factorising through* $(\mathcal{O}_{\mathrm{F}}/\mathfrak{n})^{\times}$, *new at* $\mathfrak{q}$, *such that* $\rho_{f,\lambda} \equiv \rho_{f',\lambda}$ *(mod* $\lambda$*).*

We may then assume that the fixed mod $\ell$ representation $\overline{\rho}$ arises from a form $f$ which lies in $A_{k,w}^{0}(U_0(\mathfrak{n}\mathfrak{q}_1'), \mu)$, and which is special unramified at $\mathfrak{q}_1'$. (It may be necessary to replace $\mathfrak{n}$ with some ideal dividing $\mathfrak{n}$.) Then $\mu$ is a character on $(\mathcal{O}_{\mathrm{F}}/\mathfrak{n})^{\times}$. Then, as above, $\overline{\rho}$ also arises from an automorphic representation on the quaternion algebra $B$ ramified at $\{\mathfrak{q}_1', \tau_2, \ldots, \tau_d\}$.

Again, the results and proofs of §7 go through, using this quaternion algebra to solve Cases (1)–(3), and using the algebra got by changing the invariants at $\mathfrak{p}$ and at $\tau_1$ to solve Case (4). Having removed an appropriate power of $\mathfrak{p}$ from the level, it now remains to remove the auxiliary prime $\mathfrak{q}_1'$ which we added. But for this, we may again use Conjecture 2.

# References


[1] K. Buzzard, The levels of modular representations, PhD Thesis, Cambridge 1995.

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

[3] H. Carayol, 'Sur les représentations $\ell$-adiques associées aux formes modulaires de Hilbert', *Ann. Sci. Ec. Norm. Sup., 4*$^{e}$ *série 19* (1986) 409–468

[4] H. Carayol, 'Sur les représentations galoisiennes modulo $\ell$ attachées aux formes modulaires', *Duke Math. J. 59* (1989) 785–801

[5] F. Diamond, 'The refined conjecture of Serre', *in 'Elliptic curves, modular forms and Fermat's Last Theorem', J. Coates and S.T. Yau (eds.)* (1995)

[6] F. Diamond, R. Taylor, 'Non-optimal levels of mod $\ell$ modular representations', *Invent. Math. 115* (1994) 435–462

[7] F. Diamond, R. Taylor, 'Lifting modular mod $\ell$ representations', *Duke Math. J. 74* (1994) 253–269

[8] B. Edixhoven, 'The weight in Serre's conjectures', *Invent. Math. 109* (1992) 563–594

[9] E. Freitag, *Hilbert modular forms*, Springer (1990)

[10] P. Gérardin, 'Facteurs locaux des algèbres simples de rang 4, I', *Publ. Math. Paris VII* (1978) 37–77

[11] H. Hida, 'On $p$-adic Hecke algebras for $\mathrm{GL}_2$ over totally real fields', *Ann. Math. 128* (1988) 295–384

[12] F. Jarvis, 'Mazur's Principle for totally real fields of odd degree', *To appear in Compositio Math.* (1998)

[13] R. Langlands, 'Modular forms and $\ell$-adic representations', *Springer Lecture Notes in Math. 349* (1973) 361–500

[14] R. Livné, 'On the conductors of modulo $\ell$ representations coming from modular forms', *Journal of Number Theory 31* (1989) 133–141

[15] K. Ribet, 'On modular representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms', *Invent. Math. 100* (1990) 431–476

[16] K. Ribet, 'Report on mod $\ell$ representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$', *in Proc. Symp. in Pure Math. AMS, vol. 55, part 2* (1994) 639–676

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

[18] R. Taylor, 'Representations of Galois groups associated to Hilbert modular forms', *in 'Automorphic forms, Shimura varieties and L-functions', L. Clozel and J. Milne (eds.)* (1989)