ON SERRE'S CONJECTURE FOR MOD ℓ GALOIS REPRESENTATIONS OVER TOTALLY REAL FIELDS

KEVIN BUZZARD, FRED DIAMOND, AND FRAZER JARVIS

1. Introduction

Serre conjectured in [56] that if ℓ is prime and

$$\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$$

is a continuous, odd, irreducible representation, then ρ is *modular* in the sense that it arises as the reduction of an ℓ-adic representation associated to a Hecke eigenform in the space $S_k(\Gamma_1(N))$ of cusp forms of some weight k and level N. Let us refer to this (incredibly strong) conjecture as "the weak conjecture". Serre goes on to formulate a refined conjecture which predicts the minimal weight and level of such an eigenform subject to the constraints $k \geq 2$ and $\ell \nmid N$; let us call this "Serre's refined conjecture". Note that Serre explicitly excludes weight 1 modular forms, although a further reformulation was made by Edixhoven in [20] to include them, and we refer to Edixhoven's reformulation as "Edixhoven's refined conjecture". Through the work of Ribet [51], Gross [34], Coleman-Voloch [11] and others, the equivalence between the weak conjecture and Serre's refinement was known for $\ell > 2$ (see [16]), and also when $\ell = 2$ in many cases (see [8]). The equivalence of Serre's refined conjecture and Edixhoven's refined conjecture is also essentially known, although the question does not appear to have been completely resolved: for $\ell = 2$, there still appears to be an issue regarding constructing a weight 1 form in every case that Edixhoven predicts that such a form exists.

The aim of this paper is to formulate a generalisation of Serre's refined conjecture to the context of two-dimensional representations of G_K where K is a totally real field. The details of such a formulation (assuming ℓ unramified in K) were worked out by one of us (F.D.) stemming from correspondence and conversations among the authors in 2002, and the first version of this paper appeared in 2004. At that time the weak conjecture (for $G_{\mathbf{Q}}$) appeared out of reach. Since then there has been startling progress, culminating in its recent proof by Khare and Wintenberger [44, 45], building on ideas developed by Dieulefait and themselves [19, 65, 43, 42] and relying crucially on potential modularity and modularity lifting methods and results of Taylor, Wiles and Kisin [64, 61, 60, 46, 47]. Their result also resolves the remaining case for $\ell = 2$ of Serre's refined conjecture.

Since the first version of this paper appeared, there has also been significant progress towards proving the equivalence between the "weak" and "refined" conjectures we presented over G_K. Partial results already followed from work of one of the authors [37, 39], Fujiwara [28] and Rajaei [50], and further results were subsequently obtained by Schein [54] and Gee [29]. For the most part the techniques were generalisations of ones already used in the case $K = \mathbf{Q}$ and seemed severely

Date: 10 October 2008.

limited with respect to establishing the "weight part" of the refined conjecture. However in [31, 32] Gee presented a new, much more promising approach; it remains to be seen how far the ideas there can be pushed towards a complete proof of the equivalence between weak and refined conjectures.

Another important development related to Serre's conjecture has been the recent progress on constructing ℓ-adic and mod ℓ Langlands correspondences, especially the work of Breuil, Colmez and Emerton. In particular, a correspondence between two-dimensional ℓ-adic representations of $G_{\mathbf{Q}_\ell}$ and certain ℓ-adic representations of $\mathrm{GL}_2(\mathbf{Q}_\ell)$ was constructed by Colmez [12], and a conjectural compatibility with a global correspondence was formulated and proved in many cases by Emerton in [22] (see also [21]). There is also a mod ℓ version of this compatibility, which we refer to as "Emerton's refined conjecture" (see also [23]). Most cases of Serre's refined conjecture follow from Emerton's; in particular the specification of the weight is essentially a description of the $\mathrm{GL}_2(\mathbf{Z}_\ell)$-socle of the representation of $\mathrm{GL}_2(\mathbf{Q}_\ell)$ arising as a local factor at ℓ associated to ρ. The current version of this paper includes a partial generalisation of Emerton's refined conjecture to certain forms of GL_2 over K. The sense in which it is "partial" is that at primes over ℓ we only describe the Jordan-Hölder constituents of the socle of a maximal compact for the local factor, and even that only for primes unramified over ℓ. The relevance of this socle and the difficulty of generalising the mod ℓ local correspondence to extensions of $\mathbf{Q}_\ell$ is clearly illustrated by the recent work of Breuil and Paskunas [7].

We now explain our set-up and aims in a little more detail. Suppose that K is a totally real field. Let $\mathcal{O}$ denote its ring of integers and let S_K be the set of embeddings of K in $\mathbf{R}$. Suppose that $\vec{k} \in \mathbf{Z}^{S_K}$ with $k_\tau \geq 1$ for all $\tau \in S_K$ and furthermore assume that all of the k_τ are of the same parity. Let $\mathfrak{n}$ be a non-zero ideal of $\mathcal{O}$. The space of Hilbert modular cusp forms of weight $\vec{k}$ and level $\mathfrak{n}$, denoted $S_{\vec{k}}(U_1(\mathfrak{n}))$, is a finite-dimensional complex vector space equipped with an action of commuting Hecke operators $T_\mathfrak{m}$, indexed by the non-zero ideals $\mathfrak{m}$ of $\mathcal{O}$ (to fix ideas, let us normalise our spaces and Hecke operators as in [59]). Fix once and for all embeddings $\overline{\mathbf{Q}} \to \mathbf{C}$ and $\overline{\mathbf{Q}} \to \overline{\mathbf{Q}}_\ell$, and let $0 \neq f \in S_{\vec{k}}(U_1(\mathfrak{n}))$ be an eigenform for all the $T_\mathfrak{m}$. A construction of Rogawski-Tunnell, Ohta and Carayol [53, 49, 10], completed by Taylor and Jarvis [59, 36], associates to f an ℓ-adic representation

$$\rho_f : G_K \to \mathrm{GL}_2(\overline{\mathbf{Q}}_\ell)$$

such that if $\mathfrak{p}$ is a prime of $\mathcal{O}$ not dividing $\ell\mathfrak{n}$, then ρ_f is unramified at $\mathfrak{p}$ and, if $\mathsf{Frob}_\mathfrak{p}$ denotes a geometric Frobenius, then $\mathrm{tr}\rho_f(\mathsf{Frob}_\mathfrak{p})$ is the eigenvalue of $T_\mathfrak{p}$ on f (note that Taylor does not need to specify whether his Frobenius elements are arithmetic or geometric, so we shall assume that they are geometric). Fixing an identification of the residue field of $\overline{\mathbf{Q}}_\ell$ with $\overline{\mathbf{F}}_\ell$, we obtain a representation

$$\overline{\rho}_f : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$$

defined as the semisimplification of the reduction of ρ_f. It is natural to expect the following "folklore" generalisation of Serre's weak conjecture to hold:

Conjecture 1.1. *Suppose that $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is continuous, irreducible and totally odd. Then ρ is isomorphic to $\overline{\rho}_f$ for some Hilbert modular eigenform f.*

Note that one could instead have defined ρ_f to be the representation with the property that the trace of an *arithmetic* Frobenius was equal to the corresponding

Hecke eigenvalue, which is the same as replacing ρ_f by its dual, but the "geometric" conjecture above is trivially equivalent to the "arithmetic" version (ρ is geometrically modular if and only if its dual is arithmetically modular). Although the Khare-Wintenberger approach to Serre's original conjecture may shed light on Conjecture 1.1 for a few explicit totally real fields K, it seems not (in its present form) to be able to attack the case of a general K because it relies on an induction and the fact that for certain small primes ℓ there are no 2-dimensional irreducible odd mod ℓ representations of $G_{\mathbf{Q}}$ unramified outside ℓ; however for a totally real field the analogous fact is not in general true.

The main aim of this paper is to refine Conjecture 1.1 along the lines of Serre's refinement for the case $K = \mathbf{Q}$, in the special case where the prime ℓ is unramified in K. Perhaps surprisingly, this is not as simple as it sounds. The main difficulty is in specifying the weight where, even in this unramified situation, we encounter several subtleties not present in Serre's original work. Note first of all that there is no obvious notion of a minimal weight. Moreover the possible weights and level structures at primes over ℓ are intertwined, and, contrary to the case $K = \mathbf{Q}$, one does *not* always expect a representation as in Conjecture 1.1 to arise from a classical Hilbert modular form of level prime to ℓ. Indeed, the mod ℓ representation attached to such a form has determinant equal to the product of a finite order character unramified at ℓ and some power of the mod ℓ cyclotomic character, and it is not hard to construct a mod ℓ Galois representation whose determinant is not of this form. To deal with these issues, we introduce the notion of a *Serre weight*, namely an irreducible $\overline{\mathbf{F}}_\ell$-representation σ of $\mathrm{GL}_2(\mathcal{O}/\ell)$, and define what it means for ρ to be modular of weight σ. Such a notion of weight is implicit in work of Ash and Stevens [3, 4], its relation to Serre's conjecture underlies Khare's paper [41], and its role in generalizing the conjecture to GL_n over $\mathbf{Q}$ is evident in [1, 2]. Our aim is to describe all possible Serre weights of forms giving rise to a representation ρ.

When working with classical modular forms, certain choices for normalisations and conventions have now become standard. In the Hilbert case there are various possibilities for these choices, and experience has shown the authors that things become ultimately less confusing if one works with holomorphic automorphic representations as in [10] rather than Hilbert modular forms, the advantage of this approach being that now the *only* choice one has to make is the normalisation of the local Langlands correspondence. We follow Carayol in our approach and use Hecke's normalisation rather than Langlands'. We explain our conventions more carefully later on.

Our weight conjecture (Conjecture 3.12 below) then takes the form of a recipe for the set of weights σ for which ρ is modular. Our notion of modularity is formulated in terms of ρ arising in the Jacobian (or equivalently cohomology) of Shimura curves associated to quaternion algebras over K; the weight recipe is given in terms of the local behaviour of ρ at primes over ℓ. We show (Theorem 3.15) that Conjecture 3.12 can be deduced from known results on Serre's Conjecture in the case $K = \mathbf{Q}$. It is supported for other fields by numerical evidence collected by one of the authors, Dembélé and Roberts [15], and by the theoretical evidence provided by the results of Gee, Jarvis and Schein discussed above.

Loosely speaking, Conjecture 1.1 can be thought of in the context of a "mod ℓ Langlands philosophy", with Conjecture 3.12 predicting a local-global compatibility at primes over ℓ. An interesting feature of the recipe for the weights is that if ℓ is inert in K and $\rho|_{G_{K_\ell}}$ is semisimple, then the set of weights that we associate to ρ are the Jordan–Hölder factors of the reduction of an *irreducible* $\overline{\mathbf{Q}}_\ell$-representation of $\mathrm{GL}_2(\mathcal{O}/\ell)$. This is proved in [17], where it is further shown that this association establishes a correspondence between 2-dimensional Galois representations of a local field *in its residue characteristic* and representations of GL_2 of its residue field *in characteristic zero.* Herzig [35] has shown that this phenomenon does not persist in the context of GL_n for $n > 2$, but rather is a property particular to $n \leq 2$ of a more general relation between the set of Serre weights and the reduction of a characteristic zero representation associated to ρ.

In [22, 23], Emerton made precise the sense in which Serre's refined conjecture could be viewed as part of a mod ℓ Langlands philosophy (in the case $k = \mathbf{Q}$). Using automorphic forms, he associates to ρ an $\overline{\mathbf{F}}_\ell$-representation $\pi(\rho)$ of $\mathrm{GL}_2(\mathbf{A}_\mathbf{Q})$, which is non-zero by the theorem of Khare and Wintenberger. Emerton conjectures, and shows under some technical hypotheses, that it factors as a restricted tensor product of local factors π_p, where π_p is a smooth admissible representation of $\mathrm{GL}_2(\mathbf{Q}_p)$ determined by $\rho|_{G_{\mathbf{Q}_p}}$. Serre's refined conjecture can then be recovered from properties of the π_p; moreover, results such as those in [18] and [41] describing the possible weights and levels of forms giving rise to ρ can also be extracted. We go on to formulate a conjecture in the spirit of Emerton's in the context of certain quaternion algebras over K. In order to do so, we need to associate a local factor to $\rho|_{G_{K_\mathfrak{p}}}$ when $\mathfrak{p}$ is a prime not dividing ℓ. This was already done by Emerton if the quaternion algebra is split at ℓ; we augment this with a treatment of the case where it is ramified using results of Vignéras [62].

This paper is structured as follows. In §2 we introduce the notion of a Serre weight and our notation and conventions regarding automorphic representations for GL_2 over K; we explain what it means for ρ to be modular of a given Serre weight, and relate this notion to the existence of automorphic representations π such that $\rho \sim \overline{\rho}_\pi$. In §3 we formulate Conjecture 3.12 giving a recipe for the set of Serre weights for which ρ is modular. Finally, in §4 we state our partial generalisation of Emerton's refined conjecture and derive some consequences.

Acknowledgements: Much of the research was carried out while one of the authors (F.D.) was at Brandeis University, with support from NSF grants #9996345, 0300434. K.B. was partially supported by an EPSRC Advanced Research Fellowship. We are grateful to R. Taylor for raising our interest in the questions considered here. The last section of the paper was greatly influenced by discussions at a workshop at the American Institute of Mathematics on p-adic Representations in February 2006 attended by K.B. and F.D.; we heartily thank AIM, and the workshop organisers and participants, especially M. Emerton. The influence of his ideas on that part of this work is apparent; we are also grateful to him for subsequent correspondence, in particular communicating details of [22, 23]. We also benefited from conversations and correspondence with C. Breuil, B. Conrad, L. Dembélé and T. Gee in the course of writing this paper. Finally the authors thank F. Herzig, M. Schein, R. Torrey for their helpful comments and corrections on an earlier version of the paper.

2. Serre weights

Suppose K is a totally real field (we allow $K = \mathbf{Q}$) and let $\mathcal{O}$ denote its ring of integers. Let ℓ be a prime, which we assume from the outset is unramified in K (although some of this section certainly could be made to work in more generality). Recall that we have fixed embeddings $\overline{\mathbf{Q}} \to \mathbf{C}$ and $\overline{\mathbf{Q}} \to \overline{\mathbf{Q}}_\ell$, and also an identification of $\overline{\mathbf{F}}_\ell$ with the residue field of $\overline{\mathbf{Q}}_\ell$. Let S_K denote the embeddings $K \to \mathbf{R}$ and let us fix once and for all a preferred embedding $\tau_0 : K \to \mathbf{R}$.

Consider the group

$$G = \mathrm{GL}_2(\mathcal{O}/\ell\mathcal{O}) \cong \prod_{\mathfrak{p}|\ell} \mathrm{GL}_2(\mathcal{O}/\mathfrak{p}).$$

A *Serre weight* is an isomorphism class of irreducible $\overline{\mathbf{F}}_\ell$-representations of G. These can be described explicitly as follows. For each prime $\mathfrak{p}$ of K dividing ℓ, set $k_\mathfrak{p} = \mathcal{O}/\mathfrak{p}$, $f_\mathfrak{p} = [k_\mathfrak{p} : \mathbf{F}_\ell]$ and let $S_\mathfrak{p}$ be the set of embeddings $\tau : k_\mathfrak{p} \to \overline{\mathbf{F}}_\ell$. Then every irreducible $\overline{\mathbf{F}}_\ell$-representation of $\mathrm{GL}_2(k_\mathfrak{p})$ is equivalent to one of the form

$$V_{\vec{a},\vec{b}} = \bigotimes_{\tau \in S_\mathfrak{p}} (\det{}^{a_\tau} \otimes_{k_\mathfrak{p}} \mathrm{Sym}^{b_\tau - 1} k_\mathfrak{p}^2) \otimes_\tau \overline{\mathbf{F}}_\ell,$$

where a_τ, $b_\tau \in \mathbf{Z}$ and $1 \le b_\tau \le \ell$ for each $\tau \in S_\mathfrak{p}$. Moreover we can assume that $0 \le a_\tau \le \ell - 1$ for each $\tau \in S_\mathfrak{p}$ and that $a_\tau < \ell - 1$ for some τ, in which case the resulting $(\ell^{f_\mathfrak{p}} - 1)\ell^{f_\mathfrak{p}}$ representations $V_{\vec{a},\vec{b}}$ are also inequivalent. The irreducible representations of G are thus of the form $V = \otimes_{\{\mathfrak{p}|\ell\}} V_\mathfrak{p}$, where the tensor product is over $\overline{\mathbf{F}}_\ell$ and each $V_\mathfrak{p}$ is of the form $V_{\vec{a},\vec{b}}$ for $(\vec{a}, \vec{b})$ as above.

If n is an integer then we let $\mathbf{F}_\ell(n)$ denote the 1-dimensional $\mathbf{F}_\ell$-vector space with left G-action defined by letting $g \in G$ act via $N(\det(g))^n$, where $N : (\mathcal{O}/\ell\mathcal{O})^\times \to (\mathbf{Z}/\ell\mathbf{Z})^\times$ is the norm. If $\mathbf{F}$ is a field of characteristic ℓ and V is an $\mathbf{F}$-representation of G space then we define $V(n)$ to be the $\mathbf{F}$-representation $V(n) := V \otimes_{\mathbf{F}_\ell} \mathbf{F}_\ell(n)$. Note that $V_{\vec{a},\vec{b}}(n) = V_{\vec{a}+n,\vec{b}}$.

Suppose that D is a quaternion algebra over K split at τ_0 and at no other infinite places. Fix an isomorphism $D \otimes_{K,\tau_0} \mathbf{R} = M_2(\mathbf{R})$; this induces an isomorphism of $(D \otimes_{\tau_0} \mathbf{R})^\times$ with $\mathrm{GL}_2(\mathbf{R})$, which acts on $\mathfrak{H}^\pm := \mathbf{C} \backslash \mathbf{R}$ in the usual way. Consider K as a subfield of $\mathbf{R}$ (and hence of $\mathbf{C}$) via the embedding τ_0. If $\mathbf{A}_K^f$ denotes the finite adeles of K and U is an open compact subgroup of $(D \otimes_K \mathbf{A}_K^f)^\times$ then there is a Shimura curve Y_U over K, a smooth algebraic curve whose complex points (via $\tau_0 : K \to \mathbf{C}$) are naturally identified with

$$D^\times \backslash \left((D \otimes_K \mathbf{A}_K^f)^\times \times \mathfrak{H}^\pm \right) / U$$

and such that Y_U is a canonical model for this space, in the sense of Deligne. These canonical models have the useful property that if U' is a normal compact open subgroup of U, then the natural right action of U/U' on $Y_{U'}(\mathbf{C})$ is induced by an action of U/U' on $Y_{U'}$ (that is, the action is defined over K).

Unfortunately there is more than one convention for these canonical models, and the choice that we make genuinely affects our normalisations. To fix ideas, we shall follow the conventions of Carayol in [9] and in particular our "Hodge structure" h will be that of section 0.1 of [9]. This corresponds to the choice $\epsilon = -1$ in the notation of [13]. See Section 3.3 of [13] for a discussion of the differences between this choice and the other natural choice—the key one being (Lemma 3.12 of [13]) that the choice *does* affect the Galois action on the adelic component group, by a sign. That this ambiguity exists is not surprising: for example in the elliptic curve case the modular curve $Y(\ell)$ parametrising elliptic curves equipped with generators of their ℓ-torsion exists (for $\ell > 2$) as a moduli space over $\mathbf{Q}$, and the Weil pairing gives a natural morphism $Y(\ell) \to \mathrm{Spec}(\mathbf{Q}(\zeta_\ell))$, but the two ways of normalising the Weil pairing give different morphisms.

If $K = \mathbf{Q}$ and D is split (we refer to this case as "the split case"), we let X_U denote the standard compactification of the modular curve Y_U; otherwise we simply set $X_U = Y_U$. Then X_U is a smooth projective algebraic curve over K. Note that X_U, considered as a scheme over K, will be connected (see section 1.3 of [9]) but not in general geometrically connected. Note also that, with notation as above, the natural action of U/U' on $Y_{U'}$ extends to an action on $X_{U'}$. Opting to include the split case does sometimes increase the length of a proof (we have to verify that "all errors are Eisenstein") but is arguably morally better than presenting proofs only in the non-split case and then merely asserting that they may be modified to deal with the split case too.

If U is a compact open subgroup of $(D \otimes_K \mathbf{A}_K^f)^\times$ as above, then let $\mathrm{Pic}^0(X_U)$ denote the identity component of the relative Picard scheme of $X_U \to \mathrm{Spec}(K)$. This definition is chosen specifically to deal with the fact that X_U may not be geometrically connected. In more concrete terms, if K_U denotes the ring $\Gamma(X_U, \mathcal{O}_{X_U})$ of globally-defined functions on X_U, then K_U is a number field and a finite abelian extension of K, the curve X_U is geometrically connected when regarded as a scheme over $\mathrm{Spec}(K_U)$, and $\mathrm{Pic}^0(X_U)$ is canonically isomorphic to the restriction of scalars (from K_U to K) of the Jacobian of X_U/K_U. In particular, $\mathrm{Pic}^0(X_U)$ is an abelian variety over K.

We henceforth assume that D is split at all primes of K above ℓ, and we fix an isomorphism $D \otimes_{\mathbf{Q}} \mathbf{Q}_\ell \cong M_2(K \otimes \mathbf{Q}_\ell)$. We can now regard $\mathrm{GL}_2(\mathcal{O} \otimes \mathbf{Z}_\ell)$ as a subgroup of $(D \otimes_K \mathbf{A}_K^f)^\times$. If U is a compact open subgroup of $(D \otimes_K \mathbf{A}_K^f)^\times$ as above, and if $\mathrm{GL}_2(\mathcal{O} \otimes \mathbf{Z}_\ell)$ is contained in U, then we say that U has *level prime to* ℓ. In this case, the natural map $U \to \mathrm{GL}_2(\mathcal{O}/\ell\mathcal{O}) = G$ is a surjection. Let U' denote its kernel. Then $U/U' = G$ acts naturally on the right on $Y_{U'}$ and on $X_{U'}$, and hence naturally on the left on $\mathrm{Pic}^0(X_{U'})$ and $\mathrm{Pic}^0(X_{U'})[\ell](\overline{K})$. Let us say that U is *sufficiently small* if it has level prime to ℓ and the map $Y_{U'} \to Y_U$ is étale of degree equal to the order of G. Note that any U of level prime to ℓ contains a compact open subgroup that is sufficiently small—this follows easily from 1.4.1.1–1.4.1.3 of [9] or Lemma 12.1 of [37]. The induced map $X_{U'} \to X_U$ will then be finite of degree equal to the order of G (but it may not be étale in the split case—there will usually be ramification at the cusps).

Definition 2.1. *Suppose that $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is a continuous, irreducible representation and V is a finite-dimensional $\overline{\mathbf{F}}_\ell$-vector space with a left action of G. We say that ρ is* modular of weight V *if there is a quaternion algebra D over K split at the primes above ℓ, at τ_0 and no other infinite places of K, and a sufficiently small open compact subgroup U of $(D \otimes_K \mathbf{A}_K^f)^\times$ of level prime to ℓ, such that ρ is an $\overline{\mathbf{F}}_\ell G_K$-subquotient of $\left(\mathrm{Pic}^0(X_{U'})[\ell](\overline{K}) \otimes V\right)^G$, where $U' = \ker(U \to G)$, G acts diagonally on the tensor product, and G_K acts trivially on V.*

Note that we allow subquotients with respect to the Galois action although using Hecke operators and the Eichler-Shimura relation on $X_{U'}$, one can show that if we replace "G_K-subquotient" by "G_K-submodule" then the resulting definition is equivalent. On the other hand, we really want to demand that ρ is a G_K-subquotient of the G-invariants of $\mathrm{Pic}^0(X_{U'})[\ell](\overline{K}) \otimes V$ rather than an $\overline{\mathbf{F}}_\ell[G_K \times G]$-subquotient on which G acts trivially. Our conjecture would not be correct were we to use G-subquotients; a general Galois representation would then be modular of more weights and we would not recover important subtleties of the refined conjecture.

We have allowed reducible G-representations W in the definition of modularity, for convenience; we now show that gives us no extra flexibility, in the sense that a continuous irreducible ρ is modular of weight W if and only if it is modular of weight V for V some Jordan-Hölder factor of W. The argument used to prove this involves a re-interpretation of the notion of being modular of weight W in terms of non-constant sheaves in the étale topology, which we now explain.

Say V is any finite-dimensional $\overline{\mathbf{F}}_\ell$-vector space equipped with a left G-action. The right action of G on $Y_{U'}$ enables us to identify G with a quotient of $\pi_1(Y_U, x)$ for x any geometric point of Y_U. Now a standard construction (see for example A I.7 of [26]) associates to V a locally constant étale sheaf $\mathcal{F}_V$ on Y_U, with (amongst other things) the property that the pullback of $\mathcal{F}_V$ to $Y_{U'}$ is just the constant sheaf associated to the vector space V. We abuse notation slightly by also using $\mathcal{F}_V$ to refer to the pullback of $\mathcal{F}_V$ to $Y_{U,\overline{K}}$, the base change of Y_U to $\overline{K}$.

Later on we will also need an ℓ-adic variant of this construction. Let $U \subset (D \otimes_K \mathbf{A}_K^f)^\times$ be compact and open, and let Y_U denote the associated Shimura curve over K. If $\vec{k} \in \mathbf{Z}_{\geq 2}^{S_K}$ and $w \in \mathbf{Z}$ with $w \equiv k_\tau \bmod 2$ for all τ, if $E \subset \overline{\mathbf{Q}}$ is a number field, Galois over $\mathbf{Q}$, splitting D and containing all embeddings of K into $\overline{\mathbf{Q}}$, if $\lambda | \ell$ is a prime of E, and U is sufficiently small in the sense of sections 2.1.3 and 2.1.4 of [10], then Carayol defines an $\mathcal{O}_{E_\lambda}$-sheaf $\mathcal{F}^0_{\lambda,U}$ on Y_U associated to $(\vec{k}, w)$ and an E_λ-sheaf $\mathcal{F}_{\lambda,U} = \mathcal{F}^0_{\lambda,U} \otimes \mathbf{Q}_\ell$. The sheaf $\mathcal{F}^0_{\lambda,U}$ (but not $\mathcal{F}_{\lambda,U}$) depends on a choice of lattice which we always take to be the one arising from tensor products of symmetric powers of $\mathcal{O}^2_{E_\lambda}$. We remark that our specific notion of *sufficiently small* is different from the one needed to define $\mathcal{F}^0_{\lambda,U}$ in [10], but with this choice of lattice any compact open U of level prime to ℓ contains one which is sufficiently small in both senses. Furthermore, if we demand that λ is the prime of E above ℓ induced by our embedding $E \to \overline{\mathbf{Q}} \to \overline{\mathbf{Q}}_\ell$ then there is an induced map $\mathcal{O}_{E_\lambda}/\lambda \to \overline{\mathbf{F}}_\ell$ and the induced $\overline{\mathbf{F}}_\ell$-sheaf $\mathcal{F}^0_{\lambda,U}/\lambda \otimes \overline{\mathbf{F}}_\ell$ is isomorphic to the sheaf $\mathcal{F}_W$ associated to the representation $W = \otimes_{\mathfrak{p}|\ell} \otimes_{\tau \in S_\mathfrak{p}} \det^{(w-k_\tau+2)/2} \mathrm{Symm}^{k_\tau-2}\, k_\mathfrak{p}^2 \otimes_\tau \overline{\mathbf{F}}_\ell$.

We begin by noting that Galois acts via an abelian quotient on many of the cohomology groups that show up in forthcoming arguments.

Lemma 2.2. *(a) Let U be any compact open subgroup of $(D \otimes_K \mathbf{A}_K^f)^\times$ and let $\mathcal{F}$ be a locally constant torsion sheaf on Y_U corresponding to a continuous representation of U/U' for some normal open compact $U' \subseteq U$ such that $Y_{U'} \to Y_U$ is étale with covering group U/U'. Then for $i \in \{0, 2\}$ the action of G_K on the cohomology groups $H^i(Y_{U,\overline{K}}, \mathcal{F})$ and $H^i_c(Y_{U,\overline{K}}, \mathcal{F})$ factors through an abelian quotient.*

(b) If $\mathcal{F}_{\lambda,U}$ is the sheaf associated to the data $(\vec{k}, w)$ as above, then the action of G_K on $H^1(Y_{U,\overline{K}}, \mathcal{F}_{\lambda,U})/H^1_p(Y_{U,\overline{K}}, \mathcal{F}_{\lambda,U})$ is via an abelian quotient.

(c) The action of G_K on the cokernel of the natural inclusion $H^1(X_{U,\overline{K}}, \mathbf{F}_\ell) \to H^1(Y_{U,\overline{K}}, \mathbf{F}_\ell)$ is via an abelian quotient.

Proof. (a) The pullback of $\mathcal{F}$ to $Y_{U'}$ is constant and $H^0(Y_{U,\overline{K}}, \mathcal{F})$ can be identified with a subspace of $H^0(Y_{U',\overline{K}}, \mathcal{F})$. The Galois action on this latter space is abelian, as the geometric components of $Y_{U'}$ are defined over an abelian extension of K. This proves the result for H^0, and the result for H^0_c follows as H^0_c is a subgroup of

H^0. For $i=2$ the result follows from the $i=0$ case and Poincaré duality, which pairs H^0 with H^2_c and H^2 with H^0_c.

(b) The quotient is trivial in the non-split case, so we are instantly reduced to the case $K=\mathbf{Q}$, $D=M_2(\mathbf{Q})$ and $E_\lambda=\mathbf{Q}_\ell$. In this case the result is surely well-known, but we sketch the proof for lack of a reference. The sheaf associated to the data (k,w) is $\mathbf{Q}_\ell$-dual to the sheaf associated to $(k,-w)$, so by Poincaré duality, this is equivalent to showing the Galois action factors through an abelian quotient on the kernel of the map (with w replaced by $2-w$). Recall that $H^1_c(Y_{U,\overline{\mathbf{Q}}},\mathcal{F}_{\lambda,U})=H^1(X_{U,\overline{\mathbf{Q}}},j_!\mathcal{F}_{\lambda,U})$ where $j:Y_U\to X_U$ is the natural inclusion, and that our map factors as

$$H^1(X_{U,\overline{\mathbf{Q}}},j_!\mathcal{F}_{\lambda,U})\to H^1(X_{U,\overline{\mathbf{Q}}},j_*\mathcal{F}_{\lambda,U})\to H^1(Y_{U,\overline{\mathbf{Q}}},\mathcal{F}_{\lambda,U}),$$

the first map being surjective, the second injective. Therefore it suffices to prove that the action of $G_{\mathbf{Q}}$ on

$$H^0(X_{U,\overline{\mathbf{Q}}},j_*\mathcal{F}_{\lambda,U}/j_!\mathcal{F}_{\lambda,U})=H^0(Z_{U,\overline{\mathbf{Q}}},i^*j_*\mathcal{F}_{\lambda,U})$$

factors through an abelian quotient, where $i:Z_U\to X_U$ is the reduced closed subscheme defined by the cusps. Shrinking U if necessary, we can assume there is a universal generalised elliptic curve over X_U (in the sense of [14]), and we let $s:E_U\to X_U$ denote its restriction to the open subscheme whose fibres over closed points are its identity components. Then we find that $j_*\mathcal{F}_{\lambda,U}$ is isomorphic to $\mathrm{Symm}^{k-2}(R^1s_*\mathbf{Q}_\ell)((w+k-2)/2)$, and since $E_U\times_{X_U}Z_U$ is isomorphic to $\mathbf{G}_{\mathrm{m},Z_U}$, we conclude that $i^*j_*\mathcal{F}_{\lambda,U}$ is isomorphic to $\mathbf{Q}_\ell((w-k+2)/2)$. The assertion now follows from the fact that the cusps are defined over an abelian extension of $\mathbf{Q}$.

(c) is similar to (b), but simpler. □

Let D, U, U', G and V be as in Definition 2.1. If ψ is a continuous character of G_K with values in $\overline{\mathbf{F}}_\ell^\times$ or $\overline{\mathbf{Q}}_\ell^\times$, then we let $\psi_{\mathbf{A}}$ denote the corresponding character of $\mathbf{A}_K^\times$ by class field theory (with uniformizers corresponding to geometric Frobenius elements).

Lemma 2.3. *Suppose that $\psi:G_K\to\overline{\mathbf{F}}_\ell^\times$ is a continuous character such that $\psi_{\mathbf{A}}$ is trivial on* $\det(U')$*, and let χ denote the restriction of $\psi_{\mathbf{A}}$ to $(\mathcal{O}\otimes\mathbf{Z}_\ell)^\times$. Then $H^1(Y_{U,\overline{K}},\mathcal{F}_{V\otimes\chi\circ\det})$ is isomorphic to $H^1(Y_{U,\overline{K}},\mathcal{F}_V)(\psi)$ as G_K-modules.*

Proof. The restriction of ψ to G_{K_U} corresponds to χ via the isomorphism

$$G_{K_U}/G_{K_{U'}}\cong\det(U)/\det(U')\cong(\mathcal{O}/\ell)^\times.$$

Recall from section 1.1.2 of [10] that we have a commutative diagram

$$\begin{array}{ccc} Y_{U'} & \to & \operatorname{Spec}K_{U'} \\ \downarrow & & \downarrow \\ Y_U & \to & \operatorname{Spec}K_U \end{array}$$

such that the action of $G=U/U'$ on $Y_{U'}$ is compatible via det with that of $\det(U)/\det(U')$ on $\operatorname{Spec}K_{U'}$. So if we let $\mathcal{F}_\chi$ denote the sheaf on $\operatorname{Spec}K_U$ corresponding to the character χ, we see that $\mathcal{F}_{\chi\circ\det}$ is isomorphic to the pull-back of $\mathcal{F}_\chi$ by the map $Y_U\to\operatorname{Spec}K_U$. As this map induces a bijection on geometric components, it follows that $H^0(Y_{U,\overline{K}},\mathcal{F}_{\chi\circ\det})$ is isomorphic to $H^0(\operatorname{Spec}K_U\times_K\overline{K},\mathcal{F}_\chi)$ as G_K-modules, which in turn is isomorphic to $\mathrm{Ind}_{G_{K_U}}^{G_K}\overline{\mathbf{F}}_\ell(\chi)$. Therefore $\mathrm{Hom}_{\overline{\mathbf{F}}_\ell[G_K]}(\overline{\mathbf{F}}_\ell(\psi),H^0(Y_{U,\overline{K}},\mathcal{F}_{\chi\circ\det}))$ is one-dimensional. Let α be the image in

$H^0(Y_{U,\overline{K}}, \mathcal{F}_{\chi\circ\det})$ of a non-trivial element. Note that the restriction of α to each component of $Y_{U,\overline{K}}$ is non-trivial, so cupping with α defines an isomorphism

$$H^1(Y_{U,\overline{K}}, \mathcal{F}_V) \otimes_{\overline{\mathbf{F}}_\ell} \overline{\mathbf{F}}_\ell\alpha \to H^1(Y_{U,\overline{K}}, \mathcal{F}_{V\otimes\chi\circ\det}).$$

□

Lemma 2.4. *Let $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ be continuous, irreducible and totally odd.*

(a) ρ is a G_K-subquotient of $\left(\mathrm{Pic}^0(X_{U'})[\ell](\overline{K}) \otimes V\right)^G$ if and only if ρ is a G_K-subquotient of $H^1(Y_{U,\overline{K}}, \mathcal{F}_{V(1)})$.

(b) ρ is modular of weight W (an arbitrary finite-dimensional $\overline{\mathbf{F}}_\ell[G]$-module) if and only if ρ is modular of weight V for some Jordan-Hölder factor of W.

Proof. (a) First recall that $\mathrm{Pic}^0(X_{U'})[\ell](\overline{K}) = H^1(X_{U',\overline{K}}, \mathbf{F}_\ell) \otimes \mu_\ell$ as $G_K \times G$-modules. By Lemma 2.2(c), the action of G_K on the cokernel of the natural injection

$$H^1(X_{U',\overline{K}}, \mathbf{F}_\ell) \to H^1(Y_{U',\overline{K}}, \mathbf{F}_\ell)$$

factors through an abelian quotient. It follows that ρ is modular of weight V if and only if ρ is an $\overline{\mathbf{F}}_\ell[G_K]$-subquotient of $(H^1(Y_{U',\overline{K}}, \mathbf{F}_\ell) \otimes V)^G \otimes \mu_\ell$.

The Hochschild-Serre spectral sequence now gives us an exact sequence of G_K-modules

$$H^1(G, H^0(Y_{U',\overline{K}}, \mathbf{F}_\ell) \otimes V) \to H^1(Y_{U,\overline{K}}, \mathcal{F}_V) \to$$
$$\to (H^1(Y_{U',\overline{K}}, \mathbf{F}_\ell) \otimes V)^G \to H^2(G, H^0(Y_{U',\overline{K}}, \mathbf{F}_\ell) \otimes V).$$

By Lemma 2.2(a), the action of G_K on the first and last terms factors through an abelian quotient. It follows that ρ is modular of weight V if and only if ρ is an $\overline{\mathbf{F}}_\ell[G_K]$-subquotient of $H^1(Y_{U,\overline{K}}, \mathcal{F}_V) \otimes \mu_\ell$. Finally note that by Lemma 2.3, we have

$$H^1(Y_{U,\overline{K}}, \mathcal{F}_{V(1)}) \cong H^1(Y_{U,\overline{K}}, \mathcal{F}_V) \otimes \mu_\ell,$$

and part (a) of the lemma follows.

(b) If $0 \to W_1 \to W_2 \to W_3 \to 0$ is a short exact sequence of finite-dimensional $\overline{\mathbf{F}}_\ell[G]$-modules then $0 \to \mathcal{F}_{W_1} \to \mathcal{F}_{W_2} \to \mathcal{F}_{W_3} \to 0$ is a short exact sequence of étale sheaves on Y_U, and (b) now follows from (a) and Lemma 2.2(a).

□

Our chosen embeddings $\overline{\mathbf{Q}} \to \mathbf{C}$ and $\overline{\mathbf{Q}} \to \overline{\mathbf{Q}}_\ell$ and identification of the residue field of $\overline{\mathbf{Z}}_\ell$ with $\overline{\mathbf{F}}_\ell$ allow us to identify S_K with $\bigcup_{\mathfrak{p}|\ell} S_\mathfrak{p}$. We now recall how this notion of modularity is related to the existence of automorphic representations for GL_2/K giving rise to ρ. We start by establishing some conventions.

When associating Galois representations to automorphic representations we shall follow Carayol's conventions in [10]. In particular our normalisations of local and global class field theory will send geometric Frobenius elements to uniformisers and our local-global compatibility will be Hecke's rather than Langlands' (the one that preserves fields of definition rather than the one that behaves well under all functorialities; the difference is a dual and a twist). We summarise Carayol's theorem (in fact we only need a weaker form which is essentially due to Langlands and Ohta), and its strengthening by Taylor and Jarvis.

For $k \geq 2$ and w integers of the same parity let $D_{k,w}$ denote the discrete series representation $D_{k,w}$ of $\mathrm{GL}_2(\mathbf{R})$ with central character $t \mapsto t^{-w}$ defined in section 0.2 of [10]. For $k = 1$ and w an odd integer we define $D_{1,w}$ to be the (non-unitary, irreducible) principal series $\mathrm{Ind}(\mu,\nu)$ where the induction is unitary induction and μ and ν are the (quasi-)characters of $\mathbf{R}^\times$ defined by $\mu(t) = |t|^{-w/2}\,\mathrm{sgn}(t)$ and $\nu(t) = |t|^{-w/2}$. Now for $\vec{k} \in \mathbf{Z}^{S_K}$ with each $k_\tau \geq 1$ and of the same parity, and $w \in \mathbf{Z}$ of the same parity as the k_τ, let us say that a cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbf{A}_K)$ is *holomorphic of weight* $(\vec{k}, w)$ (or "has weight $(\vec{k}, w)$" for short), if $\pi_\tau \cong D_{k_\tau,w}$ for each $\tau \in S_K$ (of course many cuspidal automorphic representations will not have any weight—we are picking out the ones that correspond to certain holomorphic Hilbert modular forms). The theorem of Eichler, Shimura, Deligne, Deligne–Serre, Langlands, Ohta, Carayol, Taylor, Blasius–Rogawski, Rogawski–Tunnell and Jarvis associates a Galois representation $\rho_\pi : G_K \to \mathrm{GL}_2(\overline{\mathbf{Q}}_\ell)$ to a cuspidal automorphic representation π for GL_2/K of weight $(\vec{k}, w)$, and the correspondence $\pi \mapsto \rho_\pi$ satisfies Hecke's local-global compatibility at the finite places of K of characteristic not dividing ℓ for which π is unramified. We remark that local-global compatibility at the ramified places away from ℓ does not quite appear to be known in this generality, although Carayol and Taylor establish it if $k_\tau \geq 2$ for all τ. See Theorem 7.2 of [36] for the current state of play if $k_\tau = 1$ for some τ.

From this compatibility one deduces easily (see section 3 of [10] for example) that if π is holomorphic of weight $(\vec{k}, w)$ then the determinant of ρ_π is the product of a finite order character and ω^{w-1}, where ω denotes the cyclotomic character. Note that twisting by an appropriate power of the norm character gives bijections between the automorphic representations of weight $(\vec{k}, w)$ and $(\vec{k}, w + 2n)$ for any integer n; this corresponds to twisting by an appropriate power of the cyclotomic character on the Galois side.

Let $\overline{\rho}_\pi : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ denote the semisimplification of the reduction of ρ_π. Our goal now in this section is to relate the two notions of being modular of some weight in the sense introduced above, and being modular in the sense of being isomorphic to $\overline{\rho}_\pi$ for some holomorphic π.

Proposition 2.5. *Let $(\vec{k}, w) \in \mathbf{Z}_{\geq 2}^{S_K} \times \mathbf{Z}$ be integers all of the same parity. Suppose $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is continuous, irreducible and totally odd. Then $\rho \sim \overline{\rho}_\pi$ for some cuspidal automorphic representation π for GL_2/K of weight $(\vec{k}, w)$ and level prime to ℓ if and only if ρ is modular of weight V for some Jordan-Hölder constituent V of*

$$V_{\vec{k},w} := \bigotimes_{\mathfrak{p}|\ell} \bigotimes_{\tau \in S_\mathfrak{p}} \det{}^{(w-k_\tau)/2}\mathrm{Sym}^{k_\tau-2} k_\mathfrak{p}^2 \otimes_\tau \overline{\mathbf{F}}_\ell.$$

Remark 2.6. Note that the representation $V_{\vec{k},w}$ above differs from the representation ψ in section 2.1.1 of [10] by a twist due to the fact that we are using Jacobians rather than étale cohomology.

Proof. Say V is a Jordan-Hölder constituent of $V_{\vec{k},w}$, and that ρ is modular of weight V. Then, by definition, there is a quaternion algebra D/K satisfying the usual conditions, and a level structure U prime to ℓ such that ρ is a subquotient of $(\mathrm{Pic}^0(X_{U'})[\ell] \otimes V)^G$ (with notation as above). By Lemma 2.4(a) and (b) and the remarks before Lemma 2.2, ρ is a G_K-subquotient of $H^1(Y_{U,\overline{K}}, \mathcal{F}^0_{\lambda,U}/\lambda) \otimes \overline{\mathbf{F}}_\ell$, where $\mathcal{F}^0_{\lambda,U}$ is the sheaf that Carayol associates to $(\vec{k}, w)$. Recall that by results of Jacquet-Langlands and Carayol [10], $H^1_p(Y_{U,\overline{K}}, \mathcal{F}_{\lambda,U})$ is a direct sum of irreducible 2-dimensional ℓ-adic representations ρ_π for π as in the statement of the proposition.

So by Lemma 2.2 and a standard cohomological argument, we see that $\rho \cong \overline{\rho}_\pi$ for some cuspidal automorphic representation π on $D^\times$ and hence for some cuspidal automorphic representation on GL_2/K.

The reverse implication is not quite so straightforward because given π one needs to see $\overline{\rho}_\pi$ in the cohomology of a Shimura curve (the problem being that K might have even degree and π might be principal series at all finite places). However, arguments of Wiles and Taylor show that this problem is surmountable via "level-raising." Indeed, Theorem 1 of [59] and the remarks following it show that $\overline{\rho}_\pi \cong \overline{\rho}_{\pi'}$ for π' an automorphic representation of GL_2/K that is special at a finite place, and the Galois representation associated to π' does indeed show up in the cohomology of a Shimura curve by the above-mentioned results of Jacquet-Langlands and Carayol. □

Corollary 2.7. *If $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is continuous, absolutely odd, irreducible, and $\rho \cong \overline{\rho}_\pi$ for some automorphic representation π of GL_2/K of level prime to ℓ and weight $(\vec{k}, w) \in \mathbf{Z}_{\geq 1}^{S_K} \times \mathbf{Z}$, then ρ is modular of weight V for some weight V.*

Proof. If $k_\tau \geq 2$ for all τ then this follows immediately from the previous proposition. If $k_\tau = 1$ for one or more τ then it suffices to construct an automorphic representation π' of level prime to ℓ with $\overline{\rho}_\pi \cong \overline{\rho}_{\pi'}$, and with π' of weight $(\vec{k}', w)$ with $k'_\tau \geq 2$ for all τ. This is done in [36] (via multiplication by an appropriate modular form congruent to 1 mod ℓ: in particular the result follows from the Deligne–Serre lemma and Lemma 5.2 of *loc. cit.*). □

We can furthermore predict the local behaviour at primes over ℓ of the automorphic representations of weight $(\vec{k}, w)$ giving rise to ρ. Before we start on this, here is a simple lemma that will be of use to us later.

Lemma 2.8. *If k is a finite field of characteristic ℓ and if V is an irreducible $\overline{\mathbf{F}}_\ell$-representation of $\mathrm{GL}_2(k)$ then there is an irreducible $\overline{\mathbf{Q}}_\ell$-representation of $\mathrm{GL}_2(k)$ whose reduction has V as a Jordan-Hölder factor. Furthermore there is a 1-dimensional $\overline{\mathbf{F}}_\ell$-representation χ of $\mathrm{GL}_2(k)$ and an irreducible $\overline{\mathbf{Q}}_\ell$-representation of $\mathrm{GL}_2(k)$ with a fixed vector for the subgroup $\left(\begin{smallmatrix} * & * \\ 0 & 1 \end{smallmatrix}\right)$ whose reduction has $\chi \otimes V$ as a Jordan-Hölder factor.*

Proof. For 1-dimensional V the result is clear (use the Teichmüller lift) and for V of dimension equal to the size of k, the Steinberg representation does the job. For other V the lemma follows immediately from Proposition 1.1 of [17] (with $J = S$ in the notation of that paper). □

We now introduce the following rather naive version of a type. If L is a finite extension of $\mathbf{Q}_\ell$, with integers $\mathcal{O}_L$, if π is a smooth irreducible complex representation of $\mathrm{GL}_2(L)$ and if σ is a smooth irreducible representation of $\mathrm{GL}_2(\mathcal{O}_L)$ (so σ is finite-dimensional and its kernel contains an open subgroup of $\mathrm{GL}_2(\mathcal{O}_L)$), then we say that π is *of type* σ if the restriction of π to a representation of $\mathrm{GL}_2(\mathcal{O}_L)$ contains a subspace isomorphic to σ (note that this is a much weaker and simpler version of the usual notion of a type).

We now need a mild refinement of a level-raising result of Richard Taylor; unfortunately this refinement does not appear to be in the literature, so we sketch a proof here. The reader who wants to follow the details is advised to have a copy of Taylor's paper [59] handy.

Lemma 2.9. *Suppose that $[K:\mathbf{Q}]$ is even and π is a cuspidal automorphic representation of GL_2/K which has weight $(\vec{k},w)$. Suppose moreover that $k_\tau \geq 2$ for all infinite places τ. For all primes $\mathfrak{p} \mid \ell$ choose a smooth irreducible representation $\sigma_\mathfrak{p}$ of $\mathrm{GL}_2(\mathcal{O}_{K_\mathfrak{p}})$ such that $\pi_\mathfrak{p}$ is of type $\sigma_\mathfrak{p}$.*

Then there is a prime $\mathfrak{q}$ of K not dividing ℓ and a cuspidal automorphic representation π' of GL_2/K, also of weight $(\vec{k},w)$, and such that

- $\overline{\rho}_\pi \sim \overline{\rho}_{\pi'}$*;*
- $\pi'_\mathfrak{q}$ *is an unramified special representation;*
- $\pi'_\mathfrak{p}$ *is of type* $\sigma_\mathfrak{p}$ *each* $\mathfrak{p} \mid \ell$.

Proof. (sketch). The main idea of the argument is contained in the proof of Theorem 1 of [59], but it would be a little disingenuous to cite this result without further comment because, as written, the proof does not keep track of types. It is not difficult to change the argument so that it does, however. We indicate what needs to be changed in order to prove the result we want. Because of our assumptions about the weight of π and the degree of K, the lemma can be deduced from a purely combinatorial statement about automorphic forms for the group of units of the totally definite quaternion algebra D over K of discriminant 1. We need to check that the system of eigenvalues associated to π_D (the transfer of π to $D^\times$) is congruent to the system of eigenvalues associated to an automorphic representation π'_D which is Steinberg at some place $\mathfrak{q}$ (this much is done in [59]) and furthermore such that $\pi'_{D,\mathfrak{p}}$ has type $\sigma_\mathfrak{p}$ at all places above ℓ. We do this by mimicking Taylor's argument with the following changes. Instead of working at level $\Gamma_1(n)$ as in [59], we work with a more general compact open level structure U, assumed for simplicity to be a product of local factors $U_\mathfrak{p} \subset \mathrm{GL}_2(K_\mathfrak{p})$, for $\mathfrak{p}$ running over the finite places of K. For $\mathfrak{p} \mid \ell$ we further assume that $U_\mathfrak{p}$ is a normal subgroup of $\mathrm{GL}_2(\mathcal{O}_\mathfrak{p})$, with $\mathcal{O}_\mathfrak{p}$ the integers in $K_\mathfrak{p}$ (all this can be achieved by shrinking U if necessary). We define G_ℓ to be the finite group $\prod_{\mathfrak{p}\mid\ell}(\mathrm{GL}_2(\mathcal{O}_\mathfrak{p})/U_\mathfrak{p})$; the group G_ℓ then acts on the space of automorphic forms of level U. Let $\sigma = \otimes_{\mathfrak{p}\mid\ell}\sigma_\mathfrak{p}$, so σ is a finite-dimensional smooth complex irreducible representation of G_ℓ, and fix a number field N which contains F, splits D, contains the trace of $\sigma(g)$ for all $g \in G_\ell$ and furthermore contains the values of the (algebraic) central character ψ of π_D. We define $S^D_{(\vec{k},w)}(U) = S^D_{(\vec{k},w)}(U;\mathbf{C})$, the space of weight $(\vec{k},w)$ automorphic forms of level U for $D^\times$, as in Taylor's paper (our $\vec{k}$ is his k and our w is his μ; strictly speaking Taylor only considers the case $w = \max\{k_\tau - 2\}$ but his arguments never assume this); then $S^D_{(\vec{k},w)}(U)$ is a finite-dimensional complex vector space. We define $S^D_{(\vec{k},w)}(U)_\psi$ to be the subspace of $S^D_{(\vec{k},w)}(U)$ where the centre Z of $D^\times(\mathbf{A}_K)$ acts via the character ψ. Note that for a fixed U and $(\vec{k},w)$ there are only finitely many characters ψ for which $S^D_{(\vec{k},w)}(U)_\psi$ is non-zero (because the infinity type of ψ is determined by w); the character ψ is an analogue of the Dirichlet character associated to a classical modular form. For R a subring of $\mathbf{C}$ containing the integers of N we define $S^D_{(\vec{k},w)}(U;R)$ as in Taylor's paper, and let $S^D_{(\vec{k},w)}(U;R)_\psi$ be $S^D_{(\vec{k},w)}(U;R) \cap S^D_{(\vec{k},w)}(U)_\psi$. These spaces all have an action of G_ℓ; we define $S^D_{(\vec{k},w)}(U)_{\sigma,\psi}$ to be the σ-eigenspace of $S^D_{(\vec{k},w)}(U)_\psi$ (that is, the $\mathbf{C}[G_\ell]$-direct summand of $S^D_{(\vec{k},w)}(U)_\psi$ cut out by the idempotent in $\mathbf{C}[G_\ell]$ corresponding to σ), and we define $S^D_{(\vec{k},w)}(U;R)_{\sigma,\psi}$ to be $S^D_{(\vec{k},w)}(U;R) \cap S^D_{(\vec{k},w)}(U)_{\sigma,\psi}$.

Note that the natural projection map $S^D_{(\vec{k},w)}(U) \to S^D_{(\vec{k},w)}(U)_{\sigma,\psi}$ typically does not extend to a projection on the integral level, but some positive integer multiple of it will be integral and defines a map $e : S^D_{(\vec{k},w)}(U;R) \to S^D_{(\vec{k},w)}(U;R)_{\sigma,\psi}$ such that the resulting map $S^D_{(\vec{k},w)}(U;R)_{\sigma,\psi} \subseteq S^D_{(\vec{k},w)}(U;R) \to S^D_{(\vec{k},w)}(U;R)_{\sigma,\psi}$ is multiplication by a positive integer which depends on the size of G_ℓ and the class group associated to $\det(U)$ but, crucially, is unchanged if one replaces U by the group $U \cap U_0(\mathfrak{q})$. Note in particular that e is not necessarily an idempotent, but $e^2 = Ce$ for some positive integer C.

Contrary to what Taylor implicitly asserts, there is in general no $\mathrm{SL}_2(R)$-invariant perfect pairing on $\mathrm{Symm}^a(R^2)$ if R is not a $\mathbf{Q}$-algebra. However if R is a subring of $\mathbf{C}$ then there is an $\mathrm{SL}_2(R)$-invariant injection from $\mathrm{Symm}^a(R^2)$ to its R-dual, with cokernel killed by some positive integer C' (which depends on a but not on R), and this induces a pairing on $\mathrm{Symm}^a(R^2)$ which is not perfect but which will suffice to prove the result we need (some of Taylor's constants need to be modified by this constant). Taylor uses this pairing to produce a perfect pairing on $S^D_{(\vec{k},w)}(U;\mathbf{C})$ and the analogue of this pairing that we shall need is the induced perfect pairing between $S^D_{(\vec{k},w)}(U;\mathbf{C})_{\sigma,\psi}$ and $S^D_{(\vec{k},w)}(U;\mathbf{C})_{\sigma^*,\psi^*}$, where $\sigma^* = \sigma.\chi \circ \det$ and $\psi^* = \psi.\chi \circ \det$, where χ is the finite order Hecke character associated to π_D on p.272 of [59] and det is the reduced norm $D^\times \to \mathrm{GL}_1$. The reason for this twist is that the SL_2-invariant pairing on the coefficient sheaves is not GL_2-invariant.

We now run through Taylor's argument on pp.272–276 of *loc. cit.*, making the following changes. If $S^D_k(U_1(n);R)$ occurs on the left hand side of a pairing, we replace it by $S^D_{(\vec{k},w)}(U;R)_{\sigma,\psi}$; if it occurs on the right hand side then we replace it by $S^D_{(\vec{k},w)}(U;R)_{\sigma^*,\psi^*}$. We replace $U_1(n;\mathfrak{q})$ with $U \cap \Gamma_0(\mathfrak{q})$ (with $\Gamma_0(\mathfrak{q})$ denoting the usual level structure, namely the matrices which are upper triangular modulo $\mathfrak{q}$), and replacing the Hecke algebra $\mathbf{T}^D_k(n)$ in [59] by the sub-$\mathbf{Z}$-algebra $\mathbf{T}_{(\vec{k},w)}(U)_{\sigma,\psi}$ of $\mathrm{End}_{\mathbf{C}}(S^D_{(\vec{k},w)}(U)_{\sigma,\psi})$ generated only by Hecke operators $T_\mathfrak{q}$ at the *unramified* primes $\mathfrak{q}$. These Hecke algebras are as big as we shall need for our application—we do not need to consider the operators $S_\mathfrak{q}$ as we have fixed a central character, and we also do not need to consider Hecke operators at the ramified places.

The analogues of the assertions about direct sums in Lemma 1 of [59] are still true on the (σ,ψ)-component of $S^D_{(\vec{k},w)}(U)$ (note that Taylor's map i commutes with the action of G_ℓ and with the action of Z (the centre of $D^\times(\mathbf{A}_K)$)), and the analogue of Lemma 2 also holds (indeed the proof given remains valid when $S^D_k(n)$ is replaced by $S^D_{(\vec{k},w)}(U)_{\sigma,\psi}$ etc., as the map $i^\dagger$ also commutes with the G_ℓ and Z-action). The analogue of Lemma 3 that we need is that for a fixed compact open subgroup $X \subset \mathrm{GL}_1(\mathbf{A}^f_K)$ there are positive integer constants C_1 and C_2 such that for any compact open U with $\det(U) = X$ we have

$$C_1\langle S^D_{(\vec{k},w)}(U;R)_{\sigma,\psi}, S^D_{(\vec{k},w)}(U;R)_{\sigma^*,\psi^*}\rangle \subseteq R$$

and for $f \in S^D_{(\vec{k},w)}(U;\mathbf{C})_{\sigma,\psi}$ with

$$\langle f, S^D_{(\vec{k},w)}(U;R)_{\sigma^*,\psi^*}\rangle \subseteq R$$

we have $C_2 f \in S^D_{(\vec{k},w)}(u;R)_{\sigma,\psi}$. The statement about C_1 follows from Taylor's Lemma 3, and that about C_2 can also be deduced from Taylor's result, the comments about C' above, the fact that the pairing $\langle,\rangle$ restricts to the zero pairing between the (σ,ψ)-eigenspace and the (σ',ψ')-eigenspace if $(\sigma',\psi')\neq(\sigma^*,\psi^*)$ (using the argument on the bottom of p.272 of [59]) and the existence of the "projector" e above.

We need the (σ^*,ψ^*)-analogue of Lemma 4 of [59] and this is true—indeed it can be deduced from Lemma 4 of [59] by restricting to the (σ^*,ψ^*)-eigenspace.

We can now prove the analogue of Theorem 1 of [59] (where we replace Taylor's Hecke algebras with ours as indicated above); we simply mimic Taylor's beautiful proof on p.276 of *loc. cit.*; the assiduous reader can check that we have explained the analogues of all the ingredients that we need. Now using a standard Cebotarev argument we deduce that given π_D of weight $(\vec{k},w)$, we can find a prime $\mathfrak{q}\nmid\ell$ of K at which π_D is unramified principal series and such that the associated system of eigenvalues of $\mathbf{T}_{(\vec{k},w)}(U)_{\sigma,\psi}$ is congruent (modulo some prime above ℓ) to a system of eigenvalues occurring in $\mathbf{T}_{(\vec{k},w)}(U_0(\mathfrak{q}))^{\mathrm{new}}_{\sigma,\psi}$. □

Using this beefed-up version of Taylor's level-raising theorem, we can deduce a beefed-up version of Proposition 2.5. Suppose that $\vec{k}\in\mathbf{Z}^{S_K}$ and $w\in\mathbf{Z}$ with $k_\tau\geq 2$ and of the same parity as w for all τ. For each $\mathfrak{p}|\ell$, suppose that $\sigma_\mathfrak{p}$ is a smooth irreducible representation of $\mathrm{GL}_2(\mathcal{O}_{K,\mathfrak{p}})$. Via our fixed embeddings $\overline{\mathbf{Q}}\to\mathbf{C}$ and $\overline{\mathbf{Q}}\to\overline{\mathbf{Q}}_\ell$, and our identification of the residue field of $\overline{\mathbf{Q}}_\ell$ with $\overline{\mathbf{F}}_\ell$, we can unambiguously define the semisimplification $\overline{\sigma}_\mathfrak{p}$ of the mod ℓ reduction of $\sigma_\mathfrak{p}$; so $\overline{\sigma}_\mathfrak{p}$ is a representation of $\mathrm{GL}_2(\mathcal{O}_{K,\mathfrak{p}})$ on a finite-dimensional $\overline{\mathbf{F}}_\ell$-vector space. Define $\sigma:=\otimes_{\mathfrak{p}|\ell}\sigma_\mathfrak{p}$ and $\overline{\sigma}:=\otimes_{\mathfrak{p}|\ell}\overline{\sigma}_\mathfrak{p}$. Finally let G_ℓ denote a finite quotient of $\mathrm{GL}_2(\mathcal{O}_K\otimes\mathbf{Z}_\ell)$ through which σ factors.

Proposition 2.10. *For an irreducible representation $\rho: G_K\to\mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$, the following are equivalent:*

- *$\rho\sim\overline{\rho}_\pi$ for some cuspidal holomorphic weight $(\vec{k},w)$ automorphic representation π of $\mathrm{GL}_2(\mathbf{A}_K)$ such that $\pi_\mathfrak{p}$ has type $\sigma_\mathfrak{p}$ for each $\mathfrak{p}|\ell$.*
- *ρ is modular of weight V for some Jordan-Hölder constituent V of*

$$W:=\overline{\sigma}^\vee\otimes\bigotimes_{\mathfrak{p}|\ell}\bigotimes_{\tau\in S_\mathfrak{p}}\left(\det{}^{(w-k_\tau)/2}\mathrm{Sym}^{k_\tau-2}k_\mathfrak{p}^2\otimes_\tau\overline{\mathbf{F}}_\ell\right).$$

Proof. Lemma 2.9 shows that $\rho\sim\overline{\rho}_\pi$ for some π as above if and only if there are D and U as usual, and a compact open subgroup U'' of U with $U/U''=G_\ell$, such that ρ is the mod ℓ reduction of an irreducible 2-dimensional $\overline{\mathbf{Q}}_\ell$-representation $\tilde{\rho}$ which is a direct summand (equivalently, a subquotient) of $\mathrm{Hom}_{G_\ell}(\sigma,H^1_p(Y_{U'',\overline{K}},\mathcal{F}_{\lambda,U''}))$, with $\mathcal{F}_{\lambda,U''}$ the E_λ-sheaf constructed by Carayol on $X_{U''}$ corresponding to $(\vec{k},w)$. Here we assume U is small enough for the projection $\phi: Y_{U''}\to Y_U$ to be étale with covering group G_ℓ. Furthermore our coefficient field E_λ is assumed large enough to satisfy the conditions that Carayol requires of it, and also large enough to afford a model for σ and to ensure that all the irreducible subquotients of $H^1(Y_{U'',\overline{K}},\mathcal{F}_{\lambda,U''})$ have dimension either 1 or 2. By Lemma 2.2(b) the latter condition is equivalent to ρ being isomorphic to the mod ℓ reduction of an irreducible 2-dimensional subquotient of $\mathrm{Hom}_{G_\ell}(\sigma,H^1(Y_{U'',\overline{K}},\mathcal{F}_{\lambda,U''}))$.

Now let let $\mathcal{F}_{\sigma^\vee}$ denote the E_λ-sheaf on Y_U associated to the dual of σ, and let $\mathcal{F}_{\lambda,U}$ denote the E_λ-sheaf constructed by Carayol on Y_U and corresponding to weight $(\vec{k}, w)$. Then

$$\begin{aligned}\mathrm{Hom}_{G_\ell}(\sigma, H^1(Y_{U'',\overline{K}}, \mathcal{F}_{\lambda,U''})) &= (H^1(Y_{U'',\overline{K}}, \mathcal{F}_{\lambda,U''}) \otimes \sigma^\vee)^{G_\ell} \\ &= H^1(Y_{U'',\overline{K}}, \mathcal{F}_{\lambda,U''} \otimes \phi^*\mathcal{F}_{\sigma^\vee})^{G_\ell} \\ &= H^1(Y_{U'',\overline{K}}, \phi^*(\mathcal{F}_{\lambda,U} \otimes \mathcal{F}_{\sigma^\vee}))^{G_\ell} \\ &= H^1(Y_{U,\overline{K}}, \mathcal{F}_{\lambda,U} \otimes \mathcal{F}_{\sigma^\vee}),\end{aligned}$$

the last equality coming, for example, from the Hochschild-Serre spectral sequence and the fact that the order of G_ℓ is invertible in the (characteristic zero) field E_λ.

Let $\mathcal{O}_\lambda$ denote the integers in E_λ, let $\mathcal{F}^0_{\lambda,U}$ and $\mathcal{F}^0_{\sigma^\vee}$ be $\mathcal{O}_\lambda$-lattices in $\mathcal{F}_{\lambda,U}$ and $\mathcal{F}_{\sigma^\vee}$ and set $\mathcal{F}^0 := \mathcal{F}^0_{\lambda,U} \otimes \mathcal{F}^0_{\sigma^\vee}$. We then deduce that an irreducible ρ is the reduction of an irreducible $\overline{\rho}_\pi$ as above if and only if it is a subquotient of $H^1(Y_{U,\overline{K}}, \mathcal{F}^0)^{tf}/\lambda$, where λ denotes the maximal ideal of $\mathcal{O}_\lambda$ and tf denotes the maximal torsion-free quotient. Now a standard argument shows that this is so if and only if ρ is a subquotient of $H^1(Y_{U,\overline{K}}, \mathcal{F}^0/\lambda)$ (the torsion in $H^1(Y_{U,\overline{K}}, \mathcal{F}^0)$ is a subquotient of $H^0(Y_{U,\overline{K}}, \mathcal{F}^0/\ell^N)$ for some N and hence the Galois representations arising as subquotients of it are all 1-dimensional by Lemma 2.2(a), and the cokernel of the injection $H^1(Y_{U,\overline{K}}, \mathcal{F}^0)/\lambda \to H^1(Y_{U,\overline{K}}, \mathcal{F}^0/\lambda)$ is contained in $H^2(Y_{U,\overline{K}}, \mathcal{F}^0)[\lambda]$ and hence in $H^2(Y_{U,\overline{K}}, \mathcal{F}^0/\lambda^N)$ for some N, and the irreducible subquotients of this group are also all 1-dimensional by Lemma 2.2(a)). To finish the proof it suffices by Lemma 2.4 to check that $\mathcal{F}^0/\lambda \otimes \overline{\mathbf{F}}_\ell$ and $W(1)$ have the same Jordan-Hölder factors, which follows immediately from the definitions. $\square$

For $\tau \in S_\mathfrak{p}$, we denote by ω_τ the fundamental character of $I_{K_\mathfrak{p}}$ defined by composing τ with the homomorphism $I_{K_\mathfrak{p}} \to k_\mathfrak{p}^\times$ obtained from local class field theory (with the convention that uniformizers correspond to geometric Frobenius elements). We then have the following compatibility among determinants, central characters and twists.

Corollary 2.11. (1) *If ρ is modular of weight V and V has central character $\otimes_{\mathfrak{p}|\ell} \prod_{\tau \in S_\mathfrak{p}} \tau^{c_\tau}$, then*

$$\det \rho|_{I_{K_\mathfrak{p}}} = \prod_{\tau \in S_\mathfrak{p}} \omega_\tau^{c_\tau + 1}$$

for each $\mathfrak{p}|\ell$.

(2) *Let $\chi : G_K \to \overline{\mathbf{F}}_\ell^\times$ be such that $\chi|_{I_{K_\mathfrak{p}}} = \prod_{\tau \in S_\mathfrak{p}} \omega_\tau^{c_\tau}$ for each $\mathfrak{p}|\ell$. Then ρ is modular of weight V if and only $\chi\rho$ is modular of weight $V \otimes V_\chi$, where*

$$V_\chi = \bigotimes_{\mathfrak{p}|\ell} \bigotimes_{\tau \in S_\mathfrak{p}} \det{}^{c_\tau} k_\mathfrak{p}^2 \otimes_\tau \overline{\mathbf{F}}_\ell.$$

Proof. (1) Let T denote the Teichmüller lift of the norm map $N : (\mathcal{O}/\ell)^\times \to \mathbf{F}_\ell^\times$. By Lemma 2.7 there is an irreducible $\overline{\mathbf{Q}}_\ell$-representation σ of G such that the reduction of $\sigma^\vee \otimes (T^{-1} \circ \det)$ contains V as a Jordan-Hölder factor. By Proposition 2.10 we see that if ρ is modular of weight V then $\rho \sim \overline{\rho}_\pi$ for some π of weight $(\vec{2}, 0)$ and type σ. Now by section 5.6.1 of [10] we see that $\det(\rho_\pi) = \chi_\pi^{-1}\omega^{-1}$ where χ_π is the central character of π and ω is the cyclotomic character. Now χ_π can be computed on $\mathcal{O}_{K_\mathfrak{p}}^\times$

because it is the central character of $\sigma_{\mathfrak{p}}$, which is the inverse of the central character $\alpha_{\mathfrak{p}}$ of $\sigma_{\mathfrak{p}}^{\vee}$. We deduce that $\det(\rho)|I_{\mathfrak{p}} = \alpha_{\mathfrak{p}}\overline{\omega}^{-1} = \alpha_{\mathfrak{p}} N_{\mathfrak{p}}^{-1}$, where $N_{\mathfrak{p}} = \prod_{\tau \in S_{\mathfrak{p}}}$ is the map $I_{K_{\mathfrak{p}}} \to k_{\mathfrak{p}}^{\times} \to \mathbf{F}_{\ell}$. Finally the fact that V is a Jordan-Hölder factor of the reduction of $\sigma^{\vee} \otimes (T^{-1} \circ \det)$ implies that the central character of V (considered as a representation of $\mathrm{GL}_2(k_{\mathfrak{p}})$) is $\alpha_{\mathfrak{p}}.N_{\mathfrak{p}}^{-2}$ and the result follows.

(2) is simpler and could have been deduced earlier; in fact it is immediate from Lemma 2.3 and Lemma 2.4(a). $\square$

Recall that for $K = \mathbf{Q}$, every modular ρ arises from a form of level prime to ℓ and some weight $k \geq 2$. Moreover, after twisting ρ, one may take the weight k to be in the range $2 \leq k \leq \ell+1$ (see [20]). This is in general false for larger K. Indeed if ρ arises from a form of weight $(\vec{k}, w)$ and level prime to ℓ, then $\det \rho|_{I_{\mathfrak{p}}} = \omega^{-1-w}$ for all $\mathfrak{p}|\ell$, and it is easy to construct representations none of whose twists have this property: choose for example an odd prime ℓ inert in a real quadratic K, and a totally odd ρ such that $\det \rho|_{I_{\ell}} = \omega_{\tau}^{a}$ for some *odd* integer a (where $\tau : \mathcal{O}_K/\ell \to \overline{\mathbf{F}}_{\ell}$); see [15] for some explicit examples. On the other hand, it is still the case that every modular ρ arises from a form of weight $(2, \ldots, 2)$ and some level not necessarily prime to ℓ. Moreover, after twisting ρ, we can assume the form has level dividing $\mathfrak{n}\ell$ for some $\mathfrak{n}$ prime to ℓ.

Corollary 2.12. *For an irreducible $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_{\ell})$ the following are equivalent:*

(1) *$\rho \sim \overline{\rho}_{\pi}$ for some holomorphic cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbf{A}_K)$;*
(2) *$\rho \sim \overline{\rho}_{\pi}$ for some cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbf{A}_K)$ of weight $(\vec{2}, 0)$;*
(3) *$\chi\rho \sim \overline{\rho}_{\pi}$ for some character χ and some cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbf{A}_K)$; of weight $(\vec{2}, 0)$ and level $U = U^{\ell}.U_1(\ell)$ (the adelic analogue of "level $\Gamma_1(\ell)$ at ℓ");*
(4) *ρ is modular of weight V for some Serre weight V.*

Proof. It is clear that (3) $\Rightarrow$ (2) $\Rightarrow$ (1). Proposition 2.10 shows that (1) $\Rightarrow$ (4) if $k_{\tau} \geq 2$ for all τ; if some of the k_{τ} are equal to 1 then one has to first multiply by an Eisenstein series to increase the weight $\vec{k}$: more formally one uses the Deligne–Serre lemma and Lemma 5.2 of [36]. Finally we need to show that (4) $\Rightarrow$ (3). If (4) holds then by Lemma 2.8 there is an irreducible $\overline{\mathbf{Q}}_{\ell}$-representation σ of G with a $\left(\begin{smallmatrix} * & * \\ 0 & 1 \end{smallmatrix}\right)$-fixed vector such that the reduction of $\sigma^{\vee} \otimes \bigotimes_{\mathfrak{p}|\ell} \otimes_{\tau \in S_{\mathfrak{p}}} \det^{-1}$ contains some twist of V. We deduce (3) from Lemma 2.11(2) and the case $(\vec{k}, w) = (\vec{2}, 0)$ of 2.10. $\square$

We remark that, in contrast to the $K = \mathbf{Q}$ case one does not appear to be able to control the tame level as much as one might conjecture, because of "unit group" obstructions that appear when choosing a global character with given local properties.

3. The weight conjecture

Suppose that $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_{\ell})$ is continuous, irreducible and totally odd. The aim of this section is to provide a conjectural recipe for the set of V such that ρ is modular of weight V.

For each prime $\mathfrak{p}$ of K dividing ℓ, we will define a set of representations $W_{\mathfrak{p}}(\rho)$ of $\mathrm{GL}_2(k_{\mathfrak{p}})$ depending only on $\rho|_{I_{K_{\mathfrak{p}}}}$, and then define the conjectural weight set $W(\rho)$ as the set of Serre weights of the form $\otimes_{\overline{\mathbf{F}}_\ell} V_{\mathfrak{p}}$ with $V_{\mathfrak{p}} \in W_{\mathfrak{p}}(\rho)$.

We need some more notation before defining $W_{\mathfrak{p}}(\rho)$. With our prime $\mathfrak{p}$ dividing ℓ fixed for now, we write simply k, f and S for $k_{\mathfrak{p}}$, $f_{\mathfrak{p}}$ and $S_{\mathfrak{p}}$. Fix an embedding $\overline{K} \to \overline{K}_{\mathfrak{p}}$ and identify $D = G_{K_{\mathfrak{p}}}$ and $I = I_{K_{\mathfrak{p}}}$ with subgroups of G_K. Let $K'_{\mathfrak{p}}$ be the unramified quadratic extension of $K_{\mathfrak{p}}$ in $\overline{K}_{\mathfrak{p}}$ and let k' denote its residue field. We let S' denote the set of embeddings $k' \to \overline{\mathbf{F}}_\ell$, let $D' = G_{K'_{\mathfrak{p}}}$ and define a map $\pi : S' \to S$ by $\tau' \mapsto \tau'|_k$.

Suppose that $L \subset \overline{K}_{\mathfrak{p}}$ is a finite unramified extension of $K_{\mathfrak{p}}$ and σ is an embedding of its residue field $\mathcal{O}_L/\ell\mathcal{O}_L$ in $\overline{\mathbf{F}}_\ell^\times$. We denote by ω_σ the fundamental character of $I = I_L$ defined by composing σ with the homomorphism $I_L \to (\mathcal{O}_L/\ell\mathcal{O}_L)^\times$ gotten from local class field theory.

In defining $W_{\mathfrak{p}}(\rho)$, we treat separately the cases where $\rho|_D$ is irreducible and where it is reducible.

3.1. **The irreducible case.** If $\rho|_D$ is irreducible, we define $W_{\mathfrak{p}}(\rho)$ by the following rule:

$$(1)\qquad V_{\vec{a},\vec{b}} \in W_{\mathfrak{p}}(\rho) \iff \begin{array}{c} \rho|_I \sim \prod_{\tau\in S} \omega_\tau^{a_\tau} \begin{pmatrix} \prod_{\tau'\in J} \omega_{\tau'}^{b_{\pi(\tau')}} & 0 \\ 0 & \prod_{\tau'\notin J} \omega_{\tau'}^{b_{\pi(\tau')}} \end{pmatrix} \\ \text{for some } J \subset S' \text{ such that } \pi : J \xrightarrow{\sim} S. \end{array}$$

Since $\rho|_D$ is irreducible, there is a character $\xi : D' \to \overline{\mathbf{F}}_\ell^\times$ such that $\rho|_D \sim \mathrm{Ind}_{D'}^D \xi$. We define

$$W'(\xi) = \{\, (V_{\vec{a},\vec{b}}, J) \mid J \subset S',\ \pi : J \xrightarrow{\sim} S,\ \xi|_I = \prod_{\tau\in S} \omega_\tau^{a_\tau} \prod_{\tau'\in J} \omega_{\tau'}^{b_{\pi(\tau')}} \,\}.$$

Thus $W_{\mathfrak{p}}(\rho) = \{\, V \mid (V, J) \in W'(\xi) \text{ for some } J \,\}$. (Note that replacing ξ by its conjugate under D/D' replaces J by its complement.) We shall see that the projection maps $W'(\xi) \to W_{\mathfrak{p}}(\rho)$ and $W'(\xi) \to \{\, J \subset S' \,|\pi : J \xrightarrow{\sim} S \,\}$ are typically bijections, so that $|W_{\mathfrak{p}}(\rho)| = 2^f$.

We now choose an element of S' which we denote τ'_0, and then let $\tau'_i = \tau'_0 \circ \mathsf{Frob}_\ell^i$ and $\tau_i = \pi(\tau'_i)$. Note that $S = \{\, \tau_i \mid i \in \mathbf{Z}/f\mathbf{Z} \,\}$ and $S' = \{\, \tau'_i \mid i \in \mathbf{Z}/2f\mathbf{Z} \,\}$. Letting $\omega = \omega_{\tau_0}$ and $\omega' = \omega_{\tau'_0}$, we have $\omega_{\tau_i} = \omega^{\ell^i}$, $\omega_{\tau'_i} = (\omega')^{\ell^i}$ and $\omega = (\omega')^{\ell^f+1}$. Note that $\xi|_I = (\omega')^n$ for some $n \bmod \ell^{2f} - 1$, and since $\rho|_D$ is irreducible, n is not divisible by $\ell^f + 1$.

For $B \subset \{0, \ldots, f-1\}$ (where the symbol $\subset$ includes the case of equality), let $J_B = \{\tau'_i | i \in B\} \cup \{\tau'_{f+i} | i \notin B\}$. If $a \in \mathbf{Z}/(\ell^f - 1)\mathbf{Z}$, $\vec{b} = (b_0, \ldots, b_{f-1})$ with each $b_i \in \{1, \ldots, \ell\}$ and $B \subset \{0, \ldots, f-1\}$, let

$$n'_{a,\vec{b},B} = a(\ell^f + 1) + \sum_{i\in B} b_i \ell^i + \sum_{i\notin B} b_i \ell^{f+i} \bmod \ell^{2f} - 1.$$

Then $W'(\xi)$ is in bijection with the set of triples $(a, \vec{b}, B)$ as above with $n \equiv n'_{a,\vec{b},B} \bmod \ell^{2f} - 1$. Now note for each $B \subset \{0, \ldots, f-1\}$, there is a unique such triple $(a, \vec{b}, B)$ with this property for each solution of

$$n \equiv \sum_{i \in B} b_i \ell^i - \sum_{i \notin B} b_i \ell^i \bmod \ell^f + 1$$

with $b_0, \dots, b_{f-1} \in \{1, \dots, \ell\}$. But the values of $\sum_{i \in B} b_i \ell^i - \sum_{i \notin B} b_i \ell^i$ are the ℓ^f consecutive integers from $n'_B - \ell^f$ to $n'_B - 1$ where

$$n'_B = \sum_{i \in B} \ell^{i+1} - \sum_{i \notin B} \ell^i + 1,$$

so there is a solution as long as $n \not\equiv n'_B \bmod \ell^f + 1$ and this solution is unique. We have thus shown that $W'(\xi)$ is in bijection with the set of B such $n \not\equiv n'_B \bmod \ell^f + 1$; moreover the projection $W'(\xi) \to \{ J \subset S' \,|\, \pi : J \xrightarrow{\sim} S \}$ is injective.

Note that if f is odd and B is either $\{0, 2, \dots, f-1\}$ or $B = \{1, 3 \dots, f-2\}$, then $n'_B \equiv 0 \bmod \ell^f + 1$. To see that the converse holds as well, observe that

$$-(\ell^f + 1) < -\ell \frac{\ell^{f-1} - 1}{\ell - 1} \leq n'_B \leq \frac{\ell^{f+1} - 1}{\ell - 1} < 2(\ell^f + 1).$$

Thus if $n'_B \equiv 0 \bmod \ell^f + 1$, then $n'_B = 0$ or $\ell^f + 1$. If $n'_B = 0$, then solving

$$\sum_{i \in B} \ell^{i+1} - \sum_{i \notin B} \ell^i + 1 \equiv 0 \bmod \ell^r$$

by induction on r, we find that f is odd and $B = \{1, 3, \dots, f-2\}$. Similarly if $n'_B = \ell^f + 1$, then f is odd and $B = \{0, 2, \dots, f-1\}$.

We now show that if f is even, then the 2^f values of $n'_B \bmod \ell^f + 1$ are distinct. First note that

$$n'_B \equiv -1 + (\ell + 1) \sum_{i \in B^*} (-1)^i \ell^i \bmod \ell^f + 1,$$

where $B^* = (\{0, 2, \dots, f-2\} \cap B) \cup (\{1, 3, \dots, f-1\} \setminus B)$. Thus $n'_{B_1} \equiv n'_{B_2} \bmod \ell^f + 1$ if and only if

$$\sum_{i \in B_1^*} (-1)^i \ell^i \equiv \sum_{i \in B_2^*} (-1)^i \ell^i \bmod (\ell^f + 1)/d,$$

where $d = \gcd(\ell + 1, \ell^f + 1) \leq \ell - 1$ (so $d = 2$ if ℓ is odd, and $d = 1$ if $\ell = 2$). But these two sums differ by at most $(\ell^f - 1)/(\ell - 1) < (\ell^f + 1)/d$, so the above congruence holds if and only if equality holds, in which case $B_1^* = B_2^*$, so $B_1 = B_2$.

Next we show that if f and ℓ are odd, then the $2^f - 2$ non-zero values of $n'_B \bmod \ell^f + 1$ are distinct. In this case we have

$$n'_B \equiv (\ell + 1) \sum_{i \in B^*} (-1)^i \ell^i \bmod \ell^f + 1,$$

where $B^* = (\{0, 2, \dots, f-1\} \cap B) \cup (\{1, 3, \dots, f-2\} \setminus B)$. Thus $n'_{B_1} \equiv n'_{B_2} \bmod \ell^f + 1$ if and only if

$$\sum_{i \in B_1^*} (-1)^i \ell^i \equiv \sum_{i \in B_2^*} (-1)^i \ell^i \bmod (\ell^f + 1)/(\ell + 1).$$

But these two sums differ by at most $(\ell^f - 1)/(\ell - 1) < 2(\ell^f + 1)/(\ell + 1)$, so if the above congruence holds then either equality holds, in which case $B_1 = B_2$, or

$$\sum_{i \in B_2^*} (-1)^i \ell^i = \sum_{i \in B_1^*} (-1)^i \ell^i + \sum_{i=0}^{f-1} (-1)^i \ell^i,$$

exchanging B_1 and B_2 if necessary. Solving modℓ^r inductively on r, we see that the only possibility is that $B_1^* = \emptyset$ and $B_2^* = \{0, 1, \dots, f-1\}$, but these are precisely the cases where $n'_B \equiv 0 \bmod \ell^f + 1$.

Finally suppose that f is odd and $\ell = 2$. In this case we have

$$n'_B \equiv 3 \sum_{i \in B^*} (-1)^i 2^i \bmod 2^f + 1,$$

where $B^* = (\{0, 2, \dots, f-1\} \cap B) \cup (\{1, 3, \dots, f-2\} \setminus B)$. In particular $n'_B \equiv 0 \bmod 3$. Moreover we have the inequality $|n'_{B_1} - n'_{B_2}| < 3(2^f + 1)$, showing that each congruence class mod $2^f + 1$ arises as n'_B for at most 3 values of B. It follows that each of the $(2^f - 2)/3$ non-zero multiples of $3 \bmod 2^f + 1$ arises as n'_B for exactly 3 values of B.

We have thus proved the following propositions:

Proposition 3.1. *Suppose that ℓ is odd. If f is even, then the congruence classes* mod $\ell^f + 1$ *of the form*

$$-1 + (\ell + 1) \sum_{i \in B^*} (-1)^i \ell^i$$

are distinct and non-zero as B^* runs through all subsets of $\{0, 1, \dots, f-1\}$. *If f is odd, then the congruence classes* mod $\ell^f + 1$ *of the form*

$$(\ell + 1) \sum_{i \in B^*} (-1)^i \ell^i$$

are distinct and non-zero as B^* runs through all non-empty proper subsets of $\{0, 1, \dots, f-1\}$. *Letting A denote the set of such classes in each case, we have*

$$|W'(\xi)| = \begin{cases} 2^f, & \text{if } n \notin A, \\ 2^f - 1, & \text{if } n \in A. \end{cases}$$

Proposition 3.2. *Suppose that $\ell = 2$ and $\xi | I = (\omega')^n$. Then*

$$|W'(\xi)| = \begin{cases} 2^f - 1, & \text{if } f \text{ is even,} \\ 2^f, & \text{if } f \text{ is odd and } 3 \nmid n, \\ 2^f - 3, & \text{if } f \text{ is odd and } 3 \mid n. \end{cases}$$

Multiple B can occur with the same $(a, \vec{b})$; for example if $f = 3$ and $n = 1$, then $(-\ell^2, (1, \ell, 1), \{0, 1\})$ and $(-\ell^2, (1, \ell, 1), \{0, 2\})$ are both in $W'(\xi)$, so the map $W'(\xi) \to W_{\mathfrak{p}}(\rho)$ is not injective. (In this case in fact, $|W'(\xi)| = 8$, but $|W_{\mathfrak{p}}(\rho)| = 6$.)

Proposition 3.3. *The map $W'(\xi) \to W_{\mathfrak{p}}(\rho)$ fails to be injective if and only if* $\ell^r n \equiv m \bmod \ell^f + 1$ *for some integers r, m with $|m| \le \ell(\ell^{f-2} - 1)/(\ell - 1)$.*

Proof. Suppose first that $W'(\xi) \to W_{\mathfrak{p}}(\rho)$ is not injective. This means that for some a, $\vec{b}$ and $B_1 \neq B_2$, we have

$$n \equiv n'_{a,\vec{b},B_1} \equiv n'_{a,\vec{b},B_2} \bmod \ell^{2f} - 1. \tag{2}$$

First we note that B_2 cannot be the complement $\overline{B}_1$ of B_1 in $\{0, \dots, f-1\}$ since

$$n'_{a,\vec{b},B_1} \equiv n'_{a,\vec{b},\overline{B}_1} \equiv \ell^f n'_{a,\vec{b},B_1} \bmod \ell^{2f} - 1$$

would imply $n \equiv 0 \bmod \ell^f + 1$, contradicting the irreducibility of $\rho|_D$. We thus have $J_2 \neq S' \setminus J_1$, where $J_1 = J_{B_1}$ and $J_2 = J_{B_2}$. One now checks that (possibly after switching J_1 and J_2) we may find $t \in \mathbf{Z}/2f\mathbf{Z}$ so that $\tau'_{t-1} \in J_1 \setminus J_2$ and $\tau'_t \in J_1 \cap J_2$. We then have

$$\ell^{f-t} n \equiv n'_{a',\vec{b}',B'_1} \equiv n'_{a',\vec{b}',B'_2} \bmod \ell^{2f} - 1,$$

where $a' \equiv \ell^{-t} a \bmod \ell^f - 1$, $b'_i = b_{i+t \bmod f}$ for $i \in \{0, \dots, f-1\}$ and B'_ν is such that $J_{B'_\nu} = J_\nu \circ \mathsf{Frob}_\ell^{f-t}$ for $\nu = 1, 2$. Replacing n with $\ell^{f-t} n$, we may thus assume that $\tau'_{f-1} \in J_1 \setminus J_2$ and $\tau'_f \in J_1 \cap J_2$, or equivalently, $f-1 \in B_1 \setminus B_2$ and $0 \notin B_1 \cup B_2$.

Returning to the congruence (2), we have

$$\sum_{i \in B_1} b_i \ell^i + \ell^f \sum_{i \notin B_1} b_i \ell^i \equiv \sum_{i \in B_2} b_i \ell^i + \ell^f \sum_{i \notin B_2} b_i \ell^i \bmod \ell^{2f} - 1,$$

or equivalently,

$$\sum_{i \in B_1 \setminus B_2} b_i \ell^i \equiv \sum_{i \in B_2 \setminus B_1} b_i \ell^i \bmod \ell^f + 1.$$

Since each sum is less than $2(\ell^f + 1)$, they must either be equal or differ by $\ell^f + 1$, and since $0 \notin B_1 \cup B_2$, each sum is divisible by ℓ, so in fact equality holds. Since $f - 1 \in B_1 \setminus B_2$, we have

$$b_{f-1} \ell^{f-1} \leq \sum_{i \in B_1 \setminus B_2} b_i \ell^i = \sum_{i \in B_2 \setminus B_1} b_i \ell^i \leq \sum_{i=2}^{f-1} \ell^i < 2\ell^{f-1},$$

so $b_{f-1} = 1$. Moreover if $b_{f-2}, b_{f-3}, \dots, b_{s+1}$ are all less than ℓ for some $s < f-2$, then we must have $b_{f-2} = b_{f-3} = \cdots = b_{s+1} = \ell - 1$ and $f-2, f-3 \dots, s \in B_2 \setminus B_1$, for if either fails, we find that

$$\sum_{i \in B_2 \setminus B_1} b_i \ell^i \leq (\ell - 1) \sum_{i=s+1}^{f-2} \ell^i + \sum_{i=2}^{s} \ell^i < \ell^{f-1}.$$

Since $0 \notin B_2 \setminus B_1$, we conclude that for some s with $0 < s < f - 1$, we have $(b_s, b_{s+1}, \dots, b_{f-1}) = (\ell, \ell - 1, \dots, \ell - 1, 1)$ and $s, s+1, \dots, f-2 \in B_2 \setminus B_1$. It follows that

$$n \equiv \sum_{i \in B_1} b_i \ell^i - \sum_{i \notin B_1} b_i \ell^i \equiv \sum_{i \in B_1, i < s} b_i \ell^i - \sum_{i \notin B_1, i < s} b_i \ell^i \bmod \ell^f + 1,$$

and this last difference has absolute value at most $\ell(\ell^{f-2} - 1)/(\ell - 1)$.

Conversely suppose that $\ell^r n \equiv m \bmod \ell^f + 1$ for some r, m with $|m| \leq \ell(\ell^{f-2} - 1)/(\ell - 1)$. Replacing r by $r + f$ if necessary, we may assume $m > 0$ and then

$$\frac{\ell^s - 1}{\ell - 1} \leq m \leq \ell \cdot \frac{\ell^s - 1}{\ell - 1}$$

for some s with $0 < s < f - 1$. We can then write $m = \sum_{i=0}^{s-1} b_i \ell^i$ for some $b_0, b_1, \dots, b_{s-1} \in \{1, \dots, \ell\}$. We can then choose $a \in \mathbf{Z}/(\ell^f - 1)\mathbf{Z}$ so that

$$\ell^r n \equiv n_{a,\vec{b},B_1} \equiv n_{a,\vec{b},B_2} \bmod \ell^{2f} - 1,$$

where $\vec{b} = (b_0, b_1, \dots, b_{s-1}, \ell, \ell - 1, \dots, \ell - 1, 1)$, $B_1 = \{0, 1, \dots, f-2\}$ and $B_2 = \{0, 1, \dots, s-1, f-1\}$. We conclude that

$$n \equiv n'_{a',\vec{b}',B'_1} \equiv n'_{a',\vec{b}',B'_2} \bmod \ell^{2f} - 1,$$

where $a' \equiv \ell^{-r} a \bmod \ell^f - 1$, $b'_i = b_{i+r \bmod f}$ for $i \in \{0, \dots, f-1\}$ and B'_ν is such that $J_{B'_\nu} = J_{B_\nu} \circ \mathsf{Frob}_\ell^{-r}$ for $\nu = 1, 2$. $\square$

3.2. **The reducible case.** Now suppose $\rho|_D$ is reducible, write $\rho|_D \sim \begin{pmatrix} \chi_1 & * \\ 0 & \chi_2 \end{pmatrix}$ and let c_ρ denote the corresponding class in $H^1(K_\mathfrak{p}, \chi_1\chi_2^{-1})$. Consider now the set of pairs

$$W'(\chi_1,\chi_2) = \{\, (V_{\vec{a},\vec{b}}, J) \mid J \subset S, \chi_1|_{I_{K_\mathfrak{p}}} = \prod_{\tau\in S} \omega_\tau^{a_\tau} \prod_{\tau\in J} \omega_\tau^{b_\tau}, \chi_2|_{I_{K_\mathfrak{p}}} = \prod_{\tau\in S} \omega_\tau^{a_\tau} \prod_{\tau\notin J} \omega_\tau^{b_\tau} \,\}.$$

Note that interchanging χ_1 and χ_2 replaces J by its complement. We shall see that the projection map $\pi_2 : W'(\chi_1,\chi_2) \to \{J \subset S\}$ is typically a bijection, so that $|W'(\chi_1,\chi_2)| = 2^f$. However $W_\mathfrak{p}(\rho)$ will be defined below as a subset of $\pi_1(W'(\chi_1,\chi_2))$ depending on c_ρ.

We now analyse the set $W'(\chi_1,\chi_2)$ in a manner analogous to the irreducible case. We write $\chi_\nu = \omega^{n_\nu}$ with $n_\nu \in \mathbf{Z}/(\ell^f-1)\mathbf{Z}$ for $\nu = 1,2$, and we let $n = n_1 - n_2$. If $a \in \mathbf{Z}/(\ell^f-1)\mathbf{Z}$, $\vec{b} = (b_0,\dots,b_{f-1})$ with each $b_i \in \{1,\dots,\ell\}$ and $B \subset \{0,\dots,f-1\}$, let

$$n_{a,\vec{b},B} = a + \sum_{i\in B} b_i\ell^i \bmod \ell^f - 1.$$

Then $W'(\chi_1,\chi_2)$ is in bijection with the set of triples $(a,\vec{b},B)$ as above with $n_1 \equiv n_{a,\vec{b},B} \bmod \ell^f - 1$ and $n_2 \equiv n_{a,\vec{b},\overline{B}} \bmod \ell^f - 1$ where $\overline{B}$ is the complement of B in $\{0,\dots,f-1\}$. Note that for each $B \subset \{0,\dots,f-1\}$ there is a unique such triple for each solution of

$$n \equiv \sum_{i\in B} b_i\ell^i - \sum_{i\notin B} b_i\ell^i \bmod \ell^f - 1.$$

with $b_0,\dots,b_{f-1} \in \{1,\dots,\ell\}$. But the values of $\sum_{i\in B} b_i\ell^i - \sum_{i\notin B} b_i\ell^i$ are the ℓ^f consecutive integers from $n_B + 1 - \ell^f$ to n_B where

$$n_B = \sum_{i\in B} \ell^{i+1} - \sum_{i\notin B} \ell^i,$$

so there is a unique solution if $n \not\equiv n_B \bmod \ell^f - 1$ and two solutions if $n \equiv n_B \bmod \ell^f - 1$. In particular the projection $W'(\chi_1,\chi_2) \to \{J \subset S\}$ is surjective and $|W'(\chi_1,\chi_2)| = 2^f + |\{\, B \mid n \equiv n_B \bmod \ell^f - 1\}|$.

We now show that if f is odd, then the 2^f values of $n_B \bmod \ell^f - 1$ are distinct, unless $\ell = 2$ or 3, in which case $n_{\{0,\dots,f-1\}} \equiv n_\emptyset \bmod \ell^f - 1$ and the rest are distinct. First note that

$$n_B \equiv -1 + (\ell+1) \sum_{i\in B^*} (-1)^i\ell^i \bmod \ell^f - 1,$$

where $B^* = (\{0,2,\dots,f-1\} \cap B) \cup (\{1,3,\dots,f-2\} \setminus B)$. Thus $n_{B_1} \equiv n_{B_2} \bmod \ell^f - 1$ if and only if

$$\sum_{i\in B_1^*} (-1)^i\ell^i \equiv \sum_{i\in B_2^*} (-1)^i\ell^i \bmod (\ell^f-1)/d,$$

where $d = \gcd(\ell+1, \ell^f-1)$. If $\ell > 3$, then $d = 2$ and the two sums differ by at most $(\ell^f-1)/(\ell-1) < (\ell^f-1)/d$, so the above congruence holds if and only if equality holds, in which case $B_1^* = B_2^*$, so $B_1 = B_2$. If $\ell = 2$ or 3, then $d = \ell - 1$, but the two sums differ by $(\ell^f-1)/(\ell-1)$ only when one of B_1 or B_2 is $\{0,\dots,f-1\}$ and the other is empty.

Now consider the case where f is even and $\ell > 3$. We then have

$$n_B \equiv (\ell+1)\sum_{i\in B^*}(-1)^i\ell^i \bmod \ell^f - 1,$$

where $B^* = (\{0,2,\ldots,f-2\}\cap B)\cup(\{1,3,\ldots,f-1\}\setminus B)$. Thus $n_{B_1} \equiv n_{B_2} \bmod \ell^f - 1$ if and only if

$$\sum_{i\in B_1^*}(-1)^i\ell^i \equiv \sum_{i\in B_2^*}(-1)^i\ell^i \bmod (\ell^f-1)/(\ell+1).$$

But these two sums differ by at most $(\ell^f-1)/(\ell-1)$, which is less than $2(\ell^f - 1)/(\ell+1)$. So if the above congruence holds then either equality holds, in which case $B_1 = B_2$, or

$$\sum_{i\in B_2^*}(-1)^i\ell^i = \sum_{i\in B_1^*}(-1)^i\ell^i + \sum_{i=0}^{f-1}(-1)^i\ell^i,$$

exchanging B_1 and B_2 if necessary. Solving $\bmod \ell^r$ inductively on r, we see that the only possibility is that $B_1^* = \emptyset$ and $B_2^* = \{0,1,\ldots,f-1\}$, in which case $n_{B_\nu} \equiv 0 \bmod \ell^f - 1$.

If f is even and $\ell = 3$, then the situation is the same, except that we have $(\ell^f-1)/(\ell-1) = 2(\ell^f-1)/(\ell+1)$, so in addition to the possibilities that arose for $\ell > 3$, we have

$$n_{\{0,\ldots,f-1\}} \equiv n_\emptyset \equiv (\ell^f-1)/2 \bmod \ell^f - 1$$

as for odd f.

Finally suppose that f is even and $\ell = 2$. In this case we have

$$n_B \equiv 3\sum_{i\in B^*}(-1)^i 2^i \bmod 2^f - 1,$$

where $B^* = (\{0,2,\ldots,f-1\}\cap B)\cup(\{1,3,\ldots,f-2\}\setminus B)$. In particular $n_B \equiv 0 \bmod 3$. Moreover we have $|n_{B_1} - n_{B_2}| \le 3(2^f-1)$ with equality possible only when one of B_1 or B_2 is $\{0,\ldots,f-1\}$ and the other is empty, in which case $n_{B_\nu} \equiv 0 \bmod 2^f - 1$. Thus each non-zero congruence class $\bmod 2^f - 1$ arises as n_B for at most 3 values of B, while 0 arises for at most 4. It follows that each of the $(2^f-4)/3$ non-zero multiples of $3 \bmod 2^f - 1$ arises as n_B for exactly 3 values of B, while 0 arises for exactly 4 values.

We have thus proved the following propositions:

Proposition 3.4. *Suppose that $\ell > 3$. If f is odd, then the congruence classes* $\bmod \ell^f - 1$ *of the form*

$$-1 + (\ell+1)\sum_{i\in B^*}(-1)^i\ell^i$$

are distinct and non-zero as B^* runs through all subsets of $\{0,1,\ldots,f-1\}$. If f is even, then the congruence classes $\bmod \ell^f - 1$ *of the form*

$$(\ell+1)\sum_{i\in B^*}(-1)^i\ell^i$$

are distinct and non-zero as B^* runs through all non-empty proper subsets of $\{0, 1, \ldots, f-1\}$. Letting A denote the set of such classes in each case, we have

$$|W'(\chi_1, \chi_2)| = \begin{cases} 2^f + 2, & \text{if } n = 0 \text{ and } f \text{ is even,} \\ 2^f + 1, & \text{if } n \in A, \\ 2^f, & \text{otherwise.} \end{cases}$$

Proposition 3.5. *Suppose that $\ell = 3$. If f is odd, then the congruence classes* $\bmod\ 3^f - 1$ *of the form*

$$-1 + 4\sum_{i \in B^*} (-1)^i 3^i$$

are distinct and non-zero $\bmod (3^f - 1)/2$ *as B^* runs through all the subsets of $\{0, 1, \ldots, f-1\}$ other than $\{0, 2, \ldots, f-1\}$ and $\{1, 3, \ldots, f-2\}$. If f is even, then the congruence classes* $\bmod 3^f - 1$ *of the form*

$$4\sum_{i \in B^*} (-1)^i 3^i$$

are distinct and non-zero $\bmod (3^f - 1)/2$ *as B^* runs through all non-empty proper subsets of $\{0, 1, \ldots, f-1\}$ other than $\{0, 2, \ldots, f-2\}$ and $\{1, 3, \ldots, f-1\}$. Letting A denote the set of such classes in each case, we have*

$$|W'(\chi_1, \chi_2)| = \begin{cases} 2^f + 2, & \text{if } n = 0 \text{ and } f \text{ is even, or } n = (\ell^f - 1)/2 \\ 2^f + 1, & \text{if } n \in A, \\ 2^f, & \text{otherwise.} \end{cases}$$

Proposition 3.6. *Suppose that $\ell = 2$. Then*

$$|W'(\chi_1, \chi_2)| = \begin{cases} 2^f + 4, & \text{if } n = 0 \text{ and } f \text{ is even,} \\ 2^f + 3 & \text{if } n \neq 0,\ 3|n \text{ and } f \text{ is even,} \\ 2^f + 2, & \text{if } n = 0 \text{ and } f \text{ is odd,} \\ 2^f + 1, & \text{if } n \neq 0 \text{ and } f \text{ is odd,} \\ 2^f, & \text{if } 3 \nmid n \text{ and } f \text{ is even.} \end{cases}$$

As in the irreducible case, multiple B can occur with the same $(a, \vec{b})$.

Proposition 3.7. *The projection map from $W'(\chi_1, \chi_2)$ onto its first component fails to be injective if and only if $\ell^r n \equiv m \bmod \ell^f - 1$ for some integers r, m with $|m| \leq \max\{0, \ell(\ell^{f-2} - 1)/(\ell - 1)\}$.*

Proof. The statement that the projection map from $W'(\chi_1, \chi_2)$ to its first component is not injective is equivalent to the statement that for some a, $\vec{b}$ and $B_1 \neq B_2$, we have

$$n_1 \equiv n_{a,\vec{b},B_1} \equiv n_{a,\vec{b},B_2} \bmod \ell^f - 1$$

and

$$n_2 \equiv n_{a,\vec{b},\overline{B}_1} \equiv n_{a,\vec{b},\overline{B}_2} \bmod \ell^f - 1,$$

where $\overline{B}$ denotes the complement of B in $\{0, 1, 2, \ldots, f-1\}$.

We first deal with the special case $n = 0$. In this case the map is never injective; take $\vec{b} = (\ell - 1, \ell - 1, \ldots, \ell - 1)$, $B_1 = \{0, 1, \ldots, f-1\}$ and $B_2 = \emptyset$ (with the appropriate value of a).

So let us now assume that $n \not\equiv 0 \bmod \ell^f - 1$. Suppose first that the projection map is not injective. Because $n \not\equiv 0$ we check that B_2 cannot be the complement of B_1 (note that this finishes the proof in the case $f = 1$). We can then assume

$f-1\in B_1\setminus B_2$ and $0\notin(B_1\setminus B_2)\cup(B_2\setminus B_1)$ after multiplying n by a power of ℓ as in the irreducible case. The rest of the argument is similar to the irreducible case, except that

$$\sum_{i\in B_1\setminus B_2} b_i\ell^i - \sum_{i\in B_2\setminus B_1} b_i\ell^i = 2(\ell^f-1)$$

is possible if $\ell=2$, but this forces $n\equiv 0 \bmod \ell^f-1$. □

For each pair $\alpha=(V_{\vec a,\vec b},J)\in W'(\chi_1,\chi_2)$ we shall define below a subspace $L_\alpha\subset H^1(K_{\mathfrak p},\overline{\mathbf F}_\ell(\chi_1\chi_2^{-1}))$ of dimension $|J|+\delta$, where $\delta=0$ except in certain cases where $\chi_1\chi_2^{-1}$ is trivial or cyclotomic. We then define $W_{\mathfrak p}(\rho)$ by the following rule:

$V_{\vec a,\vec b}\in W_{\mathfrak p}(\rho)$ if and only if $c_\rho\in L_\alpha$ for some $\alpha=(V_{\vec a,\vec b},J)\in W'(\chi_1,\chi_2)$.

Before defining the subspace L_α, we recall some facts about crystalline representations. Recall that a character $\psi: D\to\overline{\mathbf Q}_\ell^\times$ is crystalline if and only the filtered ϕ-module $D_{\mathrm{crys}}(\psi)=(B_{\mathrm{crys}}\otimes_{\mathbf Q_\ell}\overline{\mathbf Q}_\ell(\psi))^D$ is free of rank one over $K_{\mathfrak p}\otimes_{\mathbf Q_\ell}\overline{\mathbf Q}_\ell$.

For each $\tau\in S$, let $e_\tau: K_{\mathfrak p}\otimes_{\mathbf Q_\ell}\overline{\mathbf Q}_\ell\to\overline{\mathbf Q}_\ell$ denote the projection defined by $a\otimes b\mapsto\tilde\tau(a)b$ where $\tilde\tau$ is the embedding $K_{\mathfrak p}\to\overline{\mathbf Q}_\ell$ reducing to τ, and let $e_\tau D_{\mathrm{crys}}(\psi)$ denote the filtered $\overline{\mathbf Q}_\ell$-vector space $D_{\mathrm{crys}}(\psi)\otimes_{K\otimes\overline{\mathbf Q}_\ell,\tau}\overline{\mathbf Q}_\ell$.

Lemma 3.8. *Suppose that ψ is a crystalline character and for each $\tau\in S$, m_τ is the integer such that $\mathrm{gr}^{-m_\tau}e_\tau D_{\mathrm{crys}}(\psi)\neq 0$. Then $\overline\psi|_I=\prod_{\tau\in S}\omega_\tau^{m_\tau}$.*

Proof. Crystalline characters satisfying the first condition correspond to (weakly) admissible filtered ϕ-modules with the specified filtration. These exist, and any two such differ by an unramified twist. Taking tensor products, the lemma reduces to the case where $m_\tau=1$ or 0 according to whether or not $\tau=\tau_0$. The result in this case follows for example from Theorems 5.3 and 8.4 of [25]. □

Recall that $\alpha=(V_{\vec a,\vec b},J)\in W'(\chi_1,\chi_2)$ if and only if

$$\chi_1|_{I_{K_{\mathfrak p}}}=\prod_{\tau\in S}\omega_\tau^{a_\tau}\prod_{\tau\in J}\omega_\tau^{b_\tau},\qquad \chi_2|_{I_{K_{\mathfrak p}}}=\prod_{\tau\in S}\omega_\tau^{a_\tau}\prod_{\tau\notin J}\omega_\tau^{b_\tau}.$$

Lemma 3.9. *Suppose that $\alpha=(V_{\vec a,\vec b},J)\in W'(\chi_1,\chi_2)$. Let $m_{\tau,\alpha}=b_\tau$ or $-b_\tau$ according to whether or not $\tau\in J$. Then there is a unique lift χ_α of $\chi_1\chi_2^{-1}$ with the following properties:*

- *χ_α is crystalline with $\mathrm{gr}^{-m_\tau}e_\tau D_{\mathrm{crys}}(\chi_\alpha)\neq 0$ for each $\tau\in S$;*
- *if $g\in D^{\mathrm{ab}}$ corresponds via local class field theory to ℓ, then $\chi_\alpha(g)$ is the Teichmüller lift of $\chi_1\chi_2^{-1}(g)$.*

Proof. Let ψ be a character satisfying the first condition. The preceding lemma shows that the reduction of ψ is an unramified twist of $\chi_1\chi_2^{-1}$. Let $\chi_\alpha=\delta\psi$, where $\delta(g)=\tilde\chi_1\tilde\chi_2^{-1}(g)\psi^{-1}(g)$ (and the tildes denote Teichmüller lifts). □

Recall that if $\psi: G_{K_{\mathfrak p}}\to\overline{\mathbf Q}_\ell^\times$ is a crystalline representation, then Bloch and Kato define a subspace $H^1_f(K_{\mathfrak p},\overline{\mathbf Q}_\ell(\psi))$ corresponding to those extensions of topological $\overline{\mathbf Q}_\ell G_{K_{\mathfrak p}}$-modules

$$0\to\overline{\mathbf Q}_\ell(\psi)\to E\to\overline{\mathbf Q}_\ell\to 0$$

which are crystalline. By Corollary 3.8.4 of [5] for example, we have

$$\dim H^1_f(K_{\mathfrak p},\overline{\mathbf Q}_\ell(\psi))=\dim H^0(K_{\mathfrak p},\overline{\mathbf Q}_\ell(\psi))+\dim D_{\mathrm{crys}}(\psi)-\dim \mathrm{Fil}^0 D_{\mathrm{crys}}(\psi)$$

where the dimensions are over $\overline{\mathbf{Q}}_\ell$. Applying this for $\psi = \chi_\alpha$ for $\alpha = (V_{\vec{a},\vec{b}}, J) \in W'(\chi_1,\chi_2)$, we see that $\dim H^1_f(K_\mathfrak{p}, \overline{\mathbf{Q}}_\ell(\chi_\alpha)) = |J|$. We then define $H^1_f(K_\mathfrak{p}, \overline{\mathbf{Z}}_\ell(\chi_\alpha))$ as the preimage of $H^1_f(K_\mathfrak{p}, \overline{\mathbf{Q}}_\ell(\chi_\alpha))$ under the natural map

$$H^1(K_\mathfrak{p}, \overline{\mathbf{Z}}_\ell(\chi_\alpha)) \to H^1(K_\mathfrak{p}, \overline{\mathbf{Q}}_\ell(\chi_\alpha))$$

and L'_α as the image of $H^1_f(K_\mathfrak{p}, \overline{\mathbf{Z}}_\ell(\chi_\alpha))$ under the natural map

$$H^1(K_\mathfrak{p}, \overline{\mathbf{Z}}_\ell(\chi_\alpha)) \to H^1(K_\mathfrak{p}, \overline{\mathbf{F}}_\ell(\chi_1\chi_2^{-1})).$$

We then let $L_\alpha = L'_\alpha$ except in two cases:

- if $\chi_1\chi_2^{-1}$ is the cyclotomic character, $\vec{b} = (\ell, \dots, \ell)$ and $J = S$, then we let $L_\alpha = H^1(K_\mathfrak{p}, \overline{\mathbf{F}}_\ell(\chi_1\chi_2^{-1}))$;
- if $\chi_1\chi_2^{-1}$ is the trivial character and $J \neq S$, then we let $L_\alpha = L'_\alpha + L_{\mathrm{ur}}$, where L_{ur} is the one-dimensional space of unramified classes in $H^1(K_\mathfrak{p}, \overline{\mathbf{F}}_\ell)$.

Lemma 3.10. *If $\alpha = (V_{\vec{a},\vec{b}}, J) \in W'(\chi_1,\chi_2)$, then $\dim L_\alpha = |J|$ except in the following cases:*

(1) *if $\chi_1\chi_2^{-1}$ is cyclotomic, $\vec{b} = (\ell, \dots, \ell)$, $J = S$ and $\ell > 2$, then $\dim L_\alpha = |J| + 1$;*
(2) *if $\chi_1\chi_2^{-1}$ is trivial and $\ell > 2$, then $\dim L_\alpha = |J| + 1$ or $|J| + 2$ according to whether or not $L_{\mathrm{ur}} \subset L'_\alpha$;*
(3) *if $\chi_1\chi_2^{-1}$ is trivial (or equivalently, cyclotomic) and $\ell = 2$, then $\dim L_\alpha = |J| + 1$ unless either $L_{\mathrm{ur}} \not\subset L'_\alpha$ or $\vec{b} = (\ell, \dots, \ell)$ and $J = S$, in which case $\dim L_\alpha = |J| + 2$.*

Proof. Note first that $H^1_f(K_\mathfrak{p}, \overline{\mathbf{Z}}_\ell(\chi_\alpha))$ contains $H^1(K_\mathfrak{p}, \overline{\mathbf{Z}}_\ell(\chi_\alpha))_{\mathrm{tor}}$ and that the quotient is free of rank $|J| = \dim H^1_f(K_\mathfrak{p}, \overline{\mathbf{Q}}_\ell(\chi_\alpha))$. Therefore the natural map

$$H^1_f(K_\mathfrak{p}, \overline{\mathbf{Z}}_\ell(\chi_\alpha)) \otimes_{\overline{\mathbf{Z}}_\ell} \overline{\mathbf{F}}_\ell \to H^1(K_\mathfrak{p}, \overline{\mathbf{Z}}_\ell(\chi_\alpha)) \otimes_{\overline{\mathbf{Z}}_\ell} \overline{\mathbf{F}}_\ell \to H^1(K_\mathfrak{p}, \overline{\mathbf{F}}_\ell(\chi_1\chi_2^{-1}))$$

is injective and its image L'_α has dimension $|J|$, unless $\chi_1\chi_2^{-1}$ is trivial, in which case the dimension is $|J| + 1$. The lemma is now immediate from the definition of L_α and, in the cyclotomic $J = S$ case, the local Euler characteristic formula. $\square$

If $\rho|_{G_{K_\mathfrak{p}}} \sim \begin{pmatrix} \chi_1 & * \\ 0 & \chi_2 \end{pmatrix}$ and c_ρ is the corresponding class in $H^1(K_\mathfrak{p}, \overline{\mathbf{F}}_\ell(\chi_1\chi_2^{-1}))$, we now define

(3) $$W_\mathfrak{p}(\rho) = \{\, V \mid c_\rho \in L_\alpha \text{ for some } \alpha = (V, J) \in W'(\chi_1,\chi_2) \,\}.$$

Note that if $\rho|_{G_{K_\mathfrak{p}}} \sim \chi_1 \oplus \chi_2$, or equivalently $c_\rho = 0$, then $W_\mathfrak{p}(\rho) = \pi_1(W'(\chi_1,\chi_2))$ is independent of the choice of ordering of χ_1 and χ_2. Note also in this case that $|W_\mathfrak{p}(\rho)|$ has size approximately 2^f, as in the irreducible case (1).

(We remark that it is shown in [17] that in the cases where $\rho|_{G_{K_\mathfrak{p}}}$ is semisimple, the set $W_\mathfrak{p}(\rho)$ is related to the set of Jordan-Hölder constituents of the reduction of a corresponding irreducible characteristic zero representation of $\mathrm{GL}_2(k)$.)

3.3. **Basic properties of the definition.** The set $W_\mathfrak{p}(\rho)$ was defined in terms of the restriction of ρ to $G_{K_\mathfrak{p}}$. We now check that it is in fact non-empty and depends only on the restriction to inertia.

Proposition 3.11. *If $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is continuous, irreducible and totally odd, then $W_\mathfrak{p}(\rho)$ is non-empty and depends only on $\rho|_{I_{K_\mathfrak{p}}}$.*

Proof. We first prove that $W_{\mathfrak{p}}(\rho) \neq \emptyset$. If $\rho|_{G_{K_{\mathfrak{p}}}}$ is irreducible, then it is induced from a character ξ, and Propositions 3.1 and 3.2 show that $W'(\xi)$ is non-empty, and hence so is $W_{\mathfrak{p}}(\rho)$ (note that if $\ell^f = 2$, then n is not divisible by 3). If $\rho|_{G_{K_{\mathfrak{p}}}}$ is reducible, then it is of the form $\begin{pmatrix} \chi_1 & * \\ 0 & \chi_2 \end{pmatrix}$, and we showed that the projection map $W'(\chi_1, \chi_2) \to \{J \subset S\}$ is surjective. In particular, there is an element $\alpha = (V_{\vec{a},\vec{b}}, S) \in W'(\chi_1, \chi_2)$. Moreover if $\chi_1\chi_2^{-1}$ is cyclotomic, we may choose $\vec{b} = (\ell, \ldots, \ell)$, so that in all cases $L_\alpha = H^1(K_{\mathfrak{p}}, \overline{\mathbf{F}}_\ell(\chi_1\chi_2^{-1}))$ by Lemma 3.10 and the local Euler characteristic formula. It follows that $c_\rho \in L_\alpha$ and $V_{\vec{a},\vec{b}} \in W_{\mathfrak{p}}(\rho)$.

For the dependence only on inertia, first note that the irreducibility of $\rho|_{G_{K_{\mathfrak{p}}}}$ is determined by $\rho|_{I_{K_{\mathfrak{p}}}}$. In the case that $\rho|_{G_{K_{\mathfrak{p}}}}$ is irreducible, $W_{\mathfrak{p}}(\rho)$ is determined by $W'(\xi)$, which depends only on $\xi|_{I_{K_{\mathfrak{p}}}}$, which in turn depends only on $\rho|_{I_{K_{\mathfrak{p}}}}$.

Suppose now that $\rho|_{G_{K_{\mathfrak{p}}}}$ is reducible and $\rho' : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is such that $\rho|_{I_{K_{\mathfrak{p}}}} \sim \rho'|_{I_{K_{\mathfrak{p}}}}$. Changing bases, we may assume $\rho|_{G_{K_{\mathfrak{p}}}} = \begin{pmatrix} \chi_1 & * \\ 0 & \chi_2 \end{pmatrix}$ and $\rho|_{I_{K_{\mathfrak{p}}}} = \rho'|_{I_{K_{\mathfrak{p}}}}$. Note that the function $G_{K_{\mathfrak{p}}}/I_{K_{\mathfrak{p}}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ defined by $g \mapsto \rho'(g)\rho(g)^{-1}$ takes values in $Z(\rho(I_{K_{\mathfrak{p}}}))$. We divide the proof into cases according to the possible centralisers.

Suppose first that $\chi_1\chi_2^{-1}$ is ramified and c_ρ has non-trivial restriction to $I_{K_{\mathfrak{p}}}$. In this case $\rho|_{I_{K_{\mathfrak{p}}}}$ is indecomposable and has centraliser consisting only of the scalar matrices. It follows that $\rho' = \psi\rho$ for some unramified character $\psi : G_{K_{\mathfrak{p}}} \to \overline{\mathbf{F}}_\ell^\times$. Since $W'(\chi_1, \chi_2)$ depends only on the restriction of χ_1 and χ_2 to $I_{K_{\mathfrak{p}}}$, we have $W'(\psi\chi_1, \psi\chi_2) = W'(\chi_1, \chi_2)$. Moreover the subspaces L_α of $H^1(K_{\mathfrak{p}}, \overline{\mathbf{F}}_\ell(\chi_1\chi_2^{-1}))$ do not change if ρ is replaced by an unramified twist, nor does the class c_ρ. It follows that $W_{\mathfrak{p}}(\rho') = W_{\mathfrak{p}}(\rho)$.

Next suppose that $\chi_1\chi_2^{-1}$ is ramified and c_ρ has trivial restriction to $I_{K_{\mathfrak{p}}}$. Then in fact $c_\rho = 0$, so we may assume $\rho|_{G_{K_{\mathfrak{p}}}} = \begin{pmatrix} \chi_1 & 0 \\ 0 & \chi_2 \end{pmatrix}$. In this case the centraliser of $\rho(I_{K_{\mathfrak{p}}})$ consists of the diagonal matrices. It follows that $\rho'|_{G_{K_{\mathfrak{p}}}} = \begin{pmatrix} \psi_1\chi_1 & 0 \\ 0 & \psi_2\chi_2 \end{pmatrix}$ for some unramified characters ψ_1 and ψ_2. Moreover we have $c_{\rho'} = 0$ and $W'(\psi_1\chi_1, \psi_2\chi_2) = W'(\chi_1, \chi_2)$, so $W_{\mathfrak{p}}(\rho') = W_{\mathfrak{p}}(\rho)$.

Next suppose that $\chi_1\chi_2^{-1}$ is unramified and c_ρ has non-trivial restriction to $I_{K_{\mathfrak{p}}}$. In this case we have

$$Z(\rho(I_{K_{\mathfrak{p}}})) = \left\{ \begin{pmatrix} x & xy \\ 0 & x \end{pmatrix} \middle| \; x \in \overline{\mathbf{F}}_\ell^\times, y \in \overline{\mathbf{F}}_\ell \right\},$$

so if $g \in G_{K_{\mathfrak{p}}}$, then

$$\rho'(g) = \begin{pmatrix} 1 & \mu(g) \\ 0 & 1 \end{pmatrix} \psi(g)\rho(g)$$

for some unramified character ψ and cocycle $\mu : G_{K_{\mathfrak{p}}}/I_{K_{\mathfrak{p}}} \to \overline{\mathbf{F}}_\ell(\chi_1\chi_2^{-1})$. In particular, $\rho' \sim \begin{pmatrix} \chi_1' & * \\ 0 & \chi_2' \end{pmatrix}$ with $\chi_1'(\chi_2')^{-1} = \chi_1\chi_2^{-1}$. Moreover if $\chi_1 \neq \chi_2$, then $c_{\rho'} = c_\rho$ in $H^1(K_{\mathfrak{p}}, \overline{\mathbf{F}}_\ell(\chi_1\chi_2^{-1}))$, and if $\chi_1 = \chi_2$, then $c_{\rho'} - c_\rho \in L_{\mathrm{ur}}$. Since the spaces L_α are the same for ρ and ρ' and contain L_{ur} if $\chi_1 = \chi_2$, we conclude that $W_{\mathfrak{p}}(\rho') = W_{\mathfrak{p}}(\rho)$.

Finally suppose that $\chi_1\chi_2^{-1}$ is unramified and c_ρ has trivial restriction to $I_{K_\mathfrak{p}}$, so

$$\rho'(g) = \rho(g) = \begin{pmatrix} \chi_1(g) & 0 \\ 0 & \chi_1(g) \end{pmatrix}$$

for $g \in I_{K_\mathfrak{p}}$. Note that $c_\rho = 0$ unless $\chi_1 = \chi_2$, in which case $c_\rho \in L_{\mathrm{ur}}$, and similarly for $c_{\rho'}$. It follows that $W_\mathfrak{p}(\rho') = \pi_1(W'(\chi_1', \chi_2')) = \pi_1(W'(\chi_1, \chi_2)) = W_\mathfrak{p}(\rho)$. ☐

We are now ready to state the weight conjecture. Recall that $W(\rho)$ is defined as the set of representations of $\prod_{\mathfrak{p}|\ell} \mathrm{GL}_2(k_\mathfrak{p})$ of the form $\otimes_{\overline{\mathbf{F}}_\ell} V_\mathfrak{p}$ with each $V_\mathfrak{p} \in W_\mathfrak{p}(\rho)$. By the preceding proposition $W(\rho)$ is non-empty and depends only on the restrictions of ρ to inertia groups at primes over ℓ.

Conjecture 3.12. *If $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is modular, then*

$$W(\rho) = \{\, V \mid \rho \textit{ is modular of weight } V\}.$$

We now check compatibility of the conjectural weight set with twists and determinants.

Proposition 3.13. *Suppose that $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is continuous, irreducible and totally odd.*

(1) *Let $\chi : G_K \to \overline{\mathbf{F}}_\ell^\times$ be such that $\chi|_{I_{K_\mathfrak{p}}} = \prod_{\tau \in S_\mathfrak{p}} \omega_\tau^{c_\tau}$ for each $\mathfrak{p}|\ell$. Then $V \in W(\rho)$ if and only $V \otimes V_\chi \in W(\chi\rho)$, where*

$$V_\chi = \otimes_{\mathfrak{p}|\ell} \otimes_{\tau \in S_\mathfrak{p}} \det{}^{c_\tau} k_\mathfrak{p}^2 \otimes_\tau \overline{\mathbf{F}}_\ell.$$

(2) *If $V \in W(\rho)$ and V has central character $\otimes_{\mathfrak{p}|\ell} \prod_{\tau \in S_\mathfrak{p}} \tau^{c_\tau}$, then $\det \rho|_{I_{K_\mathfrak{p}}} = \prod_{\tau \in S_\mathfrak{p}} \omega_\tau^{c_\tau + 1}$ for each $\mathfrak{p}|\ell$.*

Proof. To prove the first assertion it suffices to show that $V \in W_\mathfrak{p}(\rho)$ if and only if $V \otimes V_{\chi_\mathfrak{p}} \in W_\mathfrak{p}(\chi\rho)$ where $\chi_\mathfrak{p} = \chi|_{G_{K_\mathfrak{p}}}$ and $V_{\chi_\mathfrak{p}} = \otimes_{\tau \in S_\mathfrak{p}} \det{}^{c_\tau} k_\mathfrak{p}^2 \otimes_\tau \overline{\mathbf{F}}_\ell$. If $\rho|_{G_{K_\mathfrak{p}}} \sim \mathrm{Ind}_{D'}^{D} \xi$ is irreducible, then $(V, J) \in W'(\xi)$ if and only $(V \otimes V_{\chi_\mathfrak{p}}, J) \in W'(\xi\chi_\mathfrak{p})$, yielding the assertion in this case. If $\rho|_{G_{K_\mathfrak{p}}} \sim \begin{pmatrix} \chi_1 & * \\ 0 & \chi_2 \end{pmatrix}$ is reducible, then $\alpha = (V, J) \in W'(\chi_1, \chi_2)$ if and only $\alpha' = (V \otimes V_{\chi_\mathfrak{p}}, J) \in W'(\chi_1\chi_\mathfrak{p}, \chi_2\chi_\mathfrak{p})$. Moreover, since $V_{\vec{a},\vec{b}} \otimes V_{\chi_\mathfrak{p}}$ is of the form $V_{\vec{a}',\vec{b}}$, we see that $\chi_\alpha = \chi_{\alpha'}$, so $L_\alpha = L'_\alpha \subset H^1(K_\mathfrak{p}, \overline{\mathbf{F}}_\ell(\chi_1\chi_2^{-1}))$. Since also $c_\rho = c_{\chi\rho}$, we get the assertion in this case as well.

To prove the second assertion, we can again work locally at primes $\mathfrak{p}|\ell$. Writing $V = \otimes_{\mathfrak{p}|\ell} V_\mathfrak{p}$, we have $V_\mathfrak{p} \in W_\mathfrak{p}(\rho)$ for each $\mathfrak{p}|\ell$. If $V_\mathfrak{p} = V_{\vec{a},\vec{b}}$, this gives $\det \rho|_{I_{K_\mathfrak{p}}} = \prod_{\tau \in S_\mathfrak{p}} \omega_\tau^{2a_\tau + b_\tau}$. Since $V_{\vec{a},\vec{b}}$ has central character $\prod_{\tau \in S_\mathfrak{p}} \tau^{2a_\tau + b_\tau - 1}$, we have

$$\sum_{i=0}^{f_\mathfrak{p}-1} c_{\tau_i} \ell^i \equiv \sum_{i=0}^{f_\mathfrak{p}-1} (2a_{\tau_i} + b_{\tau_i} - 1)\ell^i \bmod (\ell^{f_\mathfrak{p}} - 1).$$

Adding $\sum_{i=0}^{f_\mathfrak{p}-1} \ell^i$ to each side of the congruence, we deduce that $\prod_{\tau \in S_\mathfrak{p}} \omega_\tau^{c_\tau + 1} = \prod_{\tau \in S_\mathfrak{p}} \omega_\tau^{2a_\tau + b_\tau}$. ☐

Combining the first part of the proposition with Corollary 2.11(b), we deduce the following:

Corollary 3.14. *Suppose that $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is continuous, irreducible and totally odd and $\chi : G_K \to \overline{\mathbf{F}}_\ell^\times$ is a character. Then Conjecture 3.12 holds for ρ if and only if it holds for $\chi\rho$.*

In the case $K = \mathbf{Q}$, Conjecture 3.12 follows from known results on Serre's Conjecture (cf. [41], [2]).

Theorem 3.15. *Conjecture 3.12 holds if $K = \mathbf{Q}$.*

Proof. Replacing ρ by a twist, we can assume $\rho|_{I_\ell}$ has the form $\begin{pmatrix} \omega_2^b & 0 \\ 0 & \omega_2^{\ell b} \end{pmatrix}$ or $\begin{pmatrix} \omega^b & * \\ 0 & 1 \end{pmatrix}$ for some b with $1 \le b \le \ell - 1$. In the second case we write $\rho|_{G_{\mathbf{Q}_\ell}} \sim \begin{pmatrix} \chi_1 & * \\ 0 & \chi_2 \end{pmatrix}$. We shall also use ω to denote the mod ℓ cyclotomic character on $G_{\mathbf{Q}}$ or $G_{\mathbf{Q}_\ell}$.

In the first case we find that $W(\rho) = \{V_{0,b}, V_{b-1,\ell+1-b}\}$. In the second case, we have the following possibilities:

$$W'(\chi_1,\chi_2) = \begin{cases} \{(V_{0,b},S),(V_{b,\ell-1-b},\emptyset)\}, & \text{if } 1 < b < \ell-2, \\ \{(V_{0,\ell-1},S),(V_{0,\ell-1},\emptyset)\}, & \text{if } b = \ell-1 \text{ and } \ell > 2, \\ \{(V_{0,\ell},S),(V_{0,1},S),(V_{1,\ell-2},\emptyset)\}, & \text{if } b = 1 \text{ and } \ell > 3, \\ \{(V_{0,\ell-2},S),(V_{\ell-2,\ell},\emptyset),(V_{\ell-2,1},\emptyset)\}, & \text{if } b = \ell-2 \text{ and } \ell > 3, \\ \{(V_{0,\ell},S),(V_{0,1},S),(V_{1,\ell},\emptyset),(V_{1,1},\emptyset)\}, & \text{if } b = 1 \text{ and } \ell \le 3. \end{cases}$$

Moreover, dimension considerations show that $L_{(V,J)} = H^1(G_{\mathbf{Q}_\ell}, \overline{\mathbf{F}}(\chi_1\chi_2^{-1}))$ whenever $J = S$, unless $\chi_1\chi_2^{-1} = \omega$ and $V = V_{(0,1)}$, in which case the construction of $L_{(V,J)}$ shows that the corresponding extensions arise from finite flat group schemes over $\mathbf{Z}_\ell$, hence are peu ramifiées in the sense of Serre [56]. Considering dimensions in this case then gives that $L_{(V,J)}$ consists precisely of the classes which are peu ramifiées. Note also that $L_{(V,J)} = 0$ whenever $J = \emptyset$ unless $\chi_1\chi_2^{-1}$ is trivial, in which case we do not need an explicit description of $L_{(V,J)}$. This gives

$$W(\rho) = \begin{cases} \{V_{0,b}\}, & \text{if } 1 < b < \ell-1 \text{ and } \rho|_{G_{\mathbf{Q}_\ell}} \text{ is non-split}, \\ \{V_{0,b}, V_{b,\ell-1-b}\}, & \text{if } 1 < b < \ell-2 \text{ and } \rho|_{G_{\mathbf{Q}_\ell}} \text{ is split}, \\ \{V_{0,\ell-2}, V_{\ell-2,\ell}, V_{\ell-2,1}\}, & \text{if } b = \ell-2,\ \ell > 3 \text{ and } \rho|_{G_{\mathbf{Q}_\ell}} \text{ is split}, \\ \{V_{0,\ell-1}\}, & \text{if } b = \ell-1 \text{ and } \ell > 2, \\ \{V_{0,\ell}\}, & \text{if } b = 1,\ \chi_1\chi_2^{-1} = \omega \text{ and } \rho|_{G_{\mathbf{Q}_\ell}} \text{ is très ramifiée}, \\ \{V_{0,\ell}, V_{0,1}, V_{1,\ell-2}\}, & \text{if } b = 1,\ \ell > 3 \text{ and } \rho|_{G_{\mathbf{Q}_\ell}} \text{ is split}, \\ \{V_{0,3}, V_{0,1}, V_{1,3}, V_{1,1}\}, & \text{if } b = 1,\ \ell = 3 \text{ and } \rho|_{G_{\mathbf{Q}_\ell}} \text{ is split}, \\ \{V_{0,\ell}, V_{0,1}\}, & \text{otherwise}. \end{cases}$$

Propositions 2.5 and Corollary 2.11(b) show that ρ is modular of weight[1] $b+1$ and level prime to ℓ if and only if $\omega^{a+b}\rho$ is modular of weight $V_{a,b}$. If ρ is modular of weight 2 and level prime to ℓ, then $\omega\rho|_{G_{\mathbf{Q}_\ell}}$ arises from a finite flat group scheme over $\mathbf{Z}_\ell$, so it follows from results of Deligne and Fontaine ([20], Theorems 2.5, 2.6) and the explicit description of $W(\rho)$ above that if ρ is modular of weight V, then $V \in W(\rho)$.

[1]Most of the literature on Serre's Conjecture in the classical case uses arithmetic conventions, so for the purpose of this proof, we view ρ as "modular of weight k" if $\rho \sim \overline{\rho}_\pi$ for some cuspidal automorphic π with $\pi_\infty \cong D_{k,k}$ in the notation of §2 and [10].

To show that if $V \in W(\rho)$, then ρ is modular of weight V, we combine the following four results. The first of these is a standard consequence of multiplication by the Hasse invariant (or Eisenstein series) of weight $\ell - 1$.

Lemma 3.16. *If ρ is modular of weight 2 and level prime to ℓ, then ρ is modular of weight $\ell + 1$ and level prime to ℓ.*

The theorem in the irreducible case is then a consequence of the following result of Edixhoven; see the second paragraph of [20, 4.5].

Lemma 3.17. *Suppose that ρ is modular of weight $b + 1$ and level prime to ℓ with $2 \le b \le \ell$. If $\rho|_{G_{\mathbf{Q}_\ell}}$ is irreducible, then $\omega^{1-b}\rho$ is modular of weight $\ell + 2 - b$ and level prime to ℓ.*

To treat the reducible case, we first apply Mazur's Principle [20, 2.8].

Theorem 3.18. *Suppose that ρ is modular of weight $\ell + 1$ and level prime to ℓ. If ρ is not modular of weight 2 and level prime to ℓ, then $\rho|_{G_{\mathbf{Q}_\ell}}$ is a très ramifiée representation of the form $\begin{pmatrix} \omega\chi_2 & * \\ 0 & \chi_2 \end{pmatrix}$ for some unramified character χ_2.*

The theorem is then a consequence of the companion forms theorem, proved by Gross [34] subject to unchecked compatibilities, then by Coleman and Voloch [11], but under slightly different hypotheses than we need. Yet another proof was given by Faltings and Jordan [24], whose version we cite:

Theorem 3.19. *Suppose that ρ is modular of weight $b + 1$ and level prime to ℓ with $1 \le b \le \ell - 2$. If $\rho|_{G_{\mathbf{Q}_\ell}}$ is reducible and split, then $\omega^{-b}\rho$ is modular of weight $\ell - b$ and level prime to ℓ.*

Theorem 3.15 now follows. □

We end by remarking that Edixhoven's refinement of Serre's conjecture includes the statement that if ρ is unramified at ℓ then it should come from a mod ℓ modular form of weight 1. This refinement is not implied by our conjecture.

4. Mod ℓ Langlands correspondences

As Serre himself remarks in [56], his conjecture can be viewed as part of a "mod ℓ Langlands philosophy." Indeed the weak conjecture can be viewed as asserting the existence of a global mod ℓ Langlands correspondence for $\mathrm{GL}_2/\mathbf{Q}$, and the refinement can be viewed as a local-global compatibility statement. This was made precise by Emerton in [22] as follows. Consider the representation

$$H = \lim_{\vec{U}} H^1_{\text{ét}}(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell)$$

of $G_{\mathbf{Q}} \times \mathrm{GL}_2(\mathbf{A}_f)$. The weak form of Serre's conjecture is the statement that if $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is continuous, odd and irreducible, then

$$\pi(\rho) := \mathrm{Hom}_{\overline{\mathbf{F}}_\ell[G_{\mathbf{Q}}]}(\rho, H)$$

is non-zero. Under some technical hypotheses on $\rho|_{G_{\mathbf{Q}_\ell}}$, Emerton shows that $\pi(\rho)$ factors as a restricted tensor product of representations π_p where π_p is a representation of $\mathrm{GL}_2(\mathbf{Q}_p)$ determined by $\rho|_{G_{\mathbf{Q}_p}}$. The "level part" of Serre's refinement can then be recovered from the fact that if $p \neq \ell$ and $\rho|_{G_{\mathbf{Q}_p}}$ has Artin conductor p^{c_p}, then $\pi_p^{U_1(p^{c_p})} \neq 0$, and the "weight part" from the fact that

$$\mathrm{Hom}_{\mathrm{GL}_2(\mathbf{Z}_\ell)}(\det{}^{1-k} \otimes \mathrm{Symm}^{k-2}\,\overline{\mathbf{F}}_\ell^2, \pi_\ell) \neq 0$$

where $k = k(\rho) \geq 2$ is the weight of $\rho|_{G_{\mathbf{Q}_\ell}}$ as defined by Serre in [56]. In this section we formulate a conjectural extension of Emerton's refinement to our setting, namely that of mod ℓ representations arising from a quaternion algebra D over a totally real field K.

Suppose now that

$$\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$$

is continuous, totally odd and irreducible. We shall associate to ρ a smooth representation $\pi^D(\rho)$ of $(D \otimes \hat{\mathbf{Z}})^\times$ over $\overline{\mathbf{F}}_\ell$ and give a conjectural description for it as a product of local factors. Thus for each prime $\mathfrak{p}$ of K we would like to associate to ρ a smooth admissible representation of $D_\mathfrak{p}^\times$ defined over $\overline{\mathbf{F}}_\ell$, ideally depending only on $\rho|_{G_{K_\mathfrak{p}}}$. We are able to achieve this for primes $\mathfrak{p}$ not dividing ℓ; indeed this was already done by Emerton if D is split at $\mathfrak{p}$ and the main new ingredient of this section is to treat the case where $D_\mathfrak{p}$ is a quaternion algebra. For $\mathfrak{p}|\ell$, we are not yet able to give a conjectural description of the local factor, but the weight conjecture formulated in the preceding section can be interpreted as a description of the Jordan-Hölder factors of its socle under a maximal compact subgroup of $D_\mathfrak{p}^\times$. (We are grateful to Breuil and Emerton for this observation.)

We begin by recalling Emerton's formulation of a mod ℓ local Langlands correspondence for GL_2 over $K_\mathfrak{p}$ for $\mathfrak{p}$ not dividing ℓ; see Theorem 1.2 of [23]. Emerton's construction is a modification of one provided by Vignéras in [63], on whose results it relies, a key difference being that [23] involves reducible representations of $\mathrm{GL}_2(K_\mathfrak{p})$ in order to prove local-global compatibility.

Fix for now a prime $\mathfrak{p}$ not dividing ℓ and let $q = \mathbf{N}(\mathfrak{p}) = \#(\mathcal{O}_K/\mathfrak{p})$. For a continuous representation $\tilde\rho : G_{K_\mathfrak{p}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_\ell)$, we let $\pi(\tilde\rho)$ denote its local Langlands correspondent as modified in Section 7.1 of [23]. In particular, $\pi(\tilde\rho)$ is the usual irreducible admissible $\overline{\mathbf{Q}}_\ell$-representation of $\mathrm{GL}_2(K_\mathfrak{p})$ provided by the local Langlands correspondence, unless $\tilde\rho$ is the sum of two characters whose ratio is cyclotomic in which case $\pi(\tilde\rho)$ is a generic representation of length two.

Theorem 4.1. *(Emerton) There is a map $\rho \mapsto \pi(\rho)$ from the set of isomorphism classes of continuous representations $G_{K_\mathfrak{p}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ to the set of isomorphism classes of finite length smooth admissible $\overline{\mathbf{F}}_\ell$-representations of $\mathrm{GL}_2(K_\mathfrak{p})$, uniquely determined by the following properties:*

1. *$\pi(\rho)$ has no finite-dimensional subrepresentations.*
2. *If $g \in \mathrm{Aut}\,\overline{\mathbf{F}}_\ell$, then $\pi(\rho^g) = \pi(\rho)^g$.*
3. *If $\tilde\rho : G_{K_\mathfrak{p}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_\ell)$ is a continuous lift of ρ, then there is a $\mathrm{GL}_2(K_\mathfrak{p})$-equivariant $\overline{\mathbf{Z}}_\ell$-lattice in $\pi(\tilde\rho)$ whose reduction admits a $\mathrm{GL}_2(K_\mathfrak{p})$-equivariant embedding into $\pi(\rho)$. Moreover, this lattice is uniquely determined up to multiplication by an element of $\overline{\mathbf{Q}}_\ell^\times$.*
4. *There exists a lifting $\tilde\rho$ of ρ for which the embedding of (3) is an isomorphism.*

If D is split at $\mathfrak{p}$, we let $\pi^{D_\mathfrak{p}}(\rho)$ denote the representation of $D_\mathfrak{p}^\times \cong \mathrm{GL}_2(K_\mathfrak{p})$ given by the theorem.

We would like an analogue of the theorem which associates to ρ a representation of $D_{\mathfrak{p}}^\times$ when $D_{\mathfrak{p}}$ is a non-split quaternion algebra over $K_{\mathfrak{p}}$. Again our construction is a modification of one already provided by Vignéras (this time in [62]) on whose results we rely. Since the irreducible smooth admissible $\overline{\mathbf{F}}_\ell$-representations of $D_{\mathfrak{p}}^\times$ are finite-dimensional, one might expect a result like Theorem 4.1 but without condition (1). However results of Ribet [52] show this naive generalisation is false, and there are further complications when $\ell = 2$. On the other hand we can give a more explicit description of the desired representation $\pi^{D_{\mathfrak{p}}}(\rho)$.

Recall that the (local) Jacquet-Langlands correspondence establishes a bijection between isomorphism classes of irreducible admissible square-integrable $\overline{\mathbf{Q}}_\ell$-representations of $\mathrm{GL}_2(K_{\mathfrak{p}})$ and irreducible admissible $\overline{\mathbf{Q}}_\ell$-representations of $D_{\mathfrak{p}}^\times$. If $\tilde{\rho} : G_{K_{\mathfrak{p}}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_\ell)$ is a continuous representation, then $\pi(\tilde{\rho})$ is square-integrable if and only if $\tilde{\rho}$ is either irreducible or isomorphic to a twist of the non-split representation of the form $\left(\begin{smallmatrix} 1 & * \\ 0 & \omega^{-1} \end{smallmatrix}\right)$ where ω is the ℓ-adic cyclotomic character, and then we let $\pi^{D_{\mathfrak{p}}}(\tilde{\rho})$ the representation of $D_{\mathfrak{p}}^\times$ corresponding to $\pi(\tilde{\rho})$ via Jacquet-Langlands.

Now consider again a continuous representation $\rho : G_{K_{\mathfrak{p}}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$. It is straightforward to check that there exist continuous representations $\tilde{\rho} : G_{K_{\mathfrak{p}}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_\ell)$ containing $G_{K_{\mathfrak{p}}}$-stable lattices whose reduction is ρ. Moreover, if $\tilde{\rho}$ is irreducible, then so is ρ except in the following case: if $q \equiv -1 \bmod \ell$, L is the quadratic unramified extension of $K_{\mathfrak{p}}$ and $\tilde{\rho} \sim \chi \otimes \mathrm{Ind}_{G_L}^{G_{K_{\mathfrak{p}}}} \xi$ for some characters χ of $G_{K_{\mathfrak{p}}}$ and ξ of G_L such that ξ is different from its $\mathrm{Gal}(L/K_{\mathfrak{p}})$-conjugate and has ℓ-power order, then ρ has semi-simplification isomorphic to $\overline{\chi} \oplus \overline{\chi\omega}^{-1}$ where $\overline{\omega}$ is the mod ℓ cyclotomic character. (Note that if $q \equiv -1 \bmod \ell$, then $\overline{\omega}$ is the quadratic unramified character of $G_{K_{\mathfrak{p}}}$, unless $\ell = 2$ in which case $\overline{\omega} = 1$.)

For representations of $D_{\mathfrak{p}}^\times$ we have the following result of Vignéras (Propositions 9 and 11 and Corollary 12 of [62]):

Proposition 4.2. *(Vignéras) Suppose that $\rho : G_{K_{\mathfrak{p}}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is continuous and irreducible. If $\tilde{\rho} : G_{K_{\mathfrak{p}}} \to \mathrm{GL}_2(\overline{\mathbf{Q}}_\ell)$ is a lift of ρ, then the reduction of $\pi^{D_{\mathfrak{p}}}(\tilde{\rho})$ is irreducible and depends only on ρ.*

If $\rho : G_{K_{\mathfrak{p}}} \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ is irreducible, we define $\pi^{D_{\mathfrak{p}}}(\rho)$ to be the reduction of $\pi^{D_{\mathfrak{p}}}(\tilde{\rho})$ for any lift $\tilde{\rho}$ of ρ, well-defined by the proposition.

Suppose now that ρ is reducible. If ρ is not a twist of a representation of the form $\left(\begin{smallmatrix} 1 & * \\ 0 & \overline{\omega}^{-1} \end{smallmatrix}\right)$ (where $\overline{\omega}$ is the mod ℓ cyclotomic character), then ρ has no lifts $\tilde{\rho}$ such that $\pi(\tilde{\rho})$ is square-integrable, and we define $\pi^{D_{\mathfrak{p}}}(\rho) = 0$.

Suppose now that $\rho \sim \left(\begin{smallmatrix} \chi & * \\ 0 & \chi\overline{\omega}^{-1} \end{smallmatrix}\right)$ for some character $\chi : G_{K_{\mathfrak{p}}} \to \overline{\mathbf{F}}_\ell^\times$ (i.e., ρ is any extension of $\chi\overline{\omega}^{-1}$ by χ). If $q \not\equiv -1 \bmod \ell$, then the only lifts $\tilde{\rho}$ of ρ for which $\pi(\tilde{\rho})$ is square-integrable are non-split representations of the form $\left(\begin{smallmatrix} \tilde{\chi} & * \\ 0 & \tilde{\chi}\omega^{-1} \end{smallmatrix}\right)$ where $\tilde{\chi} : G_{K_{\mathfrak{p}}} \to \mathcal{O}_{K_{\mathfrak{p}}}^\times$ lifts χ. In this case $\pi^{D_{\mathfrak{p}}}(\tilde{\rho}) = \tilde{\chi} \circ \det$ where $\det : D_{\mathfrak{p}}^\times \to K_{\mathfrak{p}}^\times$ is the reduced norm (using $\tilde{\chi}$ and χ also to denote the characters of $K_{\mathfrak{p}}^\times$ to which they correspond via class field theory). We then define $\pi^{D_{\mathfrak{p}}}(\rho)$ to be $\chi \circ \det$, *unless* $q \equiv 1 \bmod \ell$ and ρ is the split representation $\chi \oplus \chi$ in which case we define $\pi^{D_{\mathfrak{p}}}(\rho)$ to be $\chi \circ \det \oplus \chi \circ \det$ (note that $\overline{\omega}$ is trivial).

Remark 4.3. Note that the reduction of $\pi^{D_{\mathfrak{p}}}(\tilde{\rho})$ is $\chi \circ \det$, which then coincides with $\pi^{D_{\mathfrak{p}}}(\rho)$ unless we are in the exceptional case where $q \equiv 1 \bmod \ell$ and ρ is scalar.

Work of Ribet [52] and Yang [66] suggests that in this exceptional case $\pi^{D_\mathfrak{p}}(\rho)$ will have dimension greater than 1; in particular the analogue of Theorem 4.1(4) fails.

Suppose now that $\rho \sim \left(\begin{smallmatrix} \chi & * \\ 0 & \chi\overline{\omega}^{-1} \end{smallmatrix}\right)$ for some character $\chi : G_{K_\mathfrak{p}} \to \overline{\mathbf{F}}_\ell^\times$ and that $q \equiv -1 \bmod \ell$. Let c_ρ denote the extension class associated to ρ in

$$\mathrm{Ext}^1_{\overline{\mathbf{F}}_\ell[G_{K_\mathfrak{p}}]}(\chi\overline{\omega}^{-1}, \chi) \cong H^1(G_{K_\mathfrak{p}}, \overline{\mathbf{F}}_\ell(\overline{\omega})).$$

Note that this space is 1-dimensional unless $\ell = 2$ in which case it is 2-dimensional. Recall that such ρ have irreducible lifts $\tilde{\rho}$ which are twists of tamely ramified representations induced from G_L where L is the unramified quadratic extension of $K_\mathfrak{p}$. For such $\tilde{\rho}$, $\pi^{D_\mathfrak{p}}(\tilde{\rho})$ is a two-dimensional representation of $D_\mathfrak{p}^\times$ whose reduction (which depends on a choice of lattice) has semi-simplification $\chi \circ \det \oplus (\chi\overline{\omega}^{-1}) \circ \det$ (see [62]). We will define $\pi^{D_\mathfrak{p}}(\rho)$ as a certain extension of $(\chi\overline{\omega}^{-1}) \circ \det$ by $\chi \circ \det$ depending on c_ρ. To this end we will first compute

$$\mathrm{Ext}^1_{\overline{\mathbf{F}}_\ell[D_\mathfrak{p}^\times]}((\chi\overline{\omega}^{-1}) \circ \det, \chi \circ \det) \cong H^1(D_\mathfrak{p}^\times, \overline{\mathbf{F}}_\ell(\overline{\omega} \circ \det)).$$

Although a unified treatment is possible (see Remark 4.4), it is simpler to consider separately the cases $\ell > 2$ and $\ell = 2$.

Suppose first that $\ell > 2$. Let $\mathcal{O}_{D_\mathfrak{p}}$ denote the maximal order in $D_\mathfrak{p}$ and Π a uniformizer, so $\mathrm{val}_{K_\mathfrak{p}}(\det \Pi) = 1$ and $\mathcal{O}_{D_\mathfrak{p}}/\Pi\mathcal{O}_{D_\mathfrak{p}} \cong \mathbf{F}_{q^2}$. Letting $\Gamma = D_\mathfrak{p}^\times/(1 + \Pi\mathcal{O}_{D_\mathfrak{p}})$, we have an exact sequence

$$1 \to \mathbf{F}_{q^2}^\times \to \Gamma \to \mathbf{Z} \to 0$$

where the map $\Gamma \to \mathbf{Z}$ is $\mathrm{val} \circ \det$ and $n \in \mathbf{Z}$ acts on $\mathbf{F}_{q^2}^\times$ by $x \mapsto x^{q^n}$. Note that $\mathbf{F}_{q^2}^\times$ acts trivially on $\overline{\mathbf{F}}_\ell(\overline{\omega} \circ \det)$ and the induced action of $n \in \mathbf{Z}$ is via the character $\mu(n) = (-1)^n = q^n$. Since $1 + \Pi\mathcal{O}_{D_\mathfrak{p}}$ is pro-p, we have that

$$H^1(D_\mathfrak{p}^\times, \overline{\mathbf{F}}_\ell(\overline{\omega} \circ \det)) \cong H^1(\Gamma, \overline{\mathbf{F}}_\ell(\overline{\omega} \circ \det)).$$

Since $H^1(\mathbf{Z}, \overline{\mathbf{F}}_\ell(\mu)) = H^2(\mathbf{Z}, \overline{\mathbf{F}}_\ell(\mu)) = 0$, we have that

$$H^1(\Gamma, \overline{\mathbf{F}}_\ell(\overline{\omega} \circ \det)) = \mathrm{Hom}_{\mathbf{Z}}(\mathbf{F}_{q^2}^\times, \overline{\mathbf{F}}_\ell(\mu))$$

is one-dimensional. Hence there is a unique isomorphism class of $\overline{\mathbf{F}}_\ell$-representations of $D_\mathfrak{p}^\times$ which are non-trivial extensions of $(\chi\overline{\omega}^{-1}) \circ \det$ by $\chi \circ \det$. We define $\pi^{D_\mathfrak{p}}(\rho)$ to be the extension which is trivial or not according to whether c_ρ is trivial or not.

Finally consider the case $\ell = 2$. Then $\overline{\omega}$ is trivial and

$$H^1(D_\mathfrak{p}^\times, \overline{\mathbf{F}}_2) = \mathrm{Hom}(D_\mathfrak{p}^\times/(D_\mathfrak{p}^\times)^2, \overline{\mathbf{F}}_2),$$

and one checks easily that det induces an isomorphism

$$D_\mathfrak{p}^\times/(D_\mathfrak{p}^\times)^2 \xrightarrow{\sim} K_\mathfrak{p}^\times/(K_\mathfrak{p}^\times)^2.$$

On the other hand local class field theory yields an isomorphism

$$H^1(G_{K_\mathfrak{p}}, \overline{\mathbf{F}}_2) \cong \mathrm{Hom}(K_\mathfrak{p}^\times/(K_\mathfrak{p}^\times)^2, \overline{\mathbf{F}}_2).$$

Putting these isomorphisms together yields

$$H^1(G_{K_\mathfrak{p}}, \overline{\mathbf{F}}_2) \xrightarrow{\sim} H^1(D_\mathfrak{p}^\times, \overline{\mathbf{F}}_2),$$

and we define $\pi^{D_\mathfrak{p}}(\rho)$ to be the extension obtained from the image of c_ρ.

Remark 4.4. To give a unified treatment for the cases $\ell = 2$ and $\ell > 2$ when $q \equiv -1 \bmod \ell$ and ρ as above, embed the unramified quadratic extension L of $K_\mathfrak{p}$ in $D_\mathfrak{p}$ and let N denote the normaliser of the image of $L^\times$ in $D_\mathfrak{p}^\times$. One can then check that restriction induces an isomorphism

$$H^1(D_\mathfrak{p}, \overline{\mathbf{F}}_\ell(\overline{\omega} \circ \det)) \cong H^1(N, \overline{\mathbf{F}}_\ell(\mu))$$

where $\mu(x) = \pm 1$ according to whether $x \in L^\times$. On the other hand one finds that N is isomorphic by local class field theory to the image of the Weil group of $K_\mathfrak{p}$ in $\mathrm{Gal}(L^{\mathrm{ab}}/K_\mathfrak{p})$ and that the inflation and restriction maps induce isomorphisms

$$H^1(G_{K_\mathfrak{p}}, \overline{\mathbf{F}}_\ell(\overline{\omega})) \cong H^1(\mathrm{Gal}(L^{\mathrm{ab}}/K_\mathfrak{p}), \overline{\mathbf{F}}_\ell(\overline{\omega})) \cong H^1(N, \overline{\mathbf{F}}_\ell(\mu)).$$

Remark 4.5. It is straightforward to check that if $q \equiv -1 \bmod \ell$ and ξ is a character of G_L of ℓ-power order, then every representation ρ with semisimplification $1 \oplus \overline{\omega}$ is isomorphic to the reduction of a $G_{K_\mathfrak{p}}$-stable lattice in $\mathrm{Ind}_{G_L}^{G_{K_\mathfrak{p}}} \xi$. It follows that if $\ell = 2$ (and in fact whenever $q \equiv 1 \bmod \ell$ for any ℓ), then every representation with scalar semisimplification has the same set of lifts $\tilde{\rho}$ such that $\pi(\tilde{\rho})$ is square-integrable, so $\pi^{D_\mathfrak{p}}(\rho)$ is not characterised by the set of $\pi^{D_\mathfrak{p}}(\tilde{\rho})$.

We now return to the global setting of a totally odd, continuous, irreducible

$$\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$$

and construct the $\overline{\mathbf{F}}_\ell$-representation of $D_f^\times = (D \otimes \hat{\mathbf{Z}})^\times$ whose local factors should be the $\pi^{D_\mathfrak{p}}(\rho)$. We first consider the case of a totally definite quaternion algebra D over K. Fix a maximal order $\mathcal{O}_D$ in D and isomorphisms $\mathcal{O}_{D,\mathfrak{p}} \cong \mathrm{M}_2(\mathcal{O}_{K,\mathfrak{p}})$ for each prime $\mathfrak{p}$ of K at which D is split.

For each open compact subgroup U of $D_f^\times$ we define

$$S^D(U) = \{\, f : D^\times \backslash D_f^\times / U \to \overline{\mathbf{F}}_\ell \,\}.$$

The obvious projection maps for $V \subset U$ induce inclusions $S^D(U) \to S^D(V)$ and we define S^D as the direct limit of the $S^D(U)$. Thus S^D can equivalently be defined as the set of smooth functions $f : D^\times \backslash D_f^\times \to \overline{\mathbf{F}}_\ell$. Note that this $\overline{\mathbf{F}}_\ell$-vector space admits a natural left action of $D_f^\times$ by right translation, and $S^D(U) = (S^D)^U$. Moreover for any $g \in D_f^\times$ and open compact $U, V \subset D_f^\times$ we have the double coset operator $[UgV] : S^D(V) \to S^D(U)$ defined in the usual way. In particular, for each prime $\mathfrak{p}$ at which D is split and $\mathrm{GL}_2(\mathcal{O}_{K,\mathfrak{p}}) \subset U$, we have the endomorphisms

$$T_\mathfrak{p} = \left[U \begin{pmatrix} \varpi_\mathfrak{p} & 0 \\ 0 & 1 \end{pmatrix} U \right] \quad \text{and} \quad S_\mathfrak{p} = \left[U \begin{pmatrix} \varpi_\mathfrak{p} & 0 \\ 0 & \varpi_\mathfrak{p} \end{pmatrix} U \right]$$

of $S^D(U)$, where $\varpi_\mathfrak{p}$ is any uniformizer of $\mathcal{O}_{K,\mathfrak{p}}$. If Σ is a finite set of primes of K containing all those such that:

- D is ramified at $\mathfrak{p}$,
- $\mathrm{GL}_2(\mathcal{O}_{K,\mathfrak{p}}) \not\subset U$,
- ρ is ramified at $\mathfrak{p}$, or
- $\mathfrak{p}$ divides ℓ,

then we let $\mathbf{T}^\Sigma(U)$ denote the commutative $\overline{\mathbf{F}}_\ell$-subalgebra of $\mathrm{End}_{\overline{\mathbf{F}}_\ell}(S^D(U))$ generated by the $T_\mathfrak{p}$ for $\mathfrak{p} \notin \Sigma$. We let $\mathfrak{m}_\rho^\Sigma = \mathfrak{m}_\rho^\Sigma(U)$ denote the ideal of $\mathbf{T}^\Sigma(U)$ generated by the operators

$$T_\mathfrak{p} - \mathrm{tr}(\rho(\mathsf{Frob}_\mathfrak{p})) \quad \text{and} \quad S_\mathfrak{p}\mathbf{N}(\mathfrak{p}) - \det(\rho(\mathsf{Frob}_\mathfrak{p}))$$

for all $\mathfrak{p} \notin \Sigma$. We let

$$S^D(U)[\mathfrak{m}_\rho^\Sigma] = \{ f \in S^D(U) \mid Tf = 0 \text{ for all } T \in \mathfrak{m}_\rho^\Sigma \}.$$

If $\rho = \overline{\rho}_\pi$ for some (necessarily cuspidal) automorphic representation π of $D^\times$ with weight $(\vec{2}, 0)$ and $\pi^U \neq 0$, then $\mathfrak{m}_\rho^\Sigma$ is a maximal ideal of $\mathbf{T}^\Sigma(U)$ and $S^D(U)[\mathfrak{m}_\rho^\Sigma] \neq 0$; otherwise $\mathfrak{m}_\rho^\Sigma = \mathbf{T}^\Sigma(U)$ and $S^D(U)[\mathfrak{m}_\rho^\Sigma] = 0$.

Lemma 4.6. *Suppose that D, U, ρ and Σ are as above. Then*
a) $S^D(U)[\mathfrak{m}_\rho^\Sigma]$ is independent of Σ (so we will denote it $S^D(U)[\mathfrak{m}_\rho]$);
b) if $g \in D_f^\times$ and V is an open compact subgroup of $D_f^\times$ such that $V \subset gUg^{-1}$, then g sends $S^D(U)[\mathfrak{m}_\rho]$ to $S^D(V)[\mathfrak{m}_\rho]$.

Proof. a) We may assume $\Sigma' = \Sigma \cup \{\mathfrak{p}\}$ for some $\mathfrak{p} \notin \Sigma$ and that $\mathfrak{m}_\rho^{\Sigma'} \neq \mathbf{T}^{\Sigma'}(U)$. Since ρ is irreducible, a standard argument using the representations ρ_π lifting ρ gives a representation

$$\rho' : G_K \to \mathrm{GL}_2(\mathbf{T}^{\Sigma'}(U)_{\mathfrak{m}_\rho^{\Sigma'}})$$

lifting ρ such that

$$T_\mathfrak{p} = \mathrm{tr}(\rho'(\mathrm{Frob}_\mathfrak{p})) \quad \text{and} \quad \mathbf{N}(\mathfrak{p})S_\mathfrak{p} = \det(\rho'(\mathrm{Frob}_\mathfrak{p}))$$

as endomorphisms of $S^D(U)_{\mathfrak{m}_\rho^{\Sigma'}}$. It follows that $S^D(U)[\mathfrak{m}_\rho^\Sigma] = S^D(U)[\mathfrak{m}_\rho^{\Sigma'}]$.

b) Choosing Σ sufficiently large that $\mathrm{GL}_2(\mathcal{O}_{K,\mathfrak{p}}) \subset V$ and $g_\mathfrak{p} \in \mathrm{GL}_2(\mathcal{O}_{K,\mathfrak{p}})$ for all $\mathfrak{p} \notin \Sigma$, we see that g commutes with $T_\mathfrak{p}$ for all $\mathfrak{p} \notin \Sigma$. □

We can now consider the direct limit over U of the spaces $S^D(U)[\mathfrak{m}_\rho]$; by the lemma, this makes sense and yields a representation $S^D[\mathfrak{m}_\rho]$ of $D_f^\times$.

Conjecture 4.7. *Suppose that K is a totally real field,*

$$\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$$

is a continuous, irreducible and totally odd representation, and D is a totally definite quaternion algebra over K. Then the $\overline{\mathbf{F}}_\ell$-representation $S^D[\mathfrak{m}_\rho]$ of $D_f^\times$ is isomorphic to a restricted tensor product $\otimes'_\mathfrak{p}\pi_\mathfrak{p}$ where $\pi_\mathfrak{p}$ is a smooth admissible representation of $D_\mathfrak{p}^\times$ such that

- *if $\mathfrak{p}$ does not divide ℓ then $\pi_\mathfrak{p} \cong \pi^{D_\mathfrak{p}}(\rho|_{G_{K_\mathfrak{p}}})$;*
- *if $\mathfrak{p}|\ell$ then $\pi_\mathfrak{p} \neq 0$; moreover if K and D are unramified at $\mathfrak{p}$, and σ is an irreducible $\overline{\mathbf{F}}_\ell$-representation of $\mathrm{GL}_2(\mathcal{O}_{K,\mathfrak{p}})$, then $\mathrm{Hom}_{\mathrm{GL}_2(\mathcal{O}_{K,\mathfrak{p}})}(\sigma, \pi_\mathfrak{p}) \neq 0$ if and only if $\sigma \in W_\mathfrak{p}(\rho^\vee|_{G_{K_\mathfrak{p}}})$.*

Remark 4.8. In the case $K = \mathbf{Q}$ and $D = M_2(\mathbf{Q})$ and $\rho|_{G_{\mathbf{Q}_\ell}}$ is not a twist of a representation of the form $\left(\begin{smallmatrix} 1 & * \\ 0 & \omega \end{smallmatrix}\right)$, then Emerton predicts the precise form for π_ℓ in [22] as well and goes on to prove the conjecture under technical hypotheses. It is reasonable to expect that $\pi_\mathfrak{p}$ is of the form predicted there whenever $D_\mathfrak{p} \cong M_2(\mathbf{Q}_\ell)$ and $\rho|_{G_{K_\mathfrak{p}}}$ is as above.

Under the hypotheses that $K_\mathfrak{p}$ is an unramified extension of $\mathbf{Q}_\ell$ and D is split at $\mathfrak{p}$, the relation with $W_\mathfrak{p}$ can be viewed as a description of the $\mathrm{GL}_2(\mathcal{O}_{K,\mathfrak{p}})$-socle of $\pi_\mathfrak{p}$ (which in most cases is expected to be multiplicity-free). In many cases, Breuil and Paskunas [7] construct infinitely many (isomorphism classes of) representations with the desired socle, raising the question of whether one should still expect $\pi_\mathfrak{p}$ to be completely determined by $\rho|_{G_{K_\mathfrak{p}}}$.

Some work has also been done towards defining a conjectural set of weights $W_{\mathfrak{p}}$ when $F_{\mathfrak{p}}$ is ramified extension of $\mathbf{Q}_\ell$ and $D_{\mathfrak{p}}$ is split. In particular, Schein [55] gives a definition of $W_{\mathfrak{p}}$ when $\rho|_{G_{K_{\mathfrak{p}}}}$ is tamely ramified, and Gee [33] gives a more general but less explicit definition than ours.

Suppose now that D is split at exactly one infinite place. We exclude the case $D = \mathrm{M}_2(\mathbf{Q})$ already considered by Emerton. We now define $S^D(U) = H^1(Y_{U,\bar{K}}, \overline{\mathbf{F}}_\ell)$ and $S^D = \lim S^D(U)$, the limit taken over all open compact $U \subset D_f^\times$ with respect to the maps on cohomology induced by the natural projections $Y_V \to Y_U$ for $U \subset V$. If $V \subset gUg^{-1}$, then we have a K-morphism $Y_V \to Y_U$ corresponding to right multiplication by g on complex points, inducing a homomorphism $S^D(U) \to S^D(V)$ which we also denote by g, making S^D a $G_K \times D_f^\times$-module. However the natural map $S^D(U) \to (S^D)^U$ is not necessarily an isomorphism.

For $g \in D_f^\times$, and U,V open compact subgroups of $D_f^\times$, we have the double coset operator $[VgU] : S^D(U) \to S^D(V)$ defined as the composite $S^D(U) \to S^D(V') \to S^D(V)$ where $V' = V \cap gUg^{-1}$, the first map is defined by g, and the second is the trace morphism times the integer $[V : V']/\deg(Y_{V'}/Y_V)$. We can thus define Hecke operators $T_{\mathfrak{p}}$ for $\mathfrak{p} \notin \Sigma$, algebras $\mathbf{T}^\Sigma(U) \subset \mathrm{End}_{\overline{\mathbf{F}}_\ell}(S^D(U))$, ideals $\mathfrak{m}_\rho^\Sigma$ and subspaces $S^D(U)[\mathfrak{m}_\rho^\Sigma] \subset S^D(U)$, just as in the case of totally definite D. The analogue of Lemma 4.6 is proved in exactly the same way, now yielding a representation $S^D[\mathfrak{m}_\rho]$ of $G_K \times D_f^\times$.

Conjecture 4.9. *Suppose that K is a totally real field,*

$$\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$$

is a continuous, irreducible and totally odd representation, and D is a quaternion algebra over K split at exactly one real place. Then the $\overline{\mathbf{F}}_\ell$-representation $S^D[\mathfrak{m}_\rho]$ of $G_K \times D_f^\times$ is isomorphic to $\rho \otimes \left(\otimes'_{\mathfrak{p}} \pi_{\mathfrak{p}}\right)$ where $\pi_{\mathfrak{p}}$ is a smooth admissible representation of $D_{\mathfrak{p}}^\times$ such that

- *if $\mathfrak{p}$ does not divide ℓ then $\pi_{\mathfrak{p}} \cong \pi^{D_{\mathfrak{p}}}(\rho|_{G_{K_{\mathfrak{p}}}})$;*
- *if $\mathfrak{p}|\ell$ then $\pi_{\mathfrak{p}} \neq 0$; moreover if K and D are unramified at $\mathfrak{p}$, and σ is an irreducible $\overline{\mathbf{F}}_\ell$-representation of $\mathrm{GL}_2(\mathcal{O}_{K,\mathfrak{p}})$, then $\mathrm{Hom}_{\mathrm{GL}_2(\mathcal{O}_{K,\mathfrak{p}})}(\sigma, \pi_{\mathfrak{p}}) \neq 0$ if and only if $\sigma \in W_{\mathfrak{p}}(\rho^\vee|_{G_{K_{\mathfrak{p}}}})$.*

By the following lemma, Conjecture 4.9 could be reformulated as saying that the representation $\mathrm{Hom}_{\overline{\mathbf{F}}_\ell[G_K]}(\rho, S^D)$ of $D_f^\times$ has the prescribed form, as in [22].

Lemma 4.10. *The evaluation map $\rho \otimes_{\overline{\mathbf{F}}_\ell} \mathrm{Hom}_{\overline{\mathbf{F}}_\ell[G_K]}(\rho, S^D) \to S^D$ induces a $G_K \times D_f^\times$-linear isomorphism:*

$$\rho \otimes_{\overline{\mathbf{F}}_\ell} \mathrm{Hom}_{\overline{\mathbf{F}}_\ell[G_K]}(\rho, S^D) \xrightarrow{\sim} S^D[\mathfrak{m}_\rho].$$

Proof. It suffices to prove the lemma with S^D replaced by $S^D(U)$ and take direct limits. Since ρ is irreducible, the evaluation map

$$\rho \otimes_{\overline{\mathbf{F}}_\ell} \mathrm{Hom}_{\overline{\mathbf{F}}_\ell[G_K]}(\rho, S^D(U)) \to S^D(U)$$

is injective by Schur's Lemma. Using the Eichler-Shimura relations on Y_U (in particular, that $\mathrm{Frob}_{\mathfrak{p}}^2 + T_{\mathfrak{p}} \mathrm{Frob}_{\mathfrak{p}} + \mathbf{N}(\mathfrak{p}) S_{\mathfrak{p}} = 0$ on $H^1(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell)$ for all $\mathfrak{p} \notin \Sigma$), one shows as in the proof of Prop. 6.1.7 of [23] that the image lies in $S^D(U)[\mathfrak{m}_\rho]$. Finally, the main result of [6] shows that $S^D(U)[\mathfrak{m}^\rho]$ is isomorphic to a direct sum of copies of ρ, hence the map is surjective. $\square$

Next we show how one can usually recover $S^D(U)[\mathfrak{m}_\rho]$ from $S^D[\mathfrak{m}_\rho]$. The caveat (an observation going back to Ribet) is that it is not quite true that "all errors are Eisenstein". Let us say that a representation ρ is *badly dihedral* if ρ is induced from a character of $G_{K'}$ where K' is a totally imaginary quadratic extension of K of the form $K(\delta)$ for some δ such that $\delta^\ell \in K$. For $\ell > 2$ it is not difficult to check that if ρ is badly dihedral then K must contain $\mathbf{Q}(\mu_\ell)^+$ and $K' = K(\mu_\ell)$ (the field K' contains $\zeta := \overline{\delta}/\delta$ with $\overline{\delta}$ the Galois conjugate of δ, and $\zeta \neq 1$ is an ℓth root of unity). In particular if $\ell > 3$ and ℓ is unramified in K, then there will be no badly dihedral representations at all. However for $\ell = 2$ there may be more than one possibility for K' (but only finitely many).

Lemma 4.11. *The natural map $S^D(U)[\mathfrak{m}_\rho] \to (S^D[\mathfrak{m}_\rho])^U$ is injective; moreover it is an isomorphism if either*

- *Y_U has no elliptic points of order a multiple of ℓ or*
- *ρ is not badly dihedral.*

Proof. It suffices to show that if V is any normal open compact subgroup of U then the natural map

$$H^1(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell) \to H^1(Y_{V,\overline{K}}, \overline{\mathbf{F}}_\ell)^{U/V}$$

is injective after localising at $\mathfrak{m}_\rho$, and is an isomorphism under the additional hypotheses. Equivalently we must show that ρ does not appear in the $\overline{\mathbf{F}}_\ell[G_K]$-semisimplification of the kernel of the above map, and under the additional hypotheses it does not appear in the cokernel either.

Let Z_U denote the reduced closed subscheme of Y_U defined by its elliptic points and let $Z_V = (Z_U \times_{Y_U} Y_V)^{\mathrm{red}}$, $W_U = Y_U - Z_U$ and $W_V = Y_V - Z_V$. Then the morphism $W_V \to W_U$ is étale with Galois group Γ, a quotient of U/V, and the Hochschild-Serre spectral sequence yields an exact sequence:

$$\begin{aligned} 0 \to H^1(\Gamma, H^0(W_{V,\overline{K}}, \overline{\mathbf{F}}_\ell)) &\to H^1(W_{U,\overline{K}}, \overline{\mathbf{F}}_\ell) \\ &\to H^1(W_{V,\overline{K}}, \overline{\mathbf{F}}_\ell)^U \to H^2(\Gamma, H^0(W_{V,\overline{K}}, \overline{\mathbf{F}}_\ell)). \end{aligned}$$

The inclusion $W_U \to Y_U$ yields an exact sequence:

$$\begin{aligned} H^1_{Z_{U,\overline{K}}}(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell) \to \quad & H^1(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell) \\ \to \quad & H^1(W_{U,\overline{K}}, \overline{\mathbf{F}}_\ell) \to H^2_{Z_{U,\overline{K}}}(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell) \end{aligned}$$

By the excision theorem, $H^i_{Z_{U,\overline{K}}}(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell) = \bigoplus_{z \in Z_U(\bar{K})} H^i_{\{z\}}(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell)$. By the Betti-étale comparison theorem for example, we see that each $H^1_{\{z\}}(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell) = 0$ and that each $H^2_{\{z\}}(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell)$ is one-dimensional; moreover if $\psi : X' \to X$ is a non-constant morphism of smooth proper curves over $\overline{K}$ with $\psi(x') = x$, then the induced map $H^2_{\{x\}}(X, \overline{\mathbf{F}}_\ell) \to H^2_{\{x'\}}(X', \overline{\mathbf{F}}_\ell)$ is trivial or an isomorphism according to whether the ramification degree $e(x'/x)$ is divisible by ℓ. In particular, if $z \in Z_U(\overline{K})$ is defined over L, then the morphism $Y_{U,L} \to \mathbf{P}^1_L$ gotten from a uniformizer at z induces an isomorphism

$$\overline{\mathbf{F}}_\ell(1) = H^2_{\{0\}}(\mathbf{P}^1_{\overline{L}}, \overline{\mathbf{F}}_\ell) \xrightarrow{\sim} H^2_{\{z\}}(Y_{U,\overline{L}}, \overline{\mathbf{F}}_\ell)$$

of G_L-modules. It follows that $H^2_{Z_{U,\overline{K}}}(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell) \cong \bigoplus_{P \in Z_U} \mathrm{Ind}_{G_{K(P)}}^{G_K} \overline{\mathbf{F}}_\ell(1)$ as G_K-modules. Combining this with the corresponding exact sequence arising from

$W_V \to Y_V$ yields a commutative diagram

$$\begin{array}{ccccccc} 0 \to & H^1(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell) & \to & H^1(W_{U,\overline{K}}, \overline{\mathbf{F}}_\ell) & \to & H^2_{Z_{U,\overline{K}}}(Y_{U,\overline{K}}, \overline{\mathbf{F}}_\ell) \\ & \downarrow & & \downarrow & & \downarrow \\ 0 \to & H^1(Y_{V,\overline{K}}, \overline{\mathbf{F}}_\ell)^\Gamma & \to & H^1(W_{V,\overline{K}}, \overline{\mathbf{F}}_\ell)^\Gamma & \to & H^2_{Z_{V,\overline{K}}}(Y_{V,\overline{K}}, \overline{\mathbf{F}}_\ell)^\Gamma \end{array}$$

such that the kernel of the rightmost vertical map is isomorphic to the direct sum of the $\mathrm{Ind}_{G_{K(P)}}^{G_K} \overline{\mathbf{F}}_\ell(1)$ over the $P \in Y_U$ whose ramification degree in Y_V is divisible by ℓ. If P is such an elliptic point, then it is fixed by some $\delta \in D$ such that δ has order ℓ in $D^\times/K^\times$. The extension $K' := K(\delta)$ is a commutative integral domain within D, and it is finite over K and hence a field; moreover it must be a quadratic extension of K, imaginary at our preferred infinite place τ_0 since δ has isolated fixed points in $\mathfrak{H}^\pm$, and at the other infinite places since $K' \subset D$ and hence K' splits D. The elliptic point will then be a special point for the Shimura curve Y_U with respect to the torus $\mathrm{Res}_{K'/\mathbf{Q}}(\mathbf{G}_m)$ and by Lemma 3.11 of [13] the elliptic point will be defined over an abelian extension of K'. Now under the additional hypotheses of the lemma it follows that ρ does not appear in the $\overline{\mathbf{F}}_\ell[G_K]$-semisimplification of the direct sum of the $\mathrm{Ind}_{G_{K(P)}}^{G_K} \overline{\mathbf{F}}_\ell(1)$ as above.

Recall from Lemma 2.2 that the action of G_K on $H^0(W_{V,\overline{K}}, \overline{\mathbf{F}}_\ell) = H^0(Y_{V,\overline{K}}, \overline{\mathbf{F}}_\ell)$, hence on the kernel and cokernel of the middle vertical map, factors through an abelian quotient. Finally we deduce from the snake lemma that ρ does not appear in the semisimplification of the kernel of the leftmost vertical map, nor that of the cokernel under the additional hypotheses. $\square$

Finally we record some consequences of Conjectures 4.7 and 4.9.

Proposition 4.12. *Conjecture 4.9 implies Conjecture 3.12.*

Proof. Since the conjecture is known for $\mathbf{Q}$, we can assume $K \neq \mathbf{Q}$. Now ρ is modular of weight σ if and only if $\rho(-1)$ is isomorphic to an $\overline{\mathbf{F}}_\ell[G_K]$-subquotient of $\mathrm{Hom}_{\overline{\mathbf{F}}_\ell[U]}(\sigma^\vee, H^1(Y_{U',\overline{K}}, \overline{\mathbf{F}}_\ell))$ for some D and U as in Definition 2.1. This is equivalent to saying that $\mathfrak{m}^\Sigma_{\rho(-1)}$ (for any Σ at level U') is (maximal and) in the support of $\mathrm{Hom}_{\overline{\mathbf{F}}_\ell[U]}(\sigma^\vee, S^D(U'))$, or equivalently that

$$\mathrm{Hom}_{\overline{\mathbf{F}}_\ell[U]}(\sigma^\vee, S^D(U')[\mathfrak{m}_{\rho(-1)}]) \neq 0.$$

(Note in particular that by the proof of Lemma 4.10, "$\overline{\mathbf{F}}_\ell G_K$-subquotient" can be replaced with "$\overline{\mathbf{F}}_\ell G_K$-submodule" as claimed after Definition 2.1.)

So if ρ is modular of an irreducible weight σ, then $\mathrm{Hom}_{\overline{\mathbf{F}}_\ell[U]}(\sigma^\vee, S^D[\mathfrak{m}_{\rho(-1)}]) \neq 0$ by Lemma 4.11. If Conjecture 4.9 holds, then we may write $S^D[\mathfrak{m}_{\rho(-1)}] = \rho \otimes (\otimes' \pi_\mathfrak{p})$; moreover for each $\mathfrak{p}|\ell$, we have $\mathrm{Hom}_{\overline{\mathbf{F}}_\ell[U_\mathfrak{p}]}(\sigma_\mathfrak{p}^\vee, \pi_\mathfrak{p}) \neq 0$ so that $\sigma_\mathfrak{p}^\vee \in W_\mathfrak{p}(\rho(-1)^\vee)$, or equivalently that $\sigma^\vee \in W(\rho(-1)^\vee)$. Since $\rho \cong \det(\rho) \otimes \rho^\vee$ and $\sigma \cong \psi\sigma^\vee$ where ψ is the central character of σ, it follows easily from Proposition 3.13 that this is equivalent to $\sigma \in W(\rho)$.

Conversely suppose that $\sigma \in W(\rho)$. If $[K : \mathbf{Q}]$ is odd, then let D be a quaternion algebra over K ramified at all but one infinite places and split at all finite places. If $[K : \mathbf{Q}]$ is even, then let L denote the splitting field of ρ and choose a prime $\mathfrak{q}$ unramified in $L(\mu_\ell)$ so that the conjugacy class of $\mathsf{Frob}_\mathfrak{q}$ in $\mathrm{Gal}(L(\mu_\ell)/K)$ is that of a complex conjugation. Let D be a quaternion algebra over K ramified at exactly $\mathfrak{q}$ and all but one infinite place. Then $\pi^{D_\mathfrak{p}}(\rho) \neq 0$ for all primes $\mathfrak{p}$, so we can choose U

sufficiently small (of level prime to ℓ) so that Y_U has no elliptic points and $\pi_{\mathfrak{p}}^{U_{\mathfrak{p}}} \neq 0$ for all $\mathfrak{p}$ not dividing ℓ. We can then reverse the above argument to conclude that ρ is modular of weight σ. □

Level-lowering theorems of Fujiwara [28], Rajaei [50] and the third author [37, 38] can be viewed as partial results in the direction of Conjectures 4.7 and 4.9, of which they are also consequences.

Proposition 4.13. *Let $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ be continuous, irreducible and totally odd, let $\mathfrak{n}$ be the (prime to ℓ) conductor of ρ and let $\mathfrak{n}' = \mathfrak{n}\prod_{\mathfrak{p}|\ell}\mathfrak{p}^2$.*
a) Suppose that Conjecture 4.7 holds, or that Conjecture 4.9 holds and ρ is not badly dihedral. Then $\rho \sim \overline{\rho}_\pi$ for some cuspidal automorphic representation π of GL_2/K of weight $(\vec{2},0)$ and conductor dividing $\mathfrak{n}'$.
b) If $[K:\mathbf{Q}]$ is even, suppose that Conjecture 4.7 holds for K; if $[K:\mathbf{Q}]$ is odd, suppose that Conjecture 4.9 holds for K and that ρ is not badly dihedral if $\ell = 2$ or 3. If K is unramified at ℓ and $\rho|_{G_{K_\mathfrak{p}}}$ arises from a finite flat group scheme over $\mathcal{O}_{K,\mathfrak{p}}$ for each $\mathfrak{p}|\ell$, then $\rho \sim \overline{\rho}_\pi$ for some cuspidal automorphic representation π of GL_2/K of weight $(\vec{2},2)$ and conductor $\mathfrak{n}$.

Proof. To prove (a) assuming Conjecture 4.7, let D be a quaternion algebra over K ramified at all infinite places and at most one prime $\mathfrak{p}_0$ over ℓ, but no other finite places. Let $U = \prod_{\mathfrak{q}} U_{\mathfrak{q}}$ be the open compact subgroup of $D_f^\times$ with

$$U_{\mathfrak{q}} = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathcal{O}_{K,\mathfrak{q}}) \,\middle|\, c \equiv d-1 \equiv 0 \bmod \mathfrak{n}\mathcal{O}_{K,\mathfrak{q}} \right\}$$

for $\mathfrak{q}$ not dividing ℓ, and $U_{\mathfrak{q}}$ a pro-ℓ-Sylow subgroup of a maximal compact subgroup of $D_{\mathfrak{q}}^\times$ for $\mathfrak{q}|\ell$. It follows from Emerton's characterisation of $\pi_{\mathfrak{q}}$ in Theorem 4.1 that $\pi_{\mathfrak{q}}^{U_{\mathfrak{q}}} \neq 0$ for all $\mathfrak{q}$ not dividing ℓ. The same is true for $\mathfrak{q}|\ell$ since $U_{\mathfrak{q}}$ is pro-ℓ and $\pi_{\mathfrak{q}} \neq 0$. Therefore $S^D(U)[\mathfrak{m}_\rho] = S^D[\mathfrak{m}_\rho]^U \neq 0$ and $\rho \sim \overline{\rho}_{\pi'}$ for some cuspidal automorphic representation π' of $D^\times$ of weight $(\vec{2},0)$ with $\pi'^U \neq 0$. Then the cuspidal automorphic representation π of GL_2/K corresponding to π via Jacquet-Langlands has conductor dividing $\mathfrak{n}'$.

The proof of (a) assuming Conjecture 4.9 is similar, except that we use the assumption that ρ is not badly dihedral in order to apply Lemma 4.11.

The proof of (b) is also similar, using the fact that if $\rho|_{G_{K_\mathfrak{p}}}$ arises from a finite flat group scheme, then $W_{\mathfrak{p}}(\rho)$ contains the trivial representation. (Recall that we are assuming ℓ to be unramified in K, so that badly dihedral representations only occur if $\ell = 2$ or 3, as remarked above.) □

We remark that in fact the conclusions of Proposition 4.13 follow from weaker modularity conjectures than Conjectures 4.7 and 4.9, together with known level lowering results, at least if $\ell > 2$ and ρ is not badly dihedral. Indeed, we will explain that weak modularity (Conjecture 1.1) implies Proposition 4.13(a), given these level lowering results. It seems that Proposition 4.13(b) is more subtle in that it requires more control over the level at ℓ, but these follow from Conjecture 1.1 and our weight conjecture (Conjecture 3.12), together with level lowering. (In fact, the only part of Conjecture 3.12 we need is the case where ρ is finite at $\mathfrak{p}$, so that $W_{\mathfrak{p}}(\rho)$ contains the trivial representation.)

For (a), we can assume that $\rho \sim \overline{\rho}_\pi$ for some automorphic representation π of GL_2/K of some weight and level, by weak modularity. By Corollary 2.12, we can assume that the weight is $(\vec{2},0)$, and that the level is $\mathfrak{m}\prod_{\mathfrak{p}|\ell}\mathfrak{p}^{a_\mathfrak{p}}$ for some ideal $\mathfrak{m}$

and some integers $a_{\mathfrak{p}}$. The same argument as in the proof of Proposition 4.13(a) above gives that $a_{\mathfrak{p}} \leq 2$. For primes $\mathfrak{q} \nmid \ell$, we may use existing level lowering results to deduce that we may take $\mathfrak{m} = \mathfrak{n}$, so that ρ is modular of weight $(\vec{2}, 0)$ and level $\mathfrak{n}'$, as required. For $[K : \mathbf{Q}]$ odd, these are due to the third author and to Rajaei [37, 38, 50], under the technical hypothesis that $\ell > 2$ and ρ is not badly dihedral. When $[K : \mathbf{Q}]$ is even, a similar argument applies, but in order to use the level lowering results mentioned above, one needs to begin by raising the level by adding a prime, using Taylor's theorem [59]. One can then switch to an appropriate quaternion algebra to perform the level lowering, and finally remove the prime that we added, using Fujiwara's unpublished version [28] of Mazur's Principle in the case $[K : \mathbf{Q}]$ even.

Proposition 4.13(b) would work in the same way, given sufficiently strong level lowering statements for primes $\mathfrak{p}|\ell$. However, these are not yet sufficient to deduce (b) from weak modularity and level lowering. But Conjecture 1.1 and Conjecture 3.12, together with level lowering (and the results of Taylor and Fujiwara when $[K : \mathbf{Q}]$ is even), is sufficient to deduce Proposition 4.13(b); one simply uses the observation made in the course of the proof above that the trivial representation lies in $W_{\mathfrak{p}}(\rho)$ if ρ is finite at $\mathfrak{p}$.

Corollary 4.14. *If Conjecture 4.7 or 4.9 holds, then there are only finitely many continuous, irreducible, totally odd $\rho : G_K \to \mathrm{GL}_2(\overline{\mathbf{F}}_\ell)$ of conductor dividing $\mathfrak{n}$.*

Proof. Note that by class field theory, there are only finitely many badly dihedral ρ of a given conductor. We can therefore assume ρ is not badly dihedral, so by Proposition 4.13(a), either conjecture implies ρ is modular of weight $(\vec{2}, 0)$ and level $\ell^2.\mathfrak{n}(\rho)$, where $\mathfrak{n}(\rho)$ denotes the Artin conductor of ρ. Since there are only finitely many automorphic representations of weight $(\vec{2}, 0)$ and given bounded level, the result follows. □

Corollary 4.15. *Suppose that Conjecture 4.7 holds if $[K : \mathbf{Q}]$ is even, and Conjecture 4.9 holds if $[K : \mathbf{Q}]$ is odd. If E is an elliptic curve over K, then E is modular.*

Proof. Given E of conductor $\mathfrak{n}$, let ℓ run through all primes greater than 3 and unramified in K, such that E has good reduction at all $\mathfrak{p}|\ell$. Then $\rho_{E,\ell}$ is finite at $\mathfrak{p}$, so Proposition 4.13(b) implies that $\rho_{E,\ell}$ is modular of weight $(\vec{2}, 2)$ and level equal to $\mathfrak{n}(\rho_{E,\ell})$, which divides the conductor of E. So there is an automorphic representation $\pi^{(\ell)}$ of level $U_1(\mathfrak{n})$ and weight $(\vec{2}, 2)$ whose mod ℓ representation agrees with $\rho_{E,\ell}$, or equivalently one of weight $(\vec{2}, 0)$ giving rise to $\rho_{E,\ell}(-1)$. There are only finitely many such automorphic representations, so there is a π such that $\pi = \pi^{(\ell)}$ for infinitely many ℓ. It follows that for all $\mathfrak{p}$ not dividing $\mathfrak{n}$, the action of $T_{\mathfrak{p}}$ on $\pi^{U_1(\mathfrak{n})}$ is by $a_{\mathfrak{p}}(E)$ and that of $S_{\mathfrak{p}}$ is trivial. Therefore $\rho_{E,\ell}(-1) \sim \rho_\pi$ (for any ℓ), and hence $L(E, s) = L(\pi, s)$. □

Remark 4.16. The remarks after Proposition 4.13 show that the conclusion of Corollary 4.14 for $\ell > 2$ actually follows from Conjecture 1.1 and known level lowering results. Similarly one sees that the conclusion of Corollary 4.15 follows from Conjecture 1.1, Conjecture 3.12 and level lowering results. In fact the modularity of E even follows from Conjecture 1.1 using modularity lifting results of Skinner-Wiles [58], Fujiwara [27] or Taylor [60]. Furthermore, using the lifting results of Kisin [46] and Gee [30] one can show (unconditionally) that if $\rho_{E,3}$ is irreducible and not badly dihedral, then E is modular. See also [57], [58] and [40] for additional cases where modularity of E is known.

References

[1] A.Ash, D.Doud, D.Pollack, Galois representations with conjectural connections to arithmetic cohomology, Duke Math. J. 112 (2002) 521–579.

[2] A.Ash, W.Sinnott, An analogue of Serre's conjecture for Galois representations and Hecke eigenclasses in the mod p cohomology of $\mathrm{GL}(n, \mathbf{Z})$, Duke Math. J. 105 (2000) 1–24.

[3] A.Ash, G.Stevens, Modular forms in characteristic ℓ and special values of their L-functions, Duke Math. J. 53 (1986) 849–868.

[4] A.Ash, G.Stevens, Cohomology of arithmetic groups and congruences between systems of Hecke eigenvalues, J. reine angew. Math. 365 (1986) 192-220.

[5] S.Bloch, K.Kato, L-functions and Tamagawa numbers of motives, Grothendieck Festschrift, Vol.1, Progr. Math. 86, Birkhäuser, Boston (1990) 333–400.

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

[7] C.Breuil, V.Paskunas, Towards a modulo p Langlands correspondence for GL_2, preprint.

[8] K.Buzzard, On level lowering for mod 2 representations, Mathematical Research Letters 7 (2000) 95–110.

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

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

[11] R.Coleman, J.-F.Voloch, Companion forms and Kodaira-Spencer theory, Invent. Math. 110 (1992) 263–282.

[12] P.Colmez, Représentations de $\mathrm{GL}_2(\mathbf{Q}_p)$ et (ϕ, Γ)-modules, preprint.

[13] C.Cornut, V.Vatsal, CM points and quaternion algebras, Documenta Math. 10 (2005) 263–309.

[14] P.Deligne, M.Rapoport, Les schémas de modules de courbes elliptiques, Lecture Notes in Mathematics 349, Springer-Verlag (1971) 143–317.

[15] L.Dembélé, F.Diamond, D.Roberts, Numerical evidences and examples of Serre's conjecture over totally real fields (in preparation).

[16] F.Diamond, The refined conjecture of Serre, in Elliptic Curves and Fermat's Last Theorem, J.Coates and S.-T.Yau (eds.), Hong Kong 1993, International Press, 2nd. ed. (1997) 172–186.

[17] F.Diamond, A correspondence between representations of local Galois groups and Lie-type groups, in L-functions and Galois representations, D.Burns, K.Buzzard and J.Nekovar (eds.), Cambridge University Press (2007) 187–206.

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

[19] L.Dieulefait, Existence of families of Galois representations and new cases of the Fontaine-Mazur conjecture, J. reine angew. Math. 577 (2004) 147–151.

[20] B.Edixhoven, The weight in Serre's conjectures on modular forms, Invent. Math. 109 (1992) 563–594.

[21] M. Emerton, A local-global compatibility conjecture in the p-adic Langlands programme for $\mathrm{GL}_2/\mathbf{Q}$, Pure Appl. Math. Q. 2 (2006), no. 2, part 2, 279–393.

[22] M.Emerton, Local-global compatibility in the p-adic Langlands programme for $\mathrm{GL}_{2,\mathbf{Q}}$, in preparation.

[23] M.Emerton, The local Langlands correspondence for $\mathrm{GL}_2(\mathbf{Q}_\ell)$ in p-adic families, and local-global compatibility for mod p and p-adic modular forms, in preparation.

[24] G.Faltings, B.Jordan, Crystalline cohomology and $\mathrm{GL}(2, \mathbf{Q})$. Israel J. Math. 90 (1995) 1–66.

[25] J.-M.Fontaine, G.Laffaille, Construction de représentations p-adiques, Ann. Sci. Ecole Norm. Sup. 15 (1982) 547–608.

[26] E. Freitag and R. Kiehl, Etale Cohomology and the Weil Conjecture, Ergebnisse der Mathematik und ihrer Grenzgebiete 13, Springer-Verlag (1988).

[27] K.Fujiwara, Deformation rings and Hecke algebras in the totally real case, preprint.

[28] K.Fujiwara, Level optimization in the totally real case, preprint.

[29] T.Gee, Companion forms over totally real fields, Manuscripta Math. 125 (2008) 1–41.

[30] T.Gee, A modularity lifting theorem for weight two Hilbert modular forms, Math. Res. Lett. 13, (2006) 805-811.
[31] T.Gee, Companion forms over totally real fields II, Duke Math. J. 136 (2007) 275–284.
[32] T.Gee, On the weights of mod p Hilbert modular forms, preprint.
[33] T.Gee, Automorphic lifts of prescribed types, preprint.
[34] B.Gross, A tameness criterion for Galois representations associated to modular forms (mod p), Duke Math. J. 61 (1990) 445–516.
[35] F.Herzig, The weight in a Serre-type conjecture for tame n-dimensional Galois representations, Harvard Ph.D. thesis, 2006.
[36] F.Jarvis, On Galois representations associated to Hilbert modular forms of low weight, J. reine angew. Math. 491 (1997) 199–216.
[37] F.Jarvis, Mazur's Principle for totally real fields of odd degree, Compositio Math. 116 (1999) 39–79.
[38] F.Jarvis, Level lowering for modular mod l representations over totally real fields, Math. Ann. 313 (1999) 141–160.
[39] F.Jarvis, Correspondences on Shimura curves and Mazur's principle at p, Pacific J. Math. 213 (2004) 267–280.
[40] F.Jarvis, J.Manoharmayum, On the modularity of supersingular elliptic curves over certain totally real number fields, J. Number Theory 128 (2008) 589–618.
[41] C.Khare, A local analysis of congruences in the (p, p) case II, Invent. Math. 143 (2001) 129–155.
[42] C.Khare, Serre's modularity conjecture: the level one case, Duke Math. J. 134 (2006) 557–589.
[43] C.Khare, J.-P.Wintenberger, On Serre's conjecture for 2-dimensional mod p representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, to appear in Ann. Math.
[44] C.Khare, J.-P.Wintenberger, Serre's modularity conjecture (I), preprint.
[45] C.Khare, J.-P.Wintenberger, Serre's modularity conjecture (II), preprint.
[46] M.Kisin, Moduli of finite flat group schemes, and modularity, to appear in Ann. Math.
[47] M.Kisin, Modularity of 2-adic Barsotti-Tate representations, preprint.
[48] J.S.Milne, Etale Cohomology, Princeton University Press (1980).
[49] M.Ohta, Hilbert modular forms of weight one and Galois representations, Progr. in Math. 46 (1984) 333–353.
[50] A.Rajaei, On levels of mod ℓ Hilbert modular forms, J. reine angew. Math. 537 (2001) 33–65.
[51] K.Ribet, On modular representations of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ arising from modular forms, Invent. Math. 100 (1990) 431–476.
[52] K.Ribet, Multiplicities of Galois representations in Jacobians of Shimura curves, Piatetski-Shapiro Festschrift, Part II, Israel Math. Conf. Proc. 3 (1989) 221–236.
[53] J.Rogawski, J.Tunnell, On Artin L-functions associated to Hilbert modular forms of weight 1, Invent. Math. 74 (1983) 1–42.
[54] M.Schein, Weights of Galois representations associated to Hilbert modular forms, to appear in J. reine angew. Math.
[55] M.Schein, Weights in Serre's conjecture for Hilbert modular forms: the ramified case, Israel J. Math. 166 (2008) 369–391.
[56] J.-P.Serre, Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, Duke Math. J. 54 (1987) 179–230.
[57] C.Skinner, A.Wiles, Modular forms and residually reducible representations, Publ. Math. IHES 89 (1999) 5–126.
[58] C.Skinner, A.Wiles, Nearly ordinary deformations of irreducible residual representations, Ann. Fac. Sci. Toulouse 10 (2001) 185–215.
[59] R.Taylor, On Galois representations associated to Hilbert modular forms, Invent. Math. 98 (1989) 265–280.
[60] R.Taylor, Remarks on a conjecture of Fontaine and Mazur, J. Inst. Math. Jussieu 1 (2002) 1–19.
[61] R.Taylor, A.Wiles, Ring-theoretic properties of certain Hecke algebras, Ann. Math. 141 (1995) 553–572.
[62] M.-F.Vignéras, Correspondance modulaire galois-quaternions pour un corps p-adique, Journées Arithmétiques d'Ulm, Springer-Verlag Lecture Notes 1380 (1989) 254–266.
[63] M.-F.Vignéras, Correspondance de Langlands semi-simple pour $\mathrm{GL}(n, F)$ modulo $\ell \neq p$, Invent. Math. 144 (2001) 177–223.

[64] A.Wiles, Modular elliptic curves and Fermat's last Theorem, Ann. Math. 141 (1995) 443–551.
[65] J.-P.Wintenberger, On p-adic geometric representations of $G_{\mathbf{Q}}$, Doc. Math., Extra Volume: John H. Coates' Sixtieth Birthday (2006) 819–827.
[66] L.Yang, Multiplicity of Galois representations in the higher weight sheaf cohomology associated to Shimura curves, CUNY thesis, 1996.

Department of Mathematics, Imperial College, 180 Queen's Gate, London SW7 2AZ, UK
E-mail address: k.buzzard@imperial.ac.uk

Department of Mathematics, King's College London, Strand, London WC2R 2LS, UK
E-mail address: fred.diamond@kcl.ac.uk

Department of Pure Mathematics, University of Sheffield, Sheffield S3 7RH, UK
E-mail address: A.F.Jarvis@sheffield.ac.uk