THE BREUIL–MÉZARD CONJECTURE FOR QUATERNION ALGEBRAS

TOBY GEE AND DAVID GERAGHTY

ABSTRACT. We formulate a version of the Breuil–Mézard conjecture for quaternion algebras, and show that it follows from the Breuil–Mézard conjecture for GL_2. In the course of the proof we establish a mod p analogue of the Jacquet–Langlands correspondence for representations of $GL_2(k)$, k a finite field of characteristic p.

Contents

1. INTRODUCTION.

The Breuil–Mézard conjecture ([BM02]) has proved to be one of the most important conjectures linking Galois representations and automorphic forms; indeed, Kisin's proof of (most cases of) the original formulation of the conjecture ([Kis09]) simultaneously established (most cases of) the Fontaine–Mazur conjecture for $GL_2 / \mathbb{Q}$. The original conjecture predicted the Hilbert–Samuel multiplicities of the special fibres of potentially semistable deformation rings for two-dimensional mod p representations of $G_{\mathbb{Q}_p}$, the absolute Galois group of $\mathbb{Q}_p$, in terms of the representation theory of $GL_2(\mathbb{Z}_p)$. The statement of the conjecture was generalised in [Kis10] to the case of representations of G_K, for K an arbitrary finite extension of $\mathbb{Q}_p$. This conjecture is largely open, although it has been proved for potentially Barsotti–Tate representations ([GK12]).

The connection between potentially semistable deformation rings and the representation theory of GL_2 is via the local Langlands correspondence. Given the Jacquet–Langlands correspondence, it is natural to wonder whether for potentially semistable deformation rings of discrete series type, the Hilbert–Samuel multiplicities could also be described in terms of the representation theory of the units in a non-split quaternion algebra. One advantage of such a description is that the

representation theory is much simpler in this case; all the irreducible admissible representations are finite-dimensional, and the irreducible mod p representations of the maximal compact subgroup are all one-dimensional.

In this paper we formulate such a conjecture, and show that it is a consequence of the conjecture for GL_2. In particular, we prove the conjecture (in most cases) over $\mathbb{Q}_p$, as a consequence of Kisin's proof of the conjecture for GL_2 in this case ([Kis09]). In order to do this, we have found it helpful to reformulate the conjecture slightly more abstractly in terms of linear functionals on Grothendieck groups of representations, and also to prove a general result on the reduction modulo p of the Jacquet–Langlands correspondence, or rather a version of this correspondence for types. In the remainder of the introduction, we will explain this in some detail.

Let K be a finite extension of $\mathbb{Q}_p$ with absolute Galois group G_K, and let τ be an inertial type for K (i.e. a two-dimensional representation of the inertia group I_K with open kernel, which can be extended to G_K). Let λ be a highest weight for an irreducible algebraic representation of $\mathrm{Res}_{K/\mathbb{Q}_p} \mathrm{GL}_{2/K}$. Then a recipe using Henniart's inertial local Langlands correspondence (see the appendix to [BM02]) associates to the pair (τ, λ) a finite-dimensional irreducible representation $\sigma(\tau, \lambda)$ of $\mathrm{GL}_2(\mathcal{O}_K)$ over $\overline{\mathbb{Q}}_p$. Let k denote the residue field of $\mathcal{O}_K$. Choosing a stable lattice, reducing modulo p and semisimplifying, we can write

$$\overline{\sigma}(\tau, \lambda) \cong \oplus_{\overline{\sigma}} \overline{\sigma}^{n_{\tau,\lambda}(\overline{\sigma})},$$

where $\overline{\sigma}$ runs over the equivalence classes of irreducible mod p representations of $\mathrm{GL}_2(k)$, and $n_{\tau,\lambda}(\overline{\sigma})$ is a nonnegative integer.

Let $\overline{\rho} : G_K \to \mathrm{GL}_2(\overline{\mathbb{F}}_p)$ be a continuous representation. Then there is (after fixing a sufficiently large coefficient field) a universal lifting ring $R^{\tau,\lambda}$ for lifts of $\overline{\rho}$ which are potentially semistable of inertial type τ and Hodge type λ. Let $e(R^{\tau,\lambda}/\varpi)$ denote the Hilbert–Samuel multiplicity of the special fibre of $R^{\tau,\lambda}$. Then the Breuil–Mézard conjecture asserts that there are uniquely determined nonnegative integers $\mu_{\overline{\sigma}}(\overline{\rho})$, depending only on $\overline{\rho}$ and $\overline{\sigma}$ (and not on τ or λ) such that for all τ, λ, we have

$$(1) \qquad\qquad e(R^{\tau,\lambda}/\varpi) = \sum_{\overline{\sigma}} n_{\tau,\lambda}(\overline{\sigma})\mu_{\overline{\sigma}}(\overline{\rho}).$$

Now, the right hand side of (1) depends only on $\overline{\sigma}(\tau, \lambda)$, the semisimplification of the reduction modulo p of $\sigma(\tau, \lambda)$. Let $R_{\overline{\mathbb{F}}_p}(\mathrm{GL}_2(k))$ denote the Grothendieck group of finite-dimensional $\overline{\mathbb{F}}_p$-representations of $\mathrm{GL}_2(k)$; then we may define a linear functional $\iota : R_{\overline{\mathbb{F}}_p}(\mathrm{GL}_2(k)) \to \mathbb{Z}$ by sending $\overline{\sigma}$ to $\mu_{\overline{\sigma}}(\overline{\rho})$. Then the right hand side of (1) is just $\iota(\overline{\sigma}(\tau, \lambda))$ by definition.

With this perspective in mind, let D be the non-split quaternion algebra with centre K, and let $\mathcal{O}_D$ be the maximal order in D. Suppose that τ is a discrete series type (that is, it is scalar or it can be extended to an irreducible representation of G_K). As explained in Section 3, a natural analogue of the procedure above associates a finite-dimensional representation $\sigma_D(\tau, \lambda)$ of $\mathcal{O}_D^\times$ to the pair (τ, λ). If l is the quadratic extension of k, then irreducible mod p representations of $\mathcal{O}_D^\times$ factor through $l^\times$, so we see that the natural analogue of the Breuil–Mézard conjecture for $D^\times$ is to ask for a linear functional $\iota_D : R_{\overline{\mathbb{F}}_p}(l^\times) \to \mathbb{Z}$ with the property that for all pairs (τ, λ) where τ is discrete series, we have

$$e(R^{\tau,\lambda,\mathrm{ds}}/\varpi) = \iota_D(\overline{\sigma}_D(\tau, \lambda)),$$

where $R^{\tau,\lambda,\mathrm{ds}}$ denotes the maximal quotient of $R^{\tau,\lambda}$ corresponding to discrete series lifts (see Section 5 for more details).

Our approach in this paper is to deduce the existence of ι_D from the existence of ι. The existence of such functionals for all representations $\overline{\rho}$ strongly suggests the possibility of there being a homomorphism $\mathrm{JL} : R_{\overline{\mathbb{F}}_p}(l^\times) \to R_{\overline{\mathbb{F}}_p}(\mathrm{GL}_2(k))$ such that $\iota_D = \iota \circ \mathrm{JL}$, and the construction of such a map is the main objective of this paper. Since elements of the Grothendieck group are determined by their Brauer characters, this determines a map between the class functions on the semisimple conjugacy classes of $\mathrm{GL}_2(k)$ and $l^\times$. The usual Jacquet–Langlands correspondence involves a sign-reversing relation between the characters evaluated at regular elliptic elements; our correspondence satisfies a close analogue of this relation.

Having written down this map, in order to check that $\iota \circ \mathrm{JL}$ satisfies the properties required of ι_D, the main fact we need to check is that JL takes $\overline{\sigma}_D(\tau, \lambda)$ to $\overline{\sigma}(\tau, \lambda)$ when τ is of supercuspidal type. In other words, we need to check that JL is compatible with the usual Jacquet–Langlands correspondence (or rather the induced correspondence for types) and reduction modulo p. In order to do this, we use results of Carayol [Car84] on the construction of supercuspidal representations as well as results of Kutzko [Kut87] on the characters of supercuspidal representations and the characters of the types they contain. Fred Diamond has pointed out to us that it is presumably also possible to verify this directly using the explicit formulas in the appendix to [BD12]. We suspect that the approach taken here will extend to give similar results for GL_n (and that the extension should be relatively straightforward when n is prime); we intend to return to this question in future work. Florian Herzig pointed out to us that our correspondence JL is given (up to a sign) by the reduction modulo p of Deligne–Lusztig induction from a non-split torus in GL_2 to GL_2. This immediately suggests natural analogues of JL in the case of GL_n.

We would like to thank Kevin Buzzard for asking whether there was a Breuil–Mézard conjecture for quaternion algebras. We would also like to thank Matthew Emerton, Florian Herzig, Guy Henniart, Mark Kisin, Vytautas Paškūnas, and Shaun Stevens for helpful conversations.

1.1. **Notation.** Fix a prime number p and an algebraic closure $\overline{\mathbb{Q}}_p$ of $\mathbb{Q}_p$. This determines an algebraic closure $\overline{\mathbb{F}}_p$ of $\mathbb{F}_p$.

Fix K a finite extension of $\mathbb{Q}_p$ with ring of integers $\mathcal{O}_K$, uniformiser ϖ, and residue field k of cardinality q. Write G_K for (a choice of) the absolute Galois group of K, and I_K for its inertia subgroup.

Let D be the (unique up to isomorphism) non-split quaternion algebra with centre K, and let $\mathcal{O}_D$ be the maximal order in D. Fix a uniformiser ϖ_D of D. Let L be the quadratic unramified extension of K, so that D splits over L. If l is the residue field of L, then $\mathcal{O}_D/\varpi_D \cong l$. We let ν_D denote the valuation on D defined by $\nu_D(x) = \nu_K(\mathrm{Nm}(x))$ where Nm is the reduced norm on D and ν_K the valuation on K normalised by $\nu_K(\varpi) = 1$. We define $U_D^0 = \mathcal{O}_D^\times$ and if $a \geq 1$ is an integer, we let $U_D^a = 1 + \varpi_D^a \mathcal{O}_D \subset U_D^0$.

We let rec_p be the local Langlands correspondence of [HT01], so that if π is an irreducible $\overline{\mathbb{Q}}_p$-representation of $\mathrm{GL}_n(K)$, then $\mathrm{rec}_p(\pi)$ is a Weil–Deligne representation of the Weil group W_K defined over $\overline{\mathbb{Q}}_p$. If $R = (r, N)$ is a Weil–Deligne

representation of W_K (so in particular, r is a representation of W_K with open kernel and N is a nilpotent endomorphism), then by $R|_{I_K}$ we mean $r|_{I_K}$.

If $E/\mathbb{Q}_p$ is an algebraic extension and V is a continuous representation of a compact group G on a finite-dimensional E-vector space V, then we define a semisimple representation $\overline{V}$ of G over the residue field of E as follows: since G is compact, it stabilizes an $\mathcal{O}_E$-lattice in V. Reducing such a lattice modulo the maximal ideal of $\mathcal{O}_E$ and semisimplifying gives the required representation. This representation is independent of the choice of lattice by the Brauer–Nesbitt theorem.

If K is a p-adic field, W is a de Rham representation of G_K over E, and $\kappa : K \hookrightarrow E$, then we will write $\mathrm{HT}_\kappa(W)$ for the multiset of Hodge–Tate weights of W with respect to κ. By definition, the multiset $\mathrm{HT}_\kappa(W)$ contains i with multiplicity $\dim_E(W \otimes_{\kappa,F} \widehat{\overline{F}}(i))^{G_F}$. Let $\mathbb{Z}_+^2 = \{(a_1, a_2) \in \mathbb{Z}^2 : a_1 \geq a_2\}$, and fix $\lambda \in (\mathbb{Z}_+^2)^{\mathrm{Hom}_{\mathbb{Q}_p}(K,E)}$. If W is two-dimensional, then we say that W has *Hodge type* λ if for each $\kappa : K \hookrightarrow E$, we have $\mathrm{HT}_\kappa(W) = \{\lambda_{\kappa,1} + 1, \lambda_{\kappa,2}\}$.

If G is a finite group, we let $R_{\overline{\mathbb{F}}_p}(G)$ denote the Grothendieck group of the category of finitely generated $\overline{\mathbb{F}}_p[G]$-modules.

If R is a commutative ring, we let R^{red} denote the maximal reduced quotient of R.

2. A MOD p JACQUET–LANGLANDS CORRESPONDENCE FOR FINITE GROUPS

We begin by defining an analogue of the Jacquet–Langlands correspondence for mod p representations of $\mathrm{GL}_2(\mathcal{O}_K)$ and $\mathcal{O}_D^\times$. The irreducible mod p representations of these two groups are obtained via inflation from the irreducible mod p representations of $\mathrm{GL}_2(k)$ and $l^\times$, and our correspondence is actually between the Grothendieck groups $R_{\overline{\mathbb{F}}_p}(\mathrm{GL}_2(k))$ and $R_{\overline{\mathbb{F}}_p}(l^\times)$. An element of either Grothendieck group is determined by its Brauer character so we may equivalently describe our map on the level of Brauer characters. Both descriptions are given below. Given an element $\overline{\sigma}$ of either $R_{\overline{\mathbb{F}}_p}(\mathrm{GL}_2(k))$ and $R_{\overline{\mathbb{F}}_p}(l^\times)$, we write $\chi_{\overline{\sigma}}$ for its Brauer character, which we view as being valued in our fixed $\overline{\mathbb{Q}}_p$. Recall that if G is a finite group, the Brauer character of a finite $\overline{\mathbb{F}}_p[G]$ module is a function on the p-regular conjugacy classes in G. For $G = \mathrm{GL}_2(k)$, the p-regular conjugacy classes coincide with the semisimple conjugacy classes; representative elements for these conjugacy classes are given by the diagonal matrices, and the matrices $i(z)$, where $z \in l^\times \setminus k^\times$ and $i : l \hookrightarrow M_2(k)$ denotes a choice of embedding of k-algebras. For $G = l^\times$, the p-regular conjugacy classes are just the elements of $l^\times$.

Definition 2.1. We define an additive map $\mathrm{JL} : R_{\overline{\mathbb{F}}_p}(l^\times) \to R_{\overline{\mathbb{F}}_p}(\mathrm{GL}_2(k))$ as follows:

- if $\psi : k^\times \to \overline{\mathbb{F}}_p^\times$ is a character, then
$$\mathrm{JL}([\psi \circ \mathbf{N}_{l/k}]) = [\overline{\mathrm{sp}_\psi}] - [\overline{\psi} \circ \det]$$

- if $\psi : l^\times \to \overline{\mathbb{F}}_p^\times$ is a character which does not factor through the norm $\mathbf{N}_{l/k}$, then
$$\mathrm{JL}([\psi]) = [\overline{\Theta(\psi)}].$$

Here the representations sp_ψ and $\Theta(\psi)$ are as defined in [Dia07, §1].

We let $\mathcal{C}(\mathrm{GL}_2(k))$ (resp. $\mathcal{C}(l^\times)$) be the space of $\overline{\mathbb{Q}}_p$-valued class functions on the semisimple conjugacy classes in $\mathrm{GL}_2(k)$ (resp. $l^\times$). Then we have $\mathcal{C}(\mathrm{GL}_2(k)) = R_{\overline{\mathbb{F}}_p}(\mathrm{GL}_2(k)) \otimes_{\mathbb{Z}} \overline{\mathbb{Q}}_p$ (resp. $\mathcal{C}(l^\times) = R_{\overline{\mathbb{F}}_p}(l^\times) \otimes_{\mathbb{Z}} \overline{\mathbb{Q}}_p$). Note that $\mathcal{C}(\mathrm{GL}_2(k))$ and $\mathcal{C}(l^\times)$ have natural ring structures, where multiplication corresponds to the tensor product on $R_{\overline{\mathbb{F}}_p}(\mathrm{GL}_2(k))$ and $R_{\overline{\mathbb{F}}_p}(l^\times)$. We may also describe $\mathrm{JL} : \mathcal{C}(l^\times) \to \mathcal{C}(\mathrm{GL}_2(k))$ as follows: If $\chi \in \mathcal{C}(l^\times)$ then $\mathrm{JL}(\chi)$ is defined by the following rule:

$$i(z) \longmapsto -\chi(z) - \chi(z^q) \qquad \text{if } z \in l^\times \setminus k^\times$$

$$\begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix} \longmapsto (q-1)\chi(x) \qquad \text{if } x \in k^\times$$

$$\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} \longmapsto 0 \qquad \text{if } x, y \in k^\times, x \neq y.$$

That this definition agrees with the previous one follows immediately from the table of Brauer characters in [Dia07, §1].

3. Types and supercuspidal representations

In this section we will discuss types for $\mathrm{GL}_2(\mathcal{O}_K)$ and $\mathcal{O}_D^\times$, and the inertial local Langlands and Jacquet–Langlands correspondences. All representations in this section will be over $\overline{\mathbb{Q}}_p$, unless otherwise stated. Recall that an irreducible admissible smooth representation of $\mathrm{GL}_2(K)$ is either one-dimensional, a principal series representation, a twist of the Steinberg representation, or is a supercuspidal representation. If it is either supercuspidal or a twist of the Steinberg representation, we say that it is a *discrete series* representation. There is a bijection JL (the Jacquet–Langlands correspondence) from the irreducible smooth admissible representations of $D^\times$ (which are necessarily finite-dimensional) to the discrete series representations of $\mathrm{GL}_2(K)$ (which are necessarily infinite-dimensional). Under this correspondence, the 1-dimensional representations of $D^\times$ correspond to the twists of the Steinberg representation. More precisely, for each character $\psi : K^\times \to \overline{\mathbb{Q}}_p$, we have $\mathrm{JL}(\psi \circ \mathrm{Nm}) = \mathrm{Sp}_2(\psi | \ |^{-1/2})$. (See [HT01, p.32] for this formula and the definition of $\mathrm{Sp}_2(*)$.)

3.1. **Types.** In this paper we will be particularly concerned with *types*, which are finite-dimensional representations of $\mathrm{GL}_2(\mathcal{O}_K)$ and $\mathcal{O}_D^\times$, and with their relationship to inertial types. An *inertial type* is a two-dimensional representation τ of I_K with open kernel which may be extended to a representation of G_K. We say that τ is a *discrete series* type if it is either scalar, or can be extended to an irreducible representation of G_K. In the latter case, we say that τ is *supercuspidal*.

In the GL_2 case, the theory of types is worked out explicitly in Henniart's appendix to [BM02]. We recall his main result. (We follow [Kis09] in introducing the notation $\sigma^{\mathrm{cr}}(\tau)$.)

Theorem 3.2. *For any inertial type τ, there are unique finite dimensional irreducible representations $\sigma(\tau)$ and $\sigma^{\mathrm{cr}}(\tau)$ of $\mathrm{GL}_2(\mathcal{O}_K)$, with the following properties:*

(1) *if π is an infinite dimensional smooth irreducible representation of $\mathrm{GL}_2(K)$, then $\mathrm{Hom}_{\mathrm{GL}_2(\mathcal{O}_K)}(\sigma(\tau), \pi) \neq 0$ if and only if $\mathrm{rec}_p(\pi)|_{I_K} \cong \tau$, in which case $\mathrm{Hom}_{\mathrm{GL}_2(\mathcal{O}_K)}(\sigma(\tau), \pi)$ is one-dimensional.*

(2) *if π is any smooth irreducible representation of $\mathrm{GL}_2(K)$, then we have*
$\mathrm{Hom}_{\mathrm{GL}_2(\mathcal{O}_K)}(\sigma^{\mathrm{cr}}(\tau), \pi) \neq 0$ *if and only if* $\mathrm{rec}_p(\pi)|_{I_K} \cong \tau$ *and the monodromy operator N on* $\mathrm{rec}_p(\pi)$ *is 0. In this case,* $\mathrm{Hom}_{\mathrm{GL}_2(\mathcal{O}_K)}(\sigma^{\mathrm{cr}}(\tau), \pi)$ *is one-dimensional.*

There is an analogous (but much simpler) theory for $D^\times$, which we now recall, following Section 5.2 of [GK12]. Note that $K^\times \mathcal{O}_D^\times$ has index two in $D^\times$. Thus if π_D is an admissible smooth representation of $D^\times$, then $\pi_D|_{\mathcal{O}_D^\times}$ is either irreducible or a sum of two irreducible representations which are conjugate under a uniformiser ϖ_D in $D^\times$. Moreover, we easily see that if π'_D is another smooth irreducible representation of $D^\times$, then π_D and π'_D differ by an unramified twist if and only if $\pi_D|_{\mathcal{O}_D^\times} \cong \pi'_D|_{\mathcal{O}_D^\times}$.

Let τ be a discrete series inertial type. Then by the Jacquet–Langlands correspondence, there is an irreducible smooth representation $\pi_{D,\tau}$ of $D^\times$ such that $\mathrm{rec}_p(\mathrm{JL}(\pi_{D,\tau}))|_{I_K} \cong \tau$. Define $\sigma_D(\tau)$ to be *one* of the irreducible components of $\pi_{D,\tau}|_{\mathcal{O}_D^\times}$; then by the above discussion, we have the following property.

Theorem 3.3. *Let τ be a discrete series inertial type. If π_D is a smooth irreducible $\overline{\mathbb{Q}}_p$-representation of $D^\times$ then $\mathrm{Hom}_{\mathcal{O}_D^\times}(\sigma_D(\tau), \pi_D)$ is non-zero if and only if $\mathrm{rec}_p(\mathrm{JL}(\pi))|_{I_K} \cong \tau$, in which case $\mathrm{Hom}_{\mathcal{O}_D^\times}(\sigma_D(\tau), \pi)$ is one-dimensional.*

Remark 3.4. By the above discussion, any representation satisfying the property of $\sigma_D(\tau)$ given in Theorem 3.3 is necessarily isomorphic to $\sigma_D(\tau)$ or to $\sigma_D(\tau)^{\varpi_D}$.

For our purposes, we will however require some more precise results: we will need to know exactly when $\pi_{D,\tau}|_{\mathcal{O}_D^\times}$ is irreducible and we will need to relate the characters of $\sigma(\tau)$ and $\sigma_D(\tau)$ in a sense we will make precise below.

3.5. **Supercuspidal representations.** Let π be a smooth irreducible representation of $\mathrm{GL}_2(K)$ or $D^\times$. Then π is said to be *minimal* if π has minimal conductor amongst all its twists by characters. Following [Car84], we define subgroups Z_s, K_s of $\mathrm{GL}_2(K)$ for $s = 1, 2$ as follows:

- $Z_1 = \langle \varpi \rangle$, $K_1 = \mathrm{GL}_2(\mathcal{O}_K)$,
- $Z_2 = \langle \begin{pmatrix} 0 & 1 \\ \varpi & 0 \end{pmatrix} \rangle$, $K_2 = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathcal{O}_K) : a, d \in \mathcal{O}^\times, c \in (\varpi), b \in \mathcal{O} \right\}$.

We refer to [Car84, §4] for the definition of a *very cuspidal* representation of $Z_s K_s$ $Z_2 K_2$ of type $m \geq 1$. When $s = 2$, such representations exist only when m is even.

Theorem 3.6 ([Car84] Théorèmes 4.2 & 8.1)**.**

(1) *Let $s = 1$ or 2 and set $r = 2/s$. Let ρ be a very cuspidal representation of $Z_s K_s$ of type m. Then* $\mathrm{c\text{-}Ind}_{Z_s K_s}^{\mathrm{GL}_2(K)} \rho$ *is an irreducible minimal supercuspidal representation of $\mathrm{GL}_2(K)$ of conductor $r(m-1) + 2$.*

(2) *The representations obtained in (1) are all inequivalent.*

(3) *Every irreducible minimal supercuspidal representation of $\mathrm{GL}_2(K)$ is isomorphic to* $\mathrm{c\text{-}Ind}_{Z_s K_s}^{\mathrm{GL}_2(K)} \rho$ *for a uniquely determined pair s, ρ as in (1).*

We note that the representations given by (1) have even conductor when $s = 1$ and odd conductor when $s = 2$. This result in fact allows us to give an explicit description of the types corresponding to supercuspidal inertial types.

Proposition 3.7. *Let $\tau : I_K \to \mathrm{GL}_2(\overline{\mathbb{Q}}_p)$ be a supercuspidal inertial type. More-over, assume that τ has minimal conductor amongst its twists by smooth characters that extend to G_K. Choose a (necessarily minimal and supercuspidal) representa-tion π of G with $\mathrm{rec}(\pi)$ equal to an extension of τ to W_K. Write $\pi = \mathrm{c\text{-}Ind}_{Z_s K_s}^{\mathrm{GL}_2(K)} \rho$ as in Theorem 3.6. Then*

$$\sigma(\tau) \cong \mathrm{Ind}_{K_s}^{\mathrm{GL}_2(\mathcal{O}_K)}(\rho|_{K_s}).$$

Proof. The construction of $\sigma(\tau)$ described by Henniart in [BM02, §A.3.1] is exactly the description given in the statement of the proposition. $\qquad\square$

When we pass to representations of $D^\times$ via the Jacquet–Langlands correspon-dence, there is a similar dichotomy which tells us precisely when the restriction to $\mathcal{O}_D^\times$ of a smooth irreducible representation of $D^\times$ is reducible.

Proposition 3.8. *Let π_D be a smooth irreducible minimal representation of $D^\times$ of dimension greater than 1. Write $c(\pi_D)$ for the conductor of π_D.*

(1) *If $c(\pi_D)$ is odd, then $\pi_D|_{\mathcal{O}_D^\times}$ is irreducible.*
(2) *If $c(\pi_D)$ is even, then $\pi_D|_{\mathcal{O}_D^\times} \cong \sigma_D \oplus \sigma_D^{\varpi_D}$ for some irreducible representa-tion σ_D of $\mathcal{O}_D^\times$ with $\sigma_D \not\cong \sigma_D^{\varpi_D}$.*

Proof. Suppose first of all that $c = c(\pi_D)$ is odd. Let $a = (c-1)/2$ and let χ denote a character of the abelian group U_D^a/U_D^{c-1} appearing in $\pi_D|_{U_D^a}$. In [Car84, §6.7] (where the integer a is denoted p), it is shown that the stabilizer Z_χ of χ in $D^\times$ is equal to $K(u)^\times U_D^a$ where $u \in D$ is an element generating a ramified quadratic extension of K. In [Car84, §6.8], it is shown that

$$\pi_D \cong \mathrm{Ind}_{K(u)^\times U_D^a}^{D^\times} \psi$$

for some character ψ extending χ. Now, since $\mathcal{O}_D^\times \backslash D^\times / K(u)^\times U_D^a \cong \mathbb{Z}/\nu_D(K(u)^\times) = \mathbb{Z}/\mathbb{Z} = 0$, it follows that

$$\pi_D|_{\mathcal{O}_D^\times} \cong \mathrm{Ind}_{\mathcal{O}_{K(u)}^\times U_D^a}^{\mathcal{O}_D^\times} \psi.$$

Thus, $\pi_D|_{\mathcal{O}_D^\times}$ is irreducible if and only if for each $t \in \mathcal{O}_D^\times - \mathcal{O}_{K(u)}^\times U_D^a$, the characters ψ and ψ^t of $H_t := t\mathcal{O}_{K(u)}^\times U_D^a t^{-1} \cap \mathcal{O}_{K(u)}^\times U_D^a$ are distinct. However, if $\psi = \psi^t$ on H_t for some $t \in \mathcal{O}_D^\times - \mathcal{O}_{K(u)}^\times U_D^a$, then since $U_D^a \subset H_t$, we certainly have $\psi^t|_{U_D^a} = \chi$. Thus, by definition, $t \in Z_\chi = K(u)^\times U_D^a$, a contradiction.

Suppose now that $c = c(\pi_D)$ is even and set $a = (c-2)/2$. Then in [Car84, §6.9] it is shown that

$$\pi_D \cong \mathrm{Ind}_{K(u)^\times U_D^a}^{D^\times} \rho$$

where now $u \in D$ generates the quadratic *unramified* extension of K and ρ is a representation of dimension 1 or q^2. Since

$$\mathcal{O}_D^\times \backslash D^\times / K(u)^\times U_D^a \cong \mathbb{Z}/\nu_D(K(u)^\times) = \mathbb{Z}/2,$$

we deduce immediately that $\pi_D|_{\mathcal{O}_D^\times}$ has at least 2 irreducible components. The stated result now follows easily from the fact that $K^\times \mathcal{O}_D^\times$ has index 2 in $D^\times$. $\quad\square$

We now recall some further results of Carayol.

Proposition 3.9. *Let π_D be a smooth irreducible minimal representation of $D^\times$ of dimension greater than 1. Let $\pi = \mathrm{JL}(\pi_D)$ and write*

$$\pi = \text{c-Ind}_{Z_s K_s}^{\mathrm{GL}_2(K)} \rho$$

for some uniquely determined pair s, ρ as in Theorem 3.6 (1). Then

- *If $s = 2$, then $(q-1) \dim \pi_D = (q+1) \dim \rho$.*
- *If $s = 1$, then $(q-1) \dim \pi_D = 2 \dim \rho$.*

Proof. By [Car84, Proposition 7.4], the dimension of π_D coincides with the formal degree of π (when Haar measure on $\mathrm{GL}_2(K)/K^\times$ is normalized as in [Car84, §5.10]). The stated result now follows from the formulas obtained in [Car84, §5.9 – 5.11].   □

We deduce the following formula relating the dimension of types for $\mathrm{GL}_2(K)$ and $D^\times$.

Corollary 3.10. *Let π_D be a smooth irreducible minimal representation of $D^\times$ of dimension greater than 1. Let $\pi = \mathrm{JL}(\pi_D)$ and write*

$$\pi = \text{c-Ind}_{Z_s K_s}^{\mathrm{GL}_2(K)} \rho$$

for some uniquely determined pair s, ρ as in Theorem 3.6 (1). Define

$$\sigma := \mathrm{Ind}_{K_s}^{\mathrm{GL}_2(\mathcal{O}_K)}(\rho|_{K_s})$$

and let σ_D denote an irreducible constituent of $\pi_D|_{\mathcal{O}_D^\times}$. Then

$$(q-1) \dim \sigma_D = \dim \sigma.$$

Remark 3.11. In our definition of the type $\sigma_D(\tau)$ for $\mathcal{O}_D^\times$, we arbitrarily choose one of the irreducible constituents of $\pi_{D,\tau}|_{\mathcal{O}_D^\times}$. This has the apparent disadvantage of breaking the symmetry of the situation but has the advantage that the dimension formula above holds independently of the parity of the conductor. Ultimately in our statement of the Breuil–Mézard conjecture for $D^\times$ we will consider both choices; see Conjecture 5.3 and Remark 5.4.

Keep the notation of the preceding corollary. We now proceed to show that the characteristic p reductions $\overline{\sigma}$ and $\overline{\sigma}_D$ of σ and σ_D are related by the mod p Jacquet–Langlands map defined in Definition 2.1. For this we will make use of results of Kutzko [Kut87].

We will denote the characters of σ and σ_D by χ_σ and χ_{σ_D} respectively. Note that the representation $\overline{\sigma}$ (resp. $\overline{\sigma}_D$) factors through the quotient $\mathrm{GL}_2(\mathcal{O}_K) \twoheadrightarrow \mathrm{GL}_2(k)$ (resp. $\mathcal{O}_D^\times \twoheadrightarrow l^\times$). We denote the Brauer character of $\overline{\sigma}$ (resp. $\overline{\sigma}_D$) by $\chi_{\overline{\sigma}} : (\mathrm{GL}_2(k)/\sim)^{\mathrm{ss}} \to \overline{\mathbb{Z}}_p$ (resp. $\chi_{\overline{\sigma}_D} : l^\times \to \overline{\mathbb{Z}}_p$). Here $(\mathrm{GL}_2(k)/\sim)^{\mathrm{ss}}$ is the set of semisimple (or equivalently, p-regular) conjugacy classes in $\mathrm{GL}_2(k)$.

If $x \in l^\times$, we let $\widetilde{x} \in \mathcal{O}_L^\times$ denote its Teichmüller lift. Choose an isomorphism of $\mathcal{O}_K$-modules $i : \mathcal{O}_L \xrightarrow{\sim} \mathcal{O}_K \oplus \mathcal{O}_K$. This gives rise to injections $i : \mathcal{O}_L^\times \hookrightarrow \mathrm{GL}_2(\mathcal{O}_K)$ and $i : l^\times \hookrightarrow \mathrm{GL}_2(k)$. We also fix an embedding $j : L \hookrightarrow D$ giving rise to an injection $j : \mathcal{O}_L^\times \hookrightarrow \mathcal{O}_D^\times$.

Proposition 3.12. *Let π_D be a smooth irreducible minimal representation of $D^\times$ of dimension greater than 1. Let $\pi = \mathrm{JL}(\pi_D)$ and write*

$$\pi = \text{c-Ind}_{Z_s K_s}^{\mathrm{GL}_2(K)} \rho$$

for some uniquely determined pair s, ρ as in Theorem 3.6 (1). Define

$$\sigma := \mathrm{Ind}_{K_s}^{\mathrm{GL}_2(\mathcal{O}_K)} \rho|_{K_s}$$

and let σ_D denote an irreducible constituent of $\pi_D|_{\mathcal{O}_D^\times}$. Then

$$\mathrm{JL}(\overline{\sigma}_D) = \overline{\sigma}.$$

Proof. Since both $\mathrm{JL}(\overline{\sigma}_D)$ and $\overline{\sigma}$ are semisimple, it suffices to show that we have an equality of Brauer characters $\mathrm{JL}(\chi_{\overline{\sigma}_D}) = \chi_{\overline{\sigma}}$. Let $g \in \mathrm{GL}_2(k)$ be a p-regular element. We need to check that $\mathrm{JL}(\chi_{\overline{\sigma}_D})(g) = \chi_{\overline{\sigma}}(g)$. There are three cases:

(1) *We have $g = x$ with $x \in k^\times$.* In this case, we need to show that

$$(q-1)\chi_{\overline{\sigma}_D}(x) = \chi_{\overline{\sigma}}(x)$$

or equivalently, that $(q-1)\chi_{\sigma_D}(\widetilde{x}) = \chi_\sigma(\widetilde{x})$. This follows from Corollary 3.10 and the fact that π and π_D have the same central character.

(2) *We have $g \sim \mathrm{diag}(x, y)$ with $x, y \in k^\times$ distinct.* In this case we are required to show that $\chi_{\overline{\sigma}}(g) = 0$, or equivalently, that $\chi_\sigma(\widetilde{g}) = 0$ where $\widetilde{g} = \mathrm{diag}(\widetilde{x}, \widetilde{y})$. If $s = 2$, this follows from [Kut87, Prop. 3.4] and the Frobenius formula for the trace of an induced representation. If $s = 1$, it follows from [Kut87, Lemmas 6.3 & 6.4].

(3) *We have $g \sim i(z)$ for some $z \in l^\times \setminus k^\times$.* In this case, we need to show that

$$-\chi_{\overline{\sigma}_D}(z) - \chi_{\overline{\sigma}_D}(z^q) = \chi_{\overline{\sigma}}(i(z)).$$

Let us first consider the sub-case where $s = 2$. Then $\pi_D|_{\mathcal{O}_D^\times}$ is irreducible and it suffices for us to show that

$$-2\chi_{\sigma_D}(j(\widetilde{z})) = \chi_\sigma(i(\widetilde{z})).$$

We will in fact show that both sides vanish. To see that the right hand side vanishes, recall that $\sigma = \mathrm{Ind}_{K_2}^{\mathrm{GL}_2(\mathcal{O}_K)} \rho$ and note that for all $t \in \mathrm{GL}_2(\mathcal{O}_K)$, we have $t^{-1} i(\widetilde{z}) t \notin K_2$. For the left hand side, we have

$$-\chi_{\sigma_D}(j(\widetilde{z})) = -\chi_{\pi_D}(j(\widetilde{z})) = \chi_\pi(i(\widetilde{z})),$$

where the second equality is a property of the Jacquet-Langlands correspondence. The vanishing then follows from [Kut87, Prop. 5.5(2)].

Finally, we treat the sub-case where $s = 1$. Then $\pi_D|_{\mathcal{O}_D^\times}$ is reducible and it suffices to show that

$$-\chi_{\sigma_D}(j(\widetilde{z})) - \chi_{\sigma_D}(j(\widetilde{z}^q)) = \chi_\sigma(i(\widetilde{z})).$$

For this, note that the left hand side is just $-\chi_{\pi_D}(j(\widetilde{z}))$, which in turn equals $\chi_\pi(i(\widetilde{z}))$. Thus we are required to show that $\chi_\pi(i(\widetilde{z})) = \chi_\sigma(i(\widetilde{z}))$. Yet this follows from [Kut87, Prop 6.11(1)] and the proof is complete. $\square$

4. Compatibility of Jacquet–Langlands correspondences

In this section we prove our main technical result, a generalization of Proposition 3.12 which includes the case of twists of the Steinberg representation (that is, the case where π_D as in Proposition 3.12 is one-dimensional) and incorporates algebraic representations. Again, all representations will be over $\overline{\mathbb{Q}}_p$ unless otherwise stated.

Let W_λ be an irreducible algebraic representation of $\operatorname{Res}_{K/\mathbb{Q}_p} \operatorname{GL}_{2/K}$ of highest weight λ. More precisely, we let $\lambda \in (\mathbb{Z}_+^2)^{\operatorname{Hom}_{\mathbb{Q}_p}(K,\overline{\mathbb{Q}}_p)}$, and take W_λ to be the representation

$$W_\lambda = \otimes_{\tau \in \operatorname{Hom}_{\mathbb{Q}_p}(K,\overline{\mathbb{Q}}_p)}(\operatorname{Sym}^{a_{\tau,1}-a_{\tau,2}} \otimes \det{}^{a_{\tau,2}})(\overline{\mathbb{Q}}_p^2)$$

of $\operatorname{Res}_{K/\mathbb{Q}_p} \operatorname{GL}_{2/K} \times_{\mathbb{Q}_p} \overline{\mathbb{Q}}_p = \prod_\tau \operatorname{GL}_{2/\overline{\mathbb{Q}}_p}$. We regard W_λ as a representation of $\operatorname{GL}_2(K)$ via the map $\prod_\tau \tau : \operatorname{GL}_2(K) \to \prod_\tau \operatorname{GL}_2(\overline{\mathbb{Q}}_p)$. We can also regard it as a representation of $D^\times$ as follows: choose an isomorphism $D \otimes_K L \cong M_{2\times 2}(L)$ and for each $\mathbb{Q}_p$-embedding $\tau : K \hookrightarrow \overline{\mathbb{Q}}_p$ choose an embedding $\widetilde{\tau} : L \hookrightarrow \overline{\mathbb{Q}}_p$ extending τ. Then we regard W_λ as representation of $D^\times$ via the chain of maps $D^\times \hookrightarrow (D \otimes_K L)^\times \cong \operatorname{GL}_2(L) \xrightarrow{\prod_\tau \widetilde{\tau}} \prod_\tau \operatorname{GL}_2(\overline{\mathbb{Q}}_p)$. The isomorphism class of the resulting representation is independent of any choices. We can then regard W_λ as a representation of $\operatorname{GL}_2(\mathcal{O}_K)$ or $\mathcal{O}_D^\times$, by restriction.

Fix a discrete series inertial type $\tau : I_K \to \operatorname{GL}_2(\overline{\mathbb{Q}}_p)$, so that we have finite-dimensional representations $\sigma(\tau)$ and $\sigma^{\operatorname{cr}}(\tau)$ (resp. $\sigma_D(\tau)$) of $\operatorname{GL}_2(\mathcal{O}_K)$ (resp. $\mathcal{O}_D^\times$). Define

$$\sigma(\tau,\lambda) := \sigma(\tau) \otimes W_\lambda$$
$$\sigma^{\operatorname{cr}}(\tau,\lambda) := \sigma^{\operatorname{cr}}(\tau) \otimes W_\lambda$$
$$\sigma_D(\tau,\lambda) := \sigma_D(\tau) \otimes W_\lambda,$$

regarded as representations of $\operatorname{GL}_2(\mathcal{O}_K)$ or $\mathcal{O}_D^\times$ as appropriate. Since $\operatorname{GL}_2(\mathcal{O}_K)$ and $\mathcal{O}_D^\times$ are compact, we may consider the corresponding semisimple $\overline{\mathbb{F}}_p$-representations $\overline{\sigma}(\tau,\lambda)$, $\overline{\sigma}^{\operatorname{cr}}(\tau,\lambda)$ and $\overline{\sigma}_D(\tau,\lambda)$ obtained by reducing a stable lattice and semisimplifying. These representations factor through the quotients $\operatorname{GL}_2(\mathcal{O}_K) \twoheadrightarrow \operatorname{GL}_2(k)$ and $\mathcal{O}_D^\times \twoheadrightarrow l^\times$. In the case $\lambda = 0$ (when W_λ is the trivial representation), we will write $\overline{\sigma}(\tau)$, $\overline{\sigma}^{\operatorname{cr}}(\tau)$ and $\overline{\sigma}_D(\tau)$ for $\overline{\sigma}(\tau,0)$, $\overline{\sigma}^{\operatorname{cr}}(\tau,0)$ and $\overline{\sigma}_D(\tau,0)$.

Let F_λ (respectively F_λ^D) be the representation of $\operatorname{GL}_2(k)$ (respectively $l^\times$) obtained from $W_\lambda|_{\operatorname{GL}_2(\mathcal{O}_K)}$ (resp. $W_\lambda|_{\mathcal{O}_D^\times}$) by taking a stable lattice, reducing mod p, and semisimplifying. The following lemma is trivial.

Lemma 4.1. *We have* $\chi_{F_\lambda}\left(\begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}\right) = \chi_{F_\lambda^D}(x)$ *for each* $x \in k^\times$, *and* $\chi_{F_\lambda}(i(z)) = \chi_{F_\lambda^D}(z)$ *for each* $z \in l^\times \setminus k^\times$.

The following theorem expresses the compatibility of our mod p Jacquet–Langlands correspondence with the reduction modulo p of the inertial correspondence.

Theorem 4.2. *Let* $\tau : I_K \to \operatorname{GL}_2(\overline{\mathbb{Q}}_p)$ *be a discrete series inertial type.*

(1) *Suppose τ is scalar. Then for each highest weight $\lambda \in (\mathbb{Z}_+^2)^{\operatorname{Hom}_{\mathbb{Q}_p}(K,\overline{\mathbb{Q}}_p)}$, we have*

$$\operatorname{JL}(\overline{\sigma}_D(\tau,\lambda)) = \overline{\sigma}(\tau,\lambda) - \overline{\sigma}^{\operatorname{cr}}(\tau,\lambda).$$

(2) *Suppose τ is supercuspidal. Then for each highest weight $\lambda \in (\mathbb{Z}_+^2)^{\operatorname{Hom}_{\mathbb{Q}_p}(K,\overline{\mathbb{Q}}_p)}$, we have*

$$\operatorname{JL}(\overline{\sigma}_D(\tau,\lambda)) = \overline{\sigma}(\tau,\lambda).$$

Proof. Since all of the representations involved are semisimple, it suffices to prove equalities of Brauer characters. By definition we have $\chi_{\overline{\sigma}(\tau,\lambda)} = \chi_{\overline{\sigma}(\tau)}\chi_{F_\lambda}$, $\chi_{\overline{\sigma}^{\operatorname{cr}}(\tau,\lambda)} =$

$\chi_{\overline{\sigma}^{\mathrm{cr}}(\tau)}\chi_{F_\lambda}$, and $\chi_{\overline{\sigma}_D(\tau,\lambda)} = \chi_{\overline{\sigma}_D(\tau)}\chi_{F_\lambda^D}$, so by Lemma 4.1 we may immediately reduce to the case $\lambda = 0$.

In case (1), since everything is compatible with twists by characters we may reduce to the case that τ is the trivial type; but then $\sigma_D(\tau)$ and $\sigma^{\mathrm{cr}}(\tau)$ are the trivial representation, and $\sigma(\tau) = \mathrm{sp}_1$, and the result is immediate from Definition 2.1. In case (2), after twisting we may reduce to the case that $\sigma_D(\tau)$ extends to a minimal representation of $D^\times$, and the result is immediate from Proposition 3.12. $\square$

5. The Breuil–Mézard conjecture

In this section we prove the main theorem of this paper, relating the Breuil–Mézard conjectures for $\mathrm{GL}_2(\mathcal{O}_K)$ and $\mathcal{O}_D^\times$. We begin by recalling the Breuil–Mézard conjecture, reformulated in terms of Grothendieck groups, as in the introduction.

Fix a finite $E/\mathbb{Q}_p$ with ring of integers $\mathcal{O}$, uniformiser ϖ and residue field $\mathbb{F}$, and fix a continuous representation $\overline{\rho} : G_K \to \mathrm{GL}_2(\mathbb{F})$. Let $R^\square$ be the universal lifting ring of $\overline{\rho}$ on the category of complete Noetherian local $\mathcal{O}$-algebras with residue field $\mathbb{F}$. Let τ be an inertial type and λ a weight as in Section 4. Extending E if necessary, we may assume that τ, $\sigma(\tau)$, $\sigma^{\mathrm{cr}}(\tau)$ and $\sigma_D(\tau)$ (when τ is a discrete series type) are all defined over E. Then, there is a quotient $R^{\tau,\lambda}$ of $R^\square$ which is reduced and p-torsion free, and is "universal" for liftings which are potentially semistable of Hodge type λ and inertial type τ. Specifically, we take $R^{\tau,\lambda}$ to be the image of the natural map $R^\square \to (R^\square[1/p])^{\tau,\lambda,\mathrm{red}}$ where $(R^\square[1/p])^{\tau,\lambda}$ is the quotient of $R^\square[1/p]$ constructed in [Kis08, Theorem 2.7.6] (where our λ corresponds to Kisin's $\mathbf{v}$). There is also a universal lifting ring $R^{\tau,\lambda,\mathrm{cr}}$ which is reduced and p-torsion free, and is universal for liftings which are potentially crystalline of Hodge type λ and inertial type τ. In this case, we take $R^{\tau,\lambda,\mathrm{cr}}$ to be the image of the map $R^\square \to (R^\square[1/p])^{\tau,\lambda,\mathrm{cr}}$, where the latter is ring constructed in [Kis08, Cor. 2.7.7]; it is reduced by [Kis08, Theorem 3.3.8]. If R is a complete local Noetherian $\mathcal{O}$-algebra with residue field $\mathbb{F}$, then we write $e(R/\varpi)$ for the Hilbert–Samuel multiplicity of R/ϖ.

Conjecture 5.1. *(The Breuil–Mézard Conjecture for* GL_2.*)*
(1) There is a linear functional $\iota : R_\mathbb{F}(\mathrm{GL}_2(k)) \to \mathbb{Z}$ such that for each τ, λ we have $\iota(\overline{\sigma}(\tau,\lambda)) = e(R^{\tau,\lambda}/\varpi)$.
(2) There is a linear functional $\iota_{\mathrm{cr}} : R_\mathbb{F}(\mathrm{GL}_2(k)) \to \mathbb{Z}$ such that for each τ, λ we have $\iota(\overline{\sigma}^{\mathrm{cr}}(\tau,\lambda)) = e(R^{\tau,\lambda,\mathrm{cr}}/\varpi)$.

Lemma 5.2. *If Conjecture 5.1 holds, then we necessarily have $\iota = \iota_{\mathrm{cr}}$.*

Proof. Since $R^{\tau,\lambda} = R^{\tau,\lambda,\mathrm{cr}}$ and $\sigma^{\mathrm{cr}}(\tau,\lambda) = \sigma(\tau,\lambda)$ unless τ is a scalar type, it is enough to show that ι (and thus ι_{cr}) is uniquely determined by its values on the $\overline{\sigma}(\tau,\lambda)$ for τ non-scalar. We may replace $R_\mathbb{F}(\mathrm{GL}_2(k))$ by $R_\mathbb{F}(\mathrm{GL}_2(k)) \otimes_\mathbb{Z} \overline{\mathbb{Q}}_p$, so it suffices to prove that $\mathcal{C}(\mathrm{GL}_2(k))$ is spanned by the Brauer characters $\chi_{\overline{\sigma}(\tau,\lambda)}$ for τ non-scalar. Now, $\chi_{\overline{\sigma}(\tau,\lambda)} = \chi_{\overline{\sigma}(\tau)}\chi_{F_\lambda}$, and the χ_{F_λ} span $\mathcal{C}(\mathrm{GL}_2(k))$, so the span of the $\chi_{\overline{\sigma}(\tau,\lambda)}$ for τ non-scalar is an ideal in $\mathcal{C}(\mathrm{GL}_2(k))$ (the ideal generated by the $\chi_{\overline{\sigma}(\tau)}$ for τ non-scalar).

Since the maximal ideals in $\mathcal{C}(\mathrm{GL}_2(k))$ are given by the sets of functions which vanish on some semisimple conjugacy class, it suffices to show that for each semisimple conjugacy class, there is some non-scalar type τ such that $\chi_{\overline{\sigma}(\tau)}$ does not vanish on that class; but this follows immediately from the table of Brauer characters in [Dia07, §1]. $\square$

The obvious variant for representations of $D^\times$ is as follows. Let λ and τ be as above. Let $R^{\tau,\lambda,\mathrm{ds}}$ denote the maximal reduced p-torsion free quotient of $R^{\tau,\lambda}$ which is supported on the set of irreducible components of $\mathrm{Spec}\, R^{\tau,\lambda}$ where the associated Weil–Deligne representation is generically of discrete series type. More specifically, if τ is a supercuspidal type, then $R^{\tau,\lambda,\mathrm{ds}} = R^{\tau,\lambda} = R^{\tau,\lambda,\mathrm{cr}}$; if τ is a principal series type, then $R^{\tau,\lambda,\mathrm{ds}} = 0$; while if τ is a scalar type, then $\mathrm{Spec}\, R^{\tau,\lambda,\mathrm{ds}}$ is the union the irreducible components of $\mathrm{Spec}\, R^{\tau,\lambda}$ not occurring in $\mathrm{Spec}\, R^{\tau,\lambda,\mathrm{cr}}$, with the reduced induced scheme structure. Recall that for some τ of supercuspidal type, we chose $\sigma_D(\tau)$ to be one of the two irreducible constituents of the restriction to $\mathcal{O}_D^\times$ of a certain supercuspidal representation; in the statement of the following conjecture, we consider both choices.

Conjecture 5.3. *There is a linear functional $\iota_D : R_{\mathbb{F}}(l^\times) \to \mathbb{Z}$ such that for each discrete series type τ, each algebraic weight λ, and each choice of $\sigma_D(\tau)$ we have $\iota_D(\chi_{\overline\sigma_D(\tau,\lambda)}) = e(R^{\tau,\lambda,\mathrm{ds}}/\varpi)$.*

Remark 5.4. As in the proof of Lemma 5.2, a functional ι_D as in Conjecture 5.3 is necessarily unique. Note that in the case that there are two choices of $\sigma_D(\tau)$, the two possibilities are related by conjugation by ϖ_D, and in the other case $\sigma_D(\tau)$ is invariant under conjugation by ϖ_D. The representation W_λ is also invariant under conjugation by ϖ_D (as it is a representation of $D^\times$). Conjugation by ϖ_D induces the involution $c : x \mapsto x^q$ on $l^\times$, so rather than insisting on allowing both choices of $\sigma_D(\tau)$ in the statement of Conjecture 5.3, we could equivalently have insisted that ι_D be invariant under the action of c, and only used one choice of $\sigma_D(\tau)$.

Before stating our main result, we note that in the case where τ is a scalar type, the potentially semistable deformation ring of weight λ and type τ constructed in [Kis08] is not necessarily reduced. More specifically, we denote by $\widetilde{R}^{\tau,\lambda}$ the image of the map $R^\square \to (R^\square[1/p])^{\tau,\lambda}$; it is p-torsion free, equidimensional and its generic fibre is generically reduced (by [Kis08, Theorem 3.3.4]). The ring $R^{\tau,\lambda}$ is its maximal reduced quotient. Similarly, we may consider quotients $\widetilde{R}^{\tau,\lambda,\mathrm{ds}}$ of the ring $\widetilde{R}^{\tau,\lambda}$ that are p-torsion free and have support consisting of the irreducible components generically of discrete series type. (There need not be a maximal such quotient.) The ring $R^{\tau,\lambda,\mathrm{ds}}$ is the maximal reduced quotient of any such $\widetilde{R}^{\tau,\lambda,\mathrm{ds}}$. If we work with these potentially larger rings $\widetilde{R}^{\tau,\lambda}$ and $\widetilde{R}^{\tau,\lambda,\mathrm{ds}}$, then the question arises as to whether the Hilbert Samuel multiplicities of the special fibres change. The following lemma shows that this is not the case.

Lemma 5.5. *Let R be a complete Noetherian $\mathcal{O}$-algebra with residue field $\mathbb{F}$. Suppose that R is p-torsion free, equidimensional, and that $R[1/p]$ is generically reduced. Then*

$$e(R/\varpi) = e(R^{\mathrm{red}}/\varpi).$$

Proof. Let I denote the kernel of the surjection $R \twoheadrightarrow R^{\mathrm{red}}$. Since R is assumed to be p-torsion free, R^{red} is also p-torsion free and we thus have an exact sequence

$$0 \to I/\varpi I \to R/\varpi \to R^{\mathrm{red}}/\varpi \to 0.$$

Thus $e(R/\varpi) = e(R^{\mathrm{red}}/\varpi) + e(I/\varpi I, R/\varpi)$ (notation as in [Kis09, §1.3]) and we are reduced to showing that $e(I/\varpi I, R/\varpi) = 0$. Since $R[1/p]$ is generically reduced, the localisation $I_\wp$ vanishes for every minimal prime $\wp$ of R. Thus the support of I on R is of dimension strictly smaller than that of R. Since $I \subset R \subset R[1/p]$, each

minimal prime in the support of I is p-torsion free. It follows that the support of $I/\varpi I$ is of dimension strictly smaller than that of R/ϖ. Thus $e(I/\varpi I, R/\varpi) = 0$, as required. $\square$

The main result of this paper is the following.

Theorem 5.6. *Conjecture 5.1 implies Conjecture 5.3.*

Proof. Assume that Conjecture 5.1 holds. Define $\iota_D := \iota \circ \mathrm{JL}$. If τ is supercuspidal, then by Theorem 4.2, we have $\iota_D(\overline{\sigma}_D(\tau, \lambda)) = \iota(\overline{\sigma}(\tau, \lambda)) = e(R^{\tau, \lambda}/\varpi) = e(R^{\tau, \lambda, \mathrm{ds}}/\varpi)$, as required. If τ is scalar, then we see in the same way using Lemma 5.2 that $\iota_D(\overline{\sigma}_D(\tau, \lambda)) = \iota(\overline{\sigma}(\tau, \lambda) - \overline{\sigma}^{\mathrm{cr}}(\tau, \lambda)) = e(R^{\tau, \lambda}/\varpi) - e(R^{\tau, \lambda, \mathrm{cr}}/\varpi) = e(R^{\tau, \lambda, \mathrm{ds}}/\varpi)$. (The last equality follows from [Kis09, Prop. 1.3.4], taking f to be the map $R^{\tau, \lambda} \to R^{\tau, \lambda, \mathrm{cr}} \oplus R^{\tau, \lambda, \mathrm{ds}}$.) $\square$

Corollary 5.7. *Suppose that $K = \mathbb{Q}_p$ and that $p \geq 5$. Then Conjecture 5.3 holds.*

Proof. Under these hypotheses, Conjecture 5.1 holds by the main result of [Paš12]. $\square$

Remark 5.8. It should also be possible to use the main result of [GK12] to prove that there is a functional ι satisfying the conclusion of Conjecture 5.1 whenever $\lambda = 0$ (the only issue being for scalar types, where the results of [GK12] consider only the potentially crystalline, rather than potentially semistable representations; but when $\lambda = 0$, the only representations excluded are ordinary, so it should be possible to prove the automorphy lifting theorems necessary to use the machinery of [GK12]). It would then follow that a functional ι_D as in Conjecture 5.3 exists if we restrict to the case $\lambda = 0$.

Remark 5.9. It may seem to the reader that the proof of Theorem 5.6 is a little too simple, and that we have avoided various technical issues, in particular the formulation of the weight part of Serre's conjecture for $\overline{\rho}$, which are usually present in discussions of the Breuil–Mézard conjecture. However, following [GK12], the weight part of Serre's conjecture can be formulated in terms of the Breuil–Mézard conjecture; namely, the predicted weights for $\overline{\rho}$ are precisely the irreducible representations $\overline{\sigma}$ of $\mathrm{GL}_2(k)$ for which $\iota(\overline{\sigma}) > 0$. (Note that if Conjecture 5.1 is true, then $\iota(\overline{\sigma})$ is positive whenever it is non-zero; this follows from taking τ to be trivial and W_λ to be a lift of $\overline{\sigma}$ in the second part of the conjecture.)

The analogous definition could be made for weights of $D^\times$ (that is, for irreducible representations of $l^\times$). In fact, if we translate the definition of the weight part of Serre's conjecture for quaternion algebras made in [GS11, Definition 3.4] to this language, it is easy to see that this is *precisely* the definition made there.

More precisely, let $\overline{\sigma}$ be an $\mathbb{F}^\times$-character of $l^\times$, and let $\widetilde{\sigma}$ be its Teichmüller lift. Then the discussion before Definition 3.2 of [GS11] shows that $\widetilde{\sigma} = \sigma_D(\tau)$ for some type τ (in fact, the tame type corresponding to $\widetilde{\sigma} \oplus \widetilde{\sigma}^q$ via local class field theory). Taking $\lambda = 0$, we see that $\iota_D(\overline{\sigma}) = e(R^{\tau, 0, \mathrm{ds}}/\varpi) \geq 0$, which is positive if and only if $\overline{\rho}$ has a discrete series lift of weight 0 and type τ. This recovers [GS11, Definition 3.4].

Remark 5.10. Our results in fact give rise to a formula for the predicted $D^\times$ weights of $\overline{\rho}$ in terms of the predicted GL_2 weights of $\overline{\rho}$. Under the perfect pairing

$$R_{\mathbb{F}}(\mathrm{GL}_2(k)) \times R_{\mathbb{F}}(\mathrm{GL}_2(k)) \to \mathbb{Z}$$

which sends two irreducibles $(\overline{\sigma}, \overline{\sigma}')$ to $\dim_{\mathbb{F}} \operatorname{Hom}_{\operatorname{GL}_2(k)}(\overline{\sigma}, \overline{\sigma}')$, we can identify the functional ι with an element $\sum_{\overline{\sigma}} \mu_{\overline{\rho}}(\overline{\sigma})\overline{\sigma}$ of $R_{\mathbb{F}}(\operatorname{GL}_2(k))$. We have a similar pairing

$$R_{\mathbb{F}}(l^{\times}) \times R_{\mathbb{F}}(l^{\times}) \to \mathbb{Z}$$

which allows us to think of ι_D as an element of $R_{\mathbb{F}}(l^{\times})$. Moreover, we may consider the adjoint $\operatorname{JL}^* : R_{\mathbb{F}}(\operatorname{GL}_2(k)) \to R_{\mathbb{F}}(l^{\times})$ of the map JL with respect to these pairings. Since $\iota_D = \iota \circ \operatorname{JL}$, we see that for any element V of $R_{\mathbb{F}}(l^{\times})$, we have

$$(\iota_D, V) = (\iota, \operatorname{JL}(V)) = (\operatorname{JL}^*(\iota), V).$$

In other words, $\iota_D = \operatorname{JL}^*(\iota)$. Note that for any irreducible $\mathbb{F}[\operatorname{GL}_2(k)]$-representation $\overline{\sigma}$, we have

$$\operatorname{JL}^*(\overline{\sigma}) = \sum_{\xi} m_{\xi}(\overline{\sigma})[\xi] + \sum_{\chi} m_{\chi}(\overline{\sigma})[\chi \circ \mathbf{N}_{l/k}]$$

where ξ runs over characters $l^{\times} \to \mathbb{F}^{\times}$ not factoring through $\mathbf{N}_{l/k}$ and χ runs over characters $k^{\times} \to \mathbb{F}^{\times}$ and where:

- $m_{\xi}(\overline{\sigma})$ is equal to the multiplicity with which $\overline{\sigma}$ appears in $\overline{\Theta(\xi)}$ (which is either 0 or 1 by [Dia07, Proposition 1.3]);
- $m_{\chi}(\overline{\sigma})$ is 1 if $\overline{\sigma} = \operatorname{sp}_{\chi}$; it is -1 if $\overline{\sigma} = \chi \circ \det$ and it is 0 otherwise.

Thus, we have:

$$\iota_D = \sum_{\xi} \left(\sum_{\overline{\sigma}} m_{\xi}(\overline{\sigma})\mu_{\overline{\rho}}(\overline{\sigma}) \right) [\xi] + \sum_{\chi} \left(\mu_{\overline{\rho}}(\operatorname{sp}_{\chi}) - \mu_{\overline{\rho}}(\chi \circ \det) \right) [\chi \circ \mathbf{N}_{l/k}].$$

References

[BD12]   Christophe Breuil and Fred Diamond, *Formes modulaires de Hilbert modulo p et valeurs d'extensions Galoisiennes*, 2012.

[BM02]   Christophe Breuil and Ariane Mézard, *Multiplicités modulaires et représentations de* $\operatorname{GL}_2(\mathbf{Z}_p)$ *et de* $\operatorname{Gal}(\overline{\mathbf{Q}}_p/\mathbf{Q}_p)$ *en* $l = p$, Duke Math. J. **115** (2002), no. 2, 205–310, With an appendix by Guy Henniart.

[Car84] H. Carayol, *Représentations cuspidales du groupe linéaire*, Ann. Sci. École Norm. Sup. (4) **17** (1984), no. 2, 191–225.

[Dia07] Fred Diamond, *A correspondence between representations of local Galois groups and Lie-type groups*, L-functions and Galois representations, London Math. Soc. Lecture Note Ser., vol. 320, Cambridge Univ. Press, Cambridge, 2007, pp. 187–206.

[GK12]   Toby Gee and Mark Kisin, *The Breuil–Mézard conjecture for potentially Barsotti–Tate representations*, 2012.

[GS11]   Toby Gee and David Savitt, *Serre weights for quaternion algebras*, Compos. Math. **147** (2011), no. 4, 1059–1086.

[HT01]   Michael Harris and Richard Taylor, *The geometry and cohomology of some simple Shimura varieties*, Annals of Mathematics Studies, vol. 151, Princeton University Press, Princeton, NJ, 2001, With an appendix by Vladimir G. Berkovich.

[Kis08] Mark Kisin, *Potentially semi-stable deformation rings*, J. Amer. Math. Soc. **21** (2008), no. 2, 513–546.

[Kis09] ———, *The Fontaine-Mazur conjecture for* GL_2, J. Amer. Math. Soc. **22** (2009), no. 3, 641–690.

[Kis10] ———, *The structure of potentially semi-stable deformation rings*, Proceedings of the International Congress of Mathematicians. Volume II (New Delhi), Hindustan Book Agency, 2010, pp. 294–311.

[Kut87] Philip Kutzko, *Character formulas for supercuspidal representations of* GL_l, *l a prime*, Amer. J. Math. **109** (1987), no. 2, 201–221.

[Paš12] Vytautas Paškūnas, *On the Breuil–Mézard conjecture*, 2012.

E-mail address: toby.gee@imperial.ac.uk

Department of Mathematics, Imperial College London

E-mail address: geraghty@math.princeton.edu

Princeton University and Institute for Advanced Study